IECO

Volume 9 No 3 Year 2026

University of Sistan and Baluchestan

Iranian Association of Electrical and Electronics Engineers

■ Volume 9
■ No 3
■ Year 2026

International Journal Of

Industrial Electronics Control and Optimization

In This Issue:

Research Article

International Journal Of Industrial Electronics Control and Optimization IECO

Editor-in-Chief's Message

International Journal of Industrial Electronics, Control and Optimization (IECO) Volume 9, No. 3, 2026

It is my pleasure to present Volume 9, Issue 3 (2026) of the International Journal of Industrial Electronics, Control and Optimization (IECO).

This issue reflects the evolving landscape of industrial electronics, advanced control systems, and optimization methodologies. The published articles span key domains including power converter topologies, intelligent diagnostics, optimal control techniques, antenna design, and stability analysis of power systems. Together, they illustrate the strong interplay between rigorous mathematical modeling, engineering design, and emerging data-driven approaches.

A notable trend across several contributions is the integration of artificial intelligence and computational intelligence techniques into classical engineering problems. Such developments are reshaping how industrial systems are analyzed, optimized, and monitored, enhancing both efficiency and reliability.

IECO remains committed to maintaining high scholarly standards through a transparent peer-review process and to fostering international academic collaboration. We sincerely thank our authors, reviewers, and editorial board members for their dedication and professionalism.

We look forward to continued contributions that advance both theoretical insight and practical innovation in industrial electronics, control, and optimization.

Prof. S. Masoud Barakati

Editor-in-Chief

International Journal of Industrial Electronics, Control and Optimization

About Journal

The University of Sistan and Baluchestan entered into strategic partnership with Iranian Association of Electrical and Electronic Engineers (IAEEE) to publish the **International Journal of Industrial Electronics Control and Optimization (IECO)**. The IECO is a refereed international journal which presents to the international scientific community important results of work in these fields, whether in the form of modeling simulation, analysis, fundamental research, development, application, design or real-time implementation. The scope of IECO is broad, encompassing all aspects of Industrial Electronics, Control and Optimization.

Note: International Journal of Industrial Electronics, Control and Optimization (IECO) has qualified to **ACADEMIC RESEARCH JOURNAL (ELMI-PAJOHESHI)** status certified by the ministry of Science, Research and Technology of Iran (No. 231566/3/18 dated 1396/10/09), and is published by the University of Sistan and Baluchestan through a formal partnership (No. 952/2/1500 dated 1395/11/04) with Iranian Association of Electrical and electronic Engineers (IAEEE) in order to develop scientific and research cooperation.

Aims and Scope

International Journal of Industrial Electronics, Control and Optimization (IECO) is a Peer reviewed journal of advanced and state-of-the-art in the science and engineering of Industrial Electronics, Control and Optimization. Its Scope encompasses the applications of Industrial Electronics, power systems, control, optimization and computational intelligence for the enhancement of industrial and manufacturing system and processes. The scope of the journal include the following:

I. Industrial Electronics

- Low and high-power converters
- Renewable energy
- Drive control techniques
- Techniques for advanced power semiconductor devices
- Power quality and utility applications
- Communications
- Flexible AC Transmission Systems (FACTS)
- Control in power electronics
- Electromagnetic and thermal performance of electronic power converters
- Motion control, robotics, sensors and actuators
- Fault detection and diagnosis

- Power systems
- Factory automation, communication, and computer networks

II. Control

- Adaptive control
- Control of process systems
- Control theory
- Data processing
- Design of control systems
- Hybrid systems
- Identification and observation
- Intelligent systems
- Model-predictive control
- Optimal control
- Robust control
- Fractional order systems

III. Optimization

- Ant Colony
- Chaos Theory
- Evolutionary Computing
- Fuzzy Computing
- Hybrid Methods
- Immunological Computing
- Neuro Computing
- Particle Swarm
- Probabilistic Computing
- Rough Sets
- Wavelet

International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Multi-objective Scheduling of Smart Homes Integrated with Renewable Energy Sources and Energy Storage Systems

Mehran Haghighi[1], Hamid Karimi[2], and Shahram Jadid[1]

[1]Department of Electrical Engineering, Center of Excellence for Power System Automation and Operation, Iran University of Science and Technology, Iran.
[2] Department of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran
Corresponding author's email: Karimi.h@qut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 04-May-2025
Received in revised form: 03-July-2025
Accepted: 12-August-2025
Published online:23-Sep-2026

Keywords:
Demand Response,
Home Energy Management System (HEMS),
Customer Satisfaction,
TOPSIS Method,
Household Appliances.

ABSTRACT

A home energy management system optimizes the electrical demand of household appliances according to price-based demand response programs. In this paper, we proposed a price-based demand response for a smart home with different types of appliances according to customer satisfaction, which also used electric and thermal storage systems. In the proposed method, various appliances were considered in the smart home modeled by the energy hub system. A multi-objective daily management is proposed which considered electricity costs and customer satisfaction simultaneously to provide comprehensive management for smart homes. This paper presents a multi-objective optimization approach that not only minimizes operating costs and maximizes customer satisfaction but also reduces environmental impacts through emission reduction, creating a more sustainable energy management system. After solving the multi-objective problem, the technique for order preference by similarity to the ideal solution was used to brank the solutions considering the preference of the decision-maker. The proposed model investigates the response of the smart home energy system in different conditions. Also, stochastic optimization was applied to model the probabilistic nature of demands, PV, and wind energy. The simulation results demonstrate that the proposed method reduces the consumer dissatisfaction by 79.9% and reduces the range from 199 to 40.

NOMENCLATURE

Indices		Eb_{No_x}	No_x gas emissions for the boiler.
t	Index of time periods.	Eb_{co_2}	co_2 gas emissions for the boiler.
i	Index of household appliances.	N_1	The maximum power that can be purchased from the grid.
s	Index of Scenarios.	N_2	The maximum power that can be sold to the grid.
C_i	similarity index.	Eb_{so_2}	so_2 gas emissions for the boiler.
Parameters		E_{so_2}	Amount of so_2 purchased.
A_{non}	Set of non-interruptible appliances.	C_{No_x}	The cost of purchased No_x.
$T_{C,i}$	The most comfortable temperature of the thermally controlled appliance i determined by the user.	E_{No_x}	Amount of No_x purchased.
A_{in}	Set of interruptible household appliances.	Lan_{NS}	Unsupplied energy costs.
A_{ther}	Set of thermostatically controlled appliances.	C_{so_2}	The cost of purchased so_2.
$P_{R,i}^{APP}$	Rated power of the appliance i.	**Variables**	
$T_{L,i}$	Time length of appliance i used from start to end.	J_1	Operation cost of HEMS (cents/kWh).
L_i	Allowed the beginning time of the task for the appliance i.	J_2	Objective function of the user's dissatisfaction level.
		J_3	The cost of greenhouse gas emissions.

U_i	Allowed the deadline of the task for the appliance i	J_4	Unsupplied energy costs.
R_{ESS}^{c}	Maximum charging rate of the ESUs.	$T_{u,i}$	Temperature (air, water, etc) determined by the schedule plan and directly perceived by the user for the appliance i (◦C).
R_{ESS}^{d}	Maximum discharging rate of the ESUs.	Pns	Amount of energy not supplied.
S_{ESS}^{ini}	Initial state-of-energy of the ESUs (kWh).	Hb_{TES}	The heat that the boiler gives to the thermal storage.
S_{ESS}^{min}	Minimum allowed state-of-energy of the ESUs (kWh).	Hb_{WH}	The heat that the boiler consumes for heating water
S_{ESS}^{max}	Maximum allowed state-of-energy of the ESUs (kWh).	Hb_{TL}	The energy that the boiler provides for thermal loads.
ϖ_1	The weight corresponds to J_1 J_3 J_4 .	P_{TES}^{WH}	WH power used to heat water for storage.
ϖ_2	The weight corresponds to J_2.	P_{wh}^{WH}	WH power used to heat hot water.
ε_i	Coefficient representing the importance of appliances.	P_{TL}^{WH}	WH power used for thermal loads.
E_i^{APP}	The required total energy of the task of the appliance i.	H_{TES}^{TL}	The heat that the storage gives for thermal loads.
W_{out}	Outdoor temperature (◦C).	H_{TES}^{WH}	The stored energy is used for hot water consumption.
S_{TES}^{ini}	Initial state-of-energy of the TES (kWh).	P_T^{WH}	Thermal from WH, which is generally used for hot water.
S_{TES}^{min}	Minimum allowed state-of-energy of the TES (kWh).	X_{ij}	Decision matrix.
Ro	Probability.	R_{ij}	Normalized decision matrix.
λ	Constant (1/3600000) for unit conversion.	V_{ij}	Weighted normal matrix.
m	The mass of water consumed.	V_j^+	Positive ideal value.
C_w	Specific heat of water.	V_j^-	Negative ideal value.
P_{PV}	Forecasted power generation from PV system (kW).	S_j^-	The distance from the negative ideal value.
S_{TES}^{max}	Maximum allowed state-of-energy of the TES (kWh).	S_j^+	The distance from the positive ideal value.
Nb	Boiler efficiency.	S_{ESS}	State-of-energy of the ESUs (kWh).
T_{cold}	Temperature of inlet cold water (◦C).	S_{TES}	State-of-energy of the TES (kWh).
M	Mass of water in full storage (kg).	Gb	The amount of gas purchased for the boiler.
θ_i^{up}	Maximum desired temperature (◦C).	Hb	power generated by boiler.
θ_i^{dn}	Minimum desired temperature (◦C).	μ_{grid}	Binary variable: 1 if grid supplying power, 0 else.
$T0$	Initial water temperature in the water heater storage (◦C).	ρ_{wh}	Demand for hot water drawn (kg).
γ,η	Coefficient denoting the thermal condition surrounding the air conditioner.	ζ_i	Customer dissatisfaction associated with the appliance i.
C_{co_2}	The cost of purchased co_2.	u_i^{APP}	Binary variable: 1 if the appliance is on, 0 else.
E_{co_2}	Amount of co_2 purchased.	P_i^{APP}	Power consumption of the appliance i.
Lan_G	gas price.	$W_{k,i}$	Continuous and positive numbers to calculate the temperature distance to the desired temperature.
$P_{must_{run}}$	Power of the non-controllable household appliances must run (kW).	$Z_{k,i}$	Binary variable to determine the work area.
Dh	Thermal load.	μ_{ESS}	Binary variable: 1 if ESUs is charging, 0 else.

I. Introduction

A. MOTIVATION AND BACKGROUND

With the restructuring of power systems and the creation of modern communication infrastructures, more attention has been paid to the participation of consumers in energy management system. The demand side management program is able to reduce the costs of production, transmission and distribution by changing the load profile of the grid. With the implementation of the demand response (DR) program, the amount of consumption during peak hours has been reduced by consumers who are willing to reduce their consumption. As a result, it is avoided to spend extra money to create production capacity for a short period of time every year. The primary goal of a demand response program has always been to smooth the load profile. Creating incentive programs in order to increase the participation of consumers in demand-side management is essential. In order to increase the efficiency of the system, the consumers should take a greater role. Today, people's desire to use smart homes is expanding, while the capabilities of smart homes in the market are more based on comfort and amenities for the user than relying on energy management. Therefore, one should look for solutions to identify the needs and desires of the consumer and obtain consumer satisfaction. Analyzing customer information is necessary to provide differentiated services. Smart home users face lower electricity bills in the short term and lower electricity tariffs in the long term. Experience has shown that consumers are generally not interested in reducing their consumption, especially during peak hours because they prioritize their comfort and convenience. With the development of smart homes, customers have the opportunity to manage their electricity consumption to reduce their electricity costs [1]. Smart meters and home energy management systems (HEMS) play an important role in managing demand response activities in residential areas [2]. The main motivation of the consumer in using HEMS is cost reduction, which requires changing or reducing electricity consumption, which is associated with the disruption of daily comfort [3]. HEMS can enable demand response programs for residential customers [4]. Consumers implement demand response programs in order to achieve economic benefits and prevent blackouts. In fact, although the main reason for consumers' participation can be attributed to financial and economic benefits, but indirectly, they also help to improve the reliability of the grid.In addition to economic benefits and comfort considerations, environmental sustainability has become a critical factor in smart home energy management. The utilization of smart homes can significantly reduce the emission of greenhouse gases because it facilitates the integration of renewable energy in the system. This environmental aspect adds another dimension to the value proposition of HEMS, making them essential tools not only for cost reduction and comfort enhancement but also for building a more sustainable and environmentally friendly residential sector.

B. RELATED WORKS

In recent years, several researches works had been focused on the design and discussion of demand response programs. Ref. [4] studied the impact of satisfaction factors, such as the level of satisfaction with the use of smart home appliances. However, the authors were not considered the uncertainty of demands and renewable generation. In [5], the joint evaluation of dynamic pricing and DR strategies had been studied based on peak limitation and the possibility of bidirectional operation of Electrical Vehicles (EVs) and Energy storage units (ESUs). In [6], risk- constrained bidding strategy was suggested for smart grid considering the plug-in EV and DR applications. A multi-objective decision making operation was suggested in [7] for the simultaneous optimization of water extraction and daily costs. However, the impact of demand- response programs was not studied. Also, the authors in [8] presented a method for predicting household loads. The proposed model investigated the effect of price-based demand response programs on the load pattern of smart homes. In [9], an energy management system is proposed for energy management of distributed energy resources in order to reduce residential energy consumption and efficient use of batteries. The authors in [10] provided an optimal approach for home energy management in which the consumers responded to different drivers of the DR programs. The results showed that different levels of motivation affect consumers' comfort index differently. The authors in [11] proposed a stochastic programming model to optimize the performance of a smart microgrid in the short-term scheduling to minimize the operating costs and greenhouse gas emissions simultaneously. Ref. [12] developed a comprehensive optimization method for the use of different home appliances considering the preferences of customers. In addition, a separate system was designed for charging and discharging storage devices. In [2], the authors presented a smart home model that proposed a two-level algorithm for energy management of home appliances, which also considered reactive loads. [13] studied the role of artificial intelligence in smart homes. Ref [14] investigated a real-time energy scheduling framework for a smart home considering uncertainty. Authors in [15] developed an integrated home energy management system that participated in a demand-side management program and implemented smart home computing with the help of a smart home operation platform. In [3], the authors considered both costs and satisfaction for energy management of households to maximize the user's comfort. However, energy storage and distributed energy sources were not considered. [16] used different energy sources to supply and control the operation of the appliances. Also, the authors investigated the direction of the power flow on the proposed model. The authors of [17] studied the impact of smart thermostat with a home energy management system. Ref. [18] minimized the total energy cost and thermal dissatisfaction of a sustainable smart home with heating, and air conditioning loads for a long period of time. In [19], a MILP model of energy management system based on reducing the cost of electricity consumption is presented. The considered loads were divided into two groups controlled by thermostat and non-thermostat. Battery storage systems and distributed generation systems were also integrated in the proposed model. Ref. [20] examined the use of second life battery energy storage systems (SL-BESSs) to support the home energy management system. In [21], the energy planning of a house that included solar heating, air conditioning and water heating system was investigated in real-time pricing. However, the correlation between electricity price and ambient temperature were not considered. In [22], smart thermostats were used to provide flexibility in demand-side energy management aiming to combine the use of a smart thermostat with a home energy management system to bring more benefits. In [23], a comprehensive class of household appliances was presented. Nevertheless, energy storage devices and renewable energy sources were not used in the suggested model. A multi-objective scheduling framework was developed in [24] to determine the optimal performance of the smart homes. The

objectives of the suggested framework were to minimize the cost of electricity consumption, flatten the consumption profile, and increase the convenience of customers. An energy management system was proposed in [25] that integrated plug-in EV (PEV), solar PV panel, and battery electrical energy storage (BEES), and DR programs. The suggested model evaluated the efficiency of different DR programs on the performance of smart homes. In [26], an operation model of a single house with alternating DERs and comprehensive load types was studied. In order to model and correctly manage household appliances, an intelligent planning scheme based on model predictive control was presented. However, level consumer welfare was not considered. In [27], the authors tried to minimize the electricity cost and thermal discomfort of users through HEMS considering heating, ventilation, and air conditioning (HVAC). This was done through stochastic programming that took into account the uncertainty associated with electricity prices, outdoor temperatures, and the output of distributed energy sources. The authors in [28] presented a predictive home energy management system for a residential building by integrating a plug-in electric vehicle, a photovoltaic array, and a heat pump. In [29], the optimal production structure of a multi-energy system with zero net emissions was investigated by separating the optimization problem into a two-stage investment problem and exploitation sub-problem. In [30], the performance of a home energy management system under the export rate was compared with a home energy management system under TOU tariff. In this paper, the energy hub model was not considered. In [31], an advanced satisfaction-based home energy management system using deep reinforcement learning for planning controllable and removable appliances was presented. In this reference, the proposed model was not transformed into an energy hub model and the uncertainty of renewable resources was not considered. A joint scheduling model of electric and natural gas utilities for HEMS was proposed in [32].In [33], the authors proposed a smart charging approach for off-range electric vehicle chargers in home energy hub (HEH) applications with direct current sources such as photovoltaics and battery storage systems. In [34], an integrated approach for optimal planning and operation of energy hubs considering the effects of wind energy resources was presented. The stable uncertainties of electricity demand, heating, cooling, and also wind power generation was considered in this study. In order to check the uncertainty parameters in the model, various scenarios were created by the Monte Carlo simulation method and then the scenarios was reduced using the K-means method. In the proposed model, the emission reduction was not considered. In [35], a stochastic home energy management system was proposed that integrates renewable energy sources with demand response programs using an improved ABC algorithm, achieving significant cost reductions under various grid constraints. In [36], an energy hub optimization model was proposed using the IPSO algorithm that efficiently coordinates electricity, gas, and thermal grids, resulting in significant operational cost reductions. In [37], a method for power system stability analysis was proposed that identifies coherent generator groups and their oscillation patterns, which can be applied to optimize the integration of renewable energy sources and improve islanding detection in smart energy management systems with distributed generation. In [38], a risk-based decentralized model for peer-to-peer energy trading among smart homes was proposed, which integrates demand-side management, photovoltaic systems, and energy storage while addressing uncertainties in renewable generation and market prices through conditional value-at-risk optimization. In [39], a flexibility-constrained energy management framework for smart energy hubs was developed that incorporates peer-to-peer transactive energy and demand response programs, integrating multiple energy carriers and storage systems to minimize operational and environmental costs while enhancing system flexibility. A number of related research works are compared in Table 1.

C. CONTRIBUTIONS

According to the best of our knowledge, the main contributions of the proposed model are listed as follows:

- This study develops a demand response framework for smart homes featuring diverse categories of household appliances, incorporating an accurate pricing mechanism while accounting for customer satisfaction. Distinct from prior works such as [6], [7], [9], [10], [11], [15], [22], [26–30], and [33–34], the proposed model formulates a multi-objective optimization problem that simultaneously minimizes electricity costs and customer dissatisfaction.
- In contrast to [2–4], [6], [10–15], [20], [22, 23], [25–28], and [30–32], this study models smart homes as integrated energy systems that jointly manage electricity, heating, and cooling demands. This sectorial integration enhances the overall control efficiency, enabling more effective handling of renewable energy source (RES) uncertainties and reducing operational expenditures. Furthermore, the coupled system significantly improves the flexibility of smart homes in terms of both power and energy, thereby enhancing their ability to accommodate RES fluctuations.

TABLE 1. Taxonomy of the model used for designing DR-based HEMS.

References	Hub Energy	Wide variety of appliances	Customer Satisfaction	Uncertainty	Emission reduction	DER	ESUs
[30],[26],[28], [22],[27]	N	Y	N	Y	N	Y	Y
[13]	N	N	Y	N	Y	N	N
[23]	N	Y	N	N	N	N	N
[4],[20]	N	Y	Y	N	N	Y	Y
[14]	N	Y	Y	Y	N	N	Y
[12]	N	Y	Y	Y	N	Y	Y
[32],[25],[31]	N	Y	Y	N	N	N	Y
[3]	N	Y	Y	Y	N	N	N
[9],[2]	N	Y	N	N	N	Y	Y
[33]	Y	Y	N	Y	N	N	Y
[6],[11]	N	N	N	Y	N	Y	Y
[7]	Y	N	N	N	N	Y	N
[29]	Y	N	N	Y	Y	Y	Y
[34]	Y	N	N	Y	N	Y	Y
[10],[15]	N	Y	N	N	N	N	Y
Proposed model	Y	Y	Y	Y	Y	Y	Y

Y/N denotes that the subject is/is not considered.

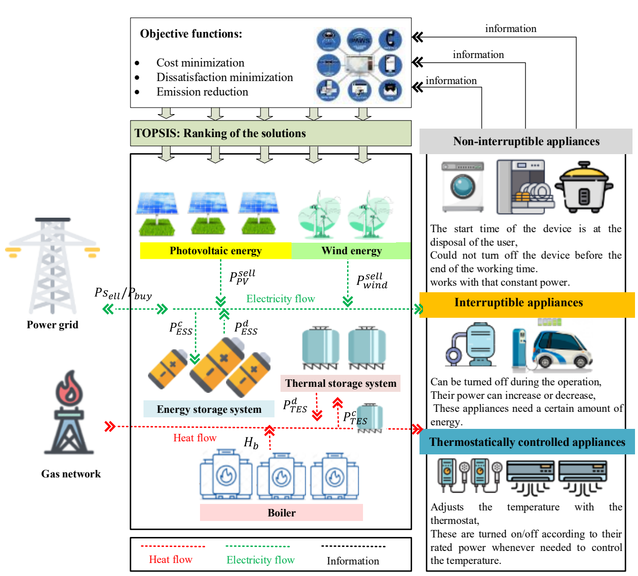


Fig. 1 Structure of the proposed smart home

- In contrast to [2–4], [6, 7], [9–12], [14, 15], [20], [22, 23], [25–28], and [30–34], this study proposes a hybrid environmental–techno-economic scheduling framework for smart homes. The proposed model aims to minimize greenhouse gas emissions alongside electricity costs and customer dissatisfaction, while ensuring compliance with all relevant technical operational constraints.

This paper is organized as follows. Section II describes the methodology and mathematical formulation of the proposed model. Section III introduces the case studies and discussion. Finally, conclusion is presented in Section IV.

II. METHODOLOGY

According to Fig. 1, the home energy management system adjusts the performance of the smart home according to the signal prices and consumer satisfaction. In the proposed model, a smart home with various electrical, heating and cooling loads is considered. The structure of the optimization is presented in Fig. 2. First, loads and PV output are predicted according to weather forecast and historical data. Then, the DR problem is considered according to the forecasted PV output, loads, electricity price, along with the desired user input by the home owner. A home with different types of appliances connected to smart home is considered as case study. Before smart home scheduling, customers

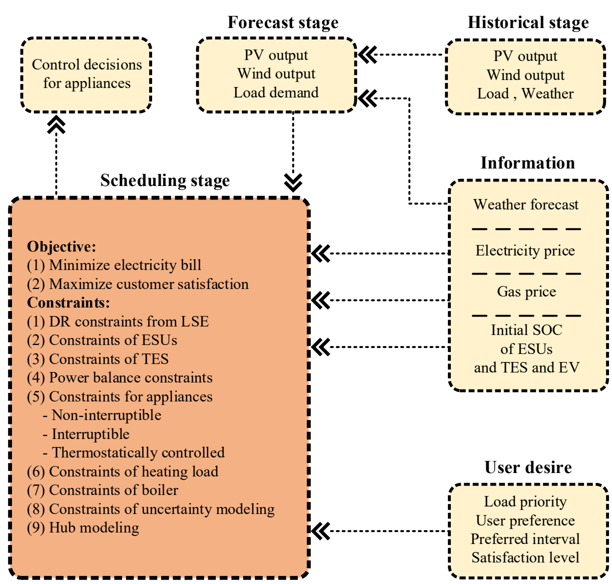


Fig. 2. The optimization framework of HEMS.

record information about the occupants' comfort and lifestyle preferences. Here, the goal is that the consumers participate in the demand response program and manage their energy systems. In this proposed model, the consumer will pays the lowest costs and has the most satisfaction in using their smart devices at the same time. The output of this work is a software tool that optimally schedules appliance usage and local energy generation based on objectives like reducing electricity costs and enhancing user comfort. It considers real-time pricing, user preferences, system limitations, and energy availability to create an efficient operation plan. This tool can be integrated into a home energy management system (HEMS) for smart and automated energy control.

A. MULTI-OBJECTIVE FUNCTION

The existing problem is a multi-objective problem where the first objective function is the house bill and the second objective function is the customer satisfaction level. Using the weighted sum approach, we convert different objective functions into one. The objective function is as Eq. (1):

$$Min\ \varpi_1(J_1 + J_3 + J_4) + \varpi_2 J_2 \qquad \varpi_1 + \varpi_2 = 1 \ , \ \varpi_1 \,\&\, \varpi_2 \in [0, 1] \tag{1}$$

The objective function creates this opportunity for the consumer to decide about optimal performance of their smart appliances. J_1, J_2, J_3, and J_4 in the above equation are respectively expressed as follows.

ϖ_1 is the weight related to J_1 , J_3 , and J_4 and also ϖ_2 is the weight related to J_2 . Also, J_1 refers operating cost of HEMS (cents) and J_2 is the objective function of the participant's level of dissatisfaction. Also, J_3 is the cost of pollution and J_4 is load shading. Eq. (2) represents the operating cost (J_1), which includes electricity purchase costs, revenue from selling excess electricity, and natural gas consumption costs. Eq. (3) quantifies customer dissatisfaction (J_2) by summing the dissatisfaction levels across all appliances, ensuring user comfort preferences are considered. The emission cost (J_3) in Eq. (4) calculates the environmental impact from grid electricity consumption and natural gas combustion, accounting for CO_2, SO_2, and NO_x emissions[29]. Finally, Eq. (5) represents the cost of unsupplied energy (J_4), calculated as the product of unsupplied power and the penalty factor to ensure reliability.

$$J_1 = \sum_{s \in S} Ro(s) * \sum_{t \in T} \left[\lambda_{buy}(t) * P_{buy}(t,s) * \Delta t - \lambda_{sell}(t) * P_{sell}(t,s) * \Delta t\right] + \sum_{t \in T} Lan_G(t) * Gb(t) * \Delta t \qquad \forall t \in T, \forall s \in S \tag{2}$$

$$J_2 = \sum_{i \in A} \zeta i \tag{3}$$

$$J_3 = \sum_{s \in S} Ro(s) * \sum_{t \in T} \left[P_{buy}(t,s) * \left(C_{CO_2} * E_{CO_2} + C_{SO_2} * E_{SO_2} + C_{NO_x} * E_{NO_x}\right)\right] + \sum_{t \in T} Hb(t) * \left(C_{CO_2} * Eb_{CO_2} + C_{SO_2} * Eb_{SO_2} + C_{NO_x} * Eb_{NO_x}\right) \qquad \forall t \in T, \forall s \in S \tag{4}$$

$$J_4 = \sum_{s \in S} Ro(s) * \sum_{t \in T} Pns(t,s) * Lan_{NS}(t) \qquad \forall t \in T, \forall s \in S \tag{5}$$

B. CONSTRAINTS OF HOUSEHOLD APPLIANCES

There are different types of devices that have several advantages and disadvantages in terms of HEMS-based operational strategy. According to the ability of control, household devices can be divided into controllable and uncontrollable appliances. Based on operational characteristics, controllable appliances are classified into:

- Non-interruptible appliances (NIA) such as rice cooker (RC), dishwasher (DW), washing machine (WM).
- Interruptible appliances (IA) such as pool pump (PP) and electric vehicle (EV).
- Thermostatically controlled appliances (TCA): such as water heaters (WH), air conditioners (AC).

1) Non-interruptible appliances (NIA) modeling:

In non-interruptible appliances, when a device is turned on, it must be operated for the same period of time as the device is supposed to work. Also, it should be turned off after finishing the work. The start time of the device is at the disposal of the user, but the user cannot turn off the device before the end of the working time. In NIA, the power is constant, which means that when the device is turned on, it works with that constant power for the number of hours specified by the user. In this paper, three NIAs are considered such as a washing machine and two dishwashers. The constraints related to NIA during the planning horizon are defined as Eq. (6) to (9):

$$u_i^{APP}(t) = 0 \quad \forall t \in [1, L_i) \cup (U_i, N_T] \ , \forall i \in A_{non} \tag{6}$$

$$P_i^{APP}(t) = u_i^{APP}(t) * P_{R,i}^{APP} \quad \forall t \in [L_i, U_i] \ , \ \forall i \in A_{non} \tag{7}$$

$$\sum_{t=j}^{j+T_{L,i}-1} u_i^{APP}(t) \geq T_{L,i} * \left(u_i^{APP}(j) - u_i^{APP}(j-1)\right) \quad \forall j \in \left(L_i, U_i - T_{L,i} + 1\right) \ \forall i \in A_{non} \tag{8}$$

$$\zeta i = \sum_{t=L_i}^{U_i} (1 + \varepsilon_i * t) * u_i^{APP}(t) \quad \forall i \in A_{non} \tag{9}$$

Equation (6) shows the time when the devices are off. u_i^{APP} is a binary variable for equipment, if it is 1, it means that the device is on, and if it is zero, it means that it is off. Equation (7) shows the power consumption of the device during operation. Equation (8) shows the minimum time the device stays on, while Eq. (9) shows the satisfaction level of consumers during interval L_i to U_i . This equation shows that the earlier the equipment is turned on, the smaller the level of consumer dissatisfaction is, and it is better for the user because he seeks to minimize the level of dissatisfaction and maximize his level of comfort in using the equipment. Index R refers to the rated (nominal) value of the corresponding parameter. For instance, $P_{R,i}^{APP}$ denotes the rated power consumption of the appliance i.

2) Interruptible appliances (IA) modeling:

In interruptible appliances, when the equipment are turned on, they can be turned off during operation, or even their power can increase or decrease. The important thing about this type of appliances is that these appliances must consume a certain amount of energy during the day and need a certain amount of energy. In this paper, A pool pump and two electric vehicles are considered. Constraints related to interruptible appliances are defined as Eq. (10) to (13):

$$P_i^{APP}(t) = 0 \quad \forall t \in [1, Li\,) \cup (Ui, N_T] \quad \forall i \in A_{in} \tag{10}$$

$$\sum_{t=Li}^{Ui} P_i^{APP}(t) * \Delta t \geq E_i^{APP} \quad \forall i \in A_{in}, \forall t \in T \tag{11}$$

$$0 \leq P_i^{APP}(t) \leq P_{R,i}^{APP} \quad \forall t \in T, \forall i \in A_{in} \tag{12}$$

$$\zeta i = \sum_{t=L_i}^{U_i} (1 + \varepsilon_i * t) * u_i^{APP}(t) \quad \forall i \in A_{in} \tag{13}$$

Equation (10) shows that if the interruptible appliances are outside of the operating range, the device must be turned off. Equation (11) states that the total power consumption of the interruptible appliances must be at least equal to the energy that the user specifies, in other words, the amount of energy required for the appliances. Eq. (12) specifies the power consumption limits. Eq. (13) shows the satisfaction level of consumers using interruptible appliances in the time interval L_i to U_i. Similar, index R refers to the rated value of the corresponding parameter.

3) Thermostatically controlled appliances (TCA) modeling: Thermostat-controlled appliances include air conditioners, electric water heaters, etc., where electrical energy is consumed to adjust the temperature. In this study, WH and AC, which are the most important thermostat-controlled appliances in a smart home, are used. These appliances adjust the temperature with the thermostat that is available in their system. For example, the electric water heater tries to keep the temperature of the heated tank constant or control it within the desired range of the user. These devices are turned on or off according to their rated power whenever needed to control the temperature. Constraints related to water heater is defined as Eqs. (14)- (18):

Equation (14) shows the power consumption limits. It is clear from Eq. (15) that the power consumption of water heater from hour 1 to t must be at least equal to the power required to heat water from hour 1 to t time. Equation (16) shows the amount of required energy for heating. Equation (17) states

$$0 \leq P_i^{APP}(t) \leq P_{R,i}^{APP} \quad \forall t \in T, \forall i \in WH \tag{14}$$

$$\sum_{k=1}^{t} P_T^{WH}(k,s) * \Delta t \geq \sum_{k=1}^{t} \rho_{wh}(t) \quad \forall t \in T, \forall i \in WH \tag{15}$$

$$\rho_{wh}(t) = \lambda * m(t) * C_w * \left(T_{u,i}(t) - T_{cold}\right) \quad \forall t \in T, \forall i \in WH \tag{16}$$

$$\sum_{k=1}^{t} P_T^{WH}(k,s) * \Delta t \leq \lambda * M * C_w * \left(\theta_i^{up} - T0\right) + \sum_{k=1}^{t} \rho_{wh}(k) \quad \forall t \in T, \forall i \in WH \tag{17}$$

$$\theta_i^{dn} \leq T_{u,i}(t) \leq \theta_i^{up} \quad \forall t \in T, \forall i \in WH \tag{18}$$

that the power consumption by the water heater from the first hour to t must be at most equal to the energy that remains in the water heater in addition to the energy needed to heat the water, Also, it is clear from Eq. (18) that the temperature of hot water should always be between the range determined by the user. Constraints related to air conditioner is defined as Eqs. (19) to (21):

$$0 \leq P_i^{APP}(t) \leq P_{R,i}^{APP} \quad \forall t \in T, \forall i \in AC \tag{19}$$

$$T_{u,i}(t) = T_{u,i}(t-1) + \eta\left[W_{out}(t) - T_{u,i}(t-1)\right] + \gamma * P_i^{APP}(t) * \Delta t \quad \forall t > 1, t \in T, \forall i \in AC \tag{20}$$

$$\theta_i^{dn} \leq T_{u,i}(t) \leq \theta_i^{up} \quad \forall t \in T, \forall i \in AC \tag{21}$$

Equation (19) shows the power consumption limit of air conditioning. Equation (20) shows the relationship between indoor temperature and AC energy consumption. Equation (21) presents that the temperature inside the building should always remain within the range determined by the user. To measure users' satisfaction with TCA, the piecewise linear function shown in Fig. 3 is used. It can be seen that the deviation from $T_{C,i}$ results in a penalty for the objective function.

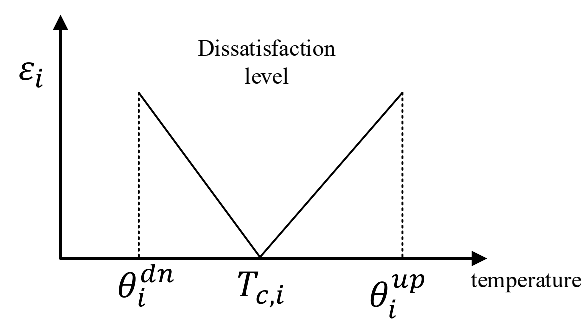


Fig. 3. Dissatisfaction function for thermostatically controlled appliances.

Satisfaction coefficients for TCAs are defined that the operating temperature of these appliances must be between the minimum and maximum temperature set by the user. An optimal temperature for the appliances, which is $T_{c.i}$, is considered as the distance from this temperature and $T_{c.i}$. Move to higher or lower temperature, the level of well-being of the residents will decrease. Now by linearizing this curve, Eqs. (22) to (24) are obtained:

$$W_{1,i} \le Z_{1,i} \quad , W_{2,i} \le Z_{1,i} \quad + Z_{2,i} \quad , \; W_{3,i} \le Z_{2,i}$$
$$W_{1,i} + W_{2,i} + W_{3,i} = 1 \quad , W_{k,i} \ge 0 \quad (k = 1,2,3) \tag{22}$$
$$Z_{1,i} \quad + Z_{2,i} \quad = 1 \, , Z_{k,i} \quad = 0 \;\; OR \; 1 \quad (k = 1,2)$$

Three coefficients W_1, W_2, and W_3 are continuous and positive numbers. These coefficients calculate the temperature distance from $T_{c.i}$. Two binary variable Z_1 and Z_2 are also taken into account, which determines the working area of the device in the curve. Based on this, the temperature inside the building is modeled according to Eq. (23) based on W_1, W_2, and W_3 which depends on the desired temperature and the minimum and maximum temperature.

$$T_{u,i}(t) = \theta_i^{dn} * W_{1,i} + T_{c,i} * W_{2,i} + \theta_i^{up} * W_{3,i} \qquad \forall i \in A_{ther} \tag{23}$$

According to Eq. (24), the temperature is equal to $T_{c.i}$, the dissatisfaction is equal to zero. Accordingly, dissatisfaction is calculated only in terms of W_1 and W_3.

$$\zeta i = W_{1,i} * \varepsilon_i + W_{3,i} * \varepsilon_i \qquad \forall i \in A_{ther} \tag{24}$$

All these three categories of mentioned appliances have a series of importance coefficients that are determined by the user, that is, when the well-being of the residents of the house is measured, the effect of using these appliances on the well-being of the residents is not the same.

C. CONSTRAINTS OF ENERGY STORAGE SYSTEM

Here the storage system refers to a set of batteries that can store electrical energy and deliver it to other devices or even to the main grid when needed. Constraints related to the electric energy storage systems are as Eqs. (25)- (30):

$$P_{ESS}^{use}(t,s) + P_{ESS}^{sold}(t,s) = \eta_{ESS}^{d} * P_{ESS}^{d}(t,s) \qquad \forall t \in T, \forall s \in S \tag{25}$$

$$0 \le P_{ESS}^{c}(t,s) \le R_{ESS}^{c} * \mu_{ESS}(t) \qquad \forall t \in T, \forall s \in S \tag{26}$$

$$0 \le P_{ESS}^{d}(t,s) \le R_{ESS}^{d} * (1 - \mu_{ESS}(t)) \qquad \forall t \in T, \forall s \in S \tag{27}$$

$$S_{ESS}(t,s) = S_{ESS}(t-1,s) + \eta_{ESS}^{c} * P_{ESS}^{c}(t,s) * \Delta t - \frac{1}{\eta_{ESS}^{d} * P_{ESS}^{d}(t,s) * \Delta t} \quad \forall t > 1, \in T, \forall s \in S \tag{28}$$

$$S_{ESS}(t,s) = S_{ESS}^{ini} + \eta_{ESS}^{c} * P_{ESS}^{c}(t,s) * \Delta t - \frac{1}{\eta_{ESS}^{d} * P_{ESS}^{d}(t,s) * \Delta t} \forall s \in S, if\ t = 1 \tag{29}$$

$$S_{ESS}^{min} \le S_{ESS}(t,s) \le S_{ESS}^{max} \qquad \forall t \in T, \forall s \in S \tag{30}$$

It is clear from Eq. (25) that if the storage system is discharged, the total discharged power should either be used at home or sold to the grid. Equations (26) and (27) show the charging power and discharging power limits, respectively. Eq (28) shows the storage energy level in each hour and scenario. Eq (29) shows the energy level at each time. Finally, Eq. (30) presents the limit of stored energy in the storage system.

D. CONSTRAINTS OF THERMAL ENERGY STORAGE

A thermal energy storage can be used in a smart home. There is a tank in this home where hot water is stored and its loss can be reduced by isolating it. Constraints related to the thermal energy storage unit are in equations 31 to 33:

$$S_{TES}(t,s) = S_{TES}(t-1,s) + \eta_{TES}^{c} * P_{TES}^{c}(t,s) * \Delta t - \frac{1}{\eta_{TES}^{d} * P_{TES}^{d}(t,s) * \Delta t} \forall t > 1, \in T, \forall s \in S \tag{31}$$

$$S_{TES}(t,s) = S_{TES}^{ini} + \eta_{TES}^{c} * P_{TES}^{c}(t,s) * \Delta t - \frac{1}{\eta_{TES}^{d} * P_{TES}^{d}(t,s) * \Delta t} \quad \forall s \in S, if\ t = 1 \tag{32}$$

$$S_{TES}^{min} \le S_{TES}(t,s) \le S_{TES}^{max} \qquad \forall t \in T, \forall s \in S \tag{33}$$

Equation (31) shows the energy level of thermal storage in every hour and every scenario. It is clear that Eq. (32) presents the energy level of thermal storage in the first hour. Equation (33) limits the stored energy between minimum and maximum values.

E. PV MODELLING

A small Photovoltaic panel with a 1 kW capacity is also used in this smart home. Equation (34) states that the generating power of PV unit can be used by indoor appliances or injected into the grid.

$$P_{PV}^{use}(t,s) + P_{PV}^{sell}(t.s) = P_{PV}(t,s) \qquad \forall t \in T, \forall s \in S \tag{34}$$

F. WIND MODELLING

A small wind turbine with a 1 kW capacity is also used in this smart home. It is clear from Eq. (35) that the actual production power of the wind turbine can be used for two purposes. It can be injected into the grid or consumed by the appliances inside the home:

$$P_{wind}^{use}(t,s) + P_{wind}^{sell}(t,s) = P_{wind}(t,s) \qquad \forall t \in T, \forall s \in S \tag{35}$$

G. POWER BALANCE CONTRAINT

Equation (36) expresses the power balance that ensures the generated energy must be equal to the consumption for each hour[7].

H. POWER EXCHANGED WITH THE GRID

Equation (37) shows the total amount of power sold to the grid:

$$P_{buy}(t,s) + P_{PV}^{use}(t,s) + P_{wind}^{use}(t,s) + P_{ESS}^{use}(t,s) + Pns(t,s) = P_{ESS}^{c}(t,s) + \sum_{i\in A} P_i^{APP}(t) + P_{must_{run}}(t) \quad \forall t \in T, \forall s \in S \tag{36}$$

$$P_{sell}(t,s) = P_{PV}^{sell}(t,s) + P_{ESS}^{sell}(t,s) + P_{wind}^{sell}(t,s) \quad \forall t \in T, \forall s \in S \tag{37}$$

Equations (38) and (39) ensure that power purchased from the grid and power injected into the grid cannot occur simultaneously:

$$P_{buy}(t,s) \le N_1 * \mu_{grid}(t) \quad \forall t \in T, \forall s \in S \tag{38}$$

$$P_{sell}(t,s) \le N_2 * \left(1 - \mu_{grid}(t)\right) \quad \forall t \in T, \forall s \in S \tag{39}$$

N_1 and N_2 are the maximum power that can be bought from the grid and the maximum power that can be sold to the grid, respectively.

I. UNCERTAINTY MODELING

In this paper, wind and solar production sources are used as renewable sources, whose output power is a function of wind speed and solar radiation, respectively. In this study, the uncertainties caused by these sources have been modeled using scenario generation and using the Weibull and Beta distribution functions.

1) Uncertainty model of PV power:

In this simulation, to check the uncertainty related to solar energy, Beta distribution function is used, which is defined by two parameters α and β. For each solar power, these two parameters are adjusted based on the distribution of the series of numbers we have and produce a scenario. In this method, the predicted value of PV power is considered as the average value. In the Beta distribution, we need two parameters, the average predicted power per hour and its standard deviation. In this method, the probability of occurrence of each scenario should be calculated as follows. After calculating the probability of occurrence of the scenario, it should be normalized[11].

$$f(x;\alpha,\beta) = f_{p_{pv}} = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\times\Gamma(\beta)} (p_{pv})^{\alpha-1} (1-p_{pv})^{\beta-1} & if\ p_{pv} \in [0, p_{pv}(si)] \\ 0 & other\ wise \end{cases} \tag{40}$$

For the Beta distribution, in order to reduce the volatility of different scenarios, the value of the standard deviation should be reduced.

2) Uncertainty model of wind power:

It is essential to use a suitable model in predicting wind turbine production. The Weibull distribution function is defined by two parameters, which are the scale parameter and the shape parameter. For each wind power, these two parameters are set based on the distribution of the series of numbers and produce a scenario. In the Weibull distribution, the probability of occurrence is calculated as follows. After calculating the probability of the scenario, it must be normalized. If the random variable X has support with non-negative and continuous values, so that its density function is written as follows, then this random variable has a Weibull distribution with parameters k and λ[11]:

$$f(x,\lambda,\mathrm{k}) = RO_{WIND} = \begin{cases} \frac{\mathrm{k}}{\lambda}\left(\frac{x}{\lambda}\right)^{\mathrm{k}-1} e^{-(x/\lambda)^{\mathrm{k}}} & x \ge 0 \\ 0 & x < 0 \end{cases} \tag{41}$$

In the above relation, λ is the scale parameter and k is the shape parameter. Also, x refers to p_{wind}. In this case, it is displayed as $x \sim w(\lambda, \mathrm{k})$ where x has a Weibull distribution with parameters λ and k. Both these parameters have positive values. If the dispersion of the scenarios is required to be further apart, the shape parameter or k should be reduced.

J. HUB MODELING

An energy hub represents an interface between different energy sub structures. In this paper, smart home appliances are considered such as dishwashers, washing machines, pool pumps, electric vehicles, uncontrollable loads, heat pumps, water heaters, boilers, thermal loads, electric storage, thermal storage, gas network, electricity grid, and renewable sources to model the smart home by an energy hub system. Constraints related to boiler and hub energy are in the form of Eqs. (42)-(49) [34]:

$$P_{TES}^{c}(t,s) = Hb_{TES}(t,s) + P_{TES}^{WH}(t,s) \quad \forall t \in T, \forall s \in S \tag{42}$$

$$P_{TES}^{d}(t,s) = H_{TES}^{TL}(t,s) + H_{TES}^{WH}(t,s) \quad \forall t \in T, \forall s \in S \tag{43}$$

$$Hb(t) = Gb(t) * Nb \quad \forall t \in T \tag{44}$$

$$0 \le Hb(t) \le Hb^{max} \quad \forall t \in T \tag{45}$$

$$Hb(t) = Hb_{WH}(t,s) + Hb_{TL}(t,s) + Hb_{TES}(t,s) \quad \forall t \in T, \forall s \in S \tag{46}$$

$$P_i^{APP}(t) = P_{TES}^{WH}(t,s) + P_{wh}^{WH}(t,s) + P_{TL}^{WH}(t,s) \quad \forall t \in T, \forall s \in S, \forall i \in WH \tag{47}$$

$$Dh(t) = H_{TES}^{TL}(t,s) + Hb_{TL}(t,s) + P_{TL}^{WH}(t,s) \quad \forall t \in T, \forall s \in S \tag{48}$$

$$P_T^{WH}(t,s) = H_{TES}^{WH}(t,s) + Hb_{WH}(t,s) + P_{wh}^{WH}(t,s) \quad \forall t \in T, \forall s \in S \tag{49}$$

It is clear from Eq. (42) that the charging power of the thermal storage must either be supplied by the boiler or WH. Equation (43) states that the power discharged by the thermal storage is consumed by the thermal loads or heats water. Equation (44) shows the production power of the boiler. Equation (45) presents the production limit of the boiler. It is clear from Eq. (46) that the generated heat of the boiler is used for heating water, thermal loads stored in the storage tank. From Eq. (47) it is understood that the power produced by WH can be stored in a storage tank, used for heat water or thermal loads. Also, Eq. (48) states that the thermal loads must either be supplied by the storage device, supplied by the boiler, or supplied by WH. Finally, Eq. (49) determines the total power consumption of heat water.

K. TOPSIS METHOD

After solving the multi-objective problem for the residential hub, a set of optimal Pareto solutions is recommended, which must be chosen from the user's point of view. TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) is a widely used and practical multi-criteria decision-making (MCDM) method. This method provides different advantages. It ranks alternatives based on their proximity to the ideal and negative-ideal solutions, making it intuitive and easy to implement. TOPSIS can simultaneously handle both beneficial and non-beneficial criteria without requiring assumptions about data distribution or utility functions. It also allows for flexible weighting of criteria and produces interpretable results. In this study, TOPSIS is employed to select the optimal demand response strategy considering multiple conflicting technical, economic, and operational criteria. In this method, m options are evaluated by n indicators. It is assumed that the desirability of each index is uniformly increasing or decreasing, which is defined for positive ideal (A^+) and negative ideal (A^-) options. The value of the indicators is between 0 and 1 and is mostly based on mathematical calculations. The biggest value shows the higher priority, which means that it is a suitable option. The best option has the minimum distance from the positive ideal solution and the furthest distance from the negative ideal solution. This TOPSIS model is implemented in eight steps according to the following steps:

1) Creating a decision matrix from the collected information and creating a data matrix based on n indicators and m options in Eq. (50) where i options and j represent indicators. X_{ij} is the i-th index value for the j-th option.

2) The decision-making matrix must be normalized by Eq. (51) so that its data is descaled. Then, by standardizing the indices, the domain of X_{ij} values is converted into a standard domain and the standardized values of R_{ij} indices are obtained as the standard matrix of Eq. (52).

$$X_{ij} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \quad (i = 1,\dots,n)\ ,\ \ (j = 1,\dots,n) \tag{50}$$

$$r_{ij} = \frac{X_{ij}}{\sqrt{\sum_{i=1}^{n} {X_{ij}}^2}} \tag{51}$$

$$R_{ij} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix} \tag{52}$$

3) Determining the weight of each index based on Eq. (53). Indices with more weight are more important, and the sum of the weights must be equal to 1, where W_j is the weight of the i-th index. It should be noted that the weights of the objectives are assigned according to the decision-maker's priorities and can vary depending on the specific requirements and preferences of the system under study.

$$\sum_{j=1}^{n} W_j = 1 \tag{53}$$

4) Converting the step 2 decision matrix into a weighted unscaled matrix. V_{ij} is the weighted normalized matrix elements which is created based on Eqs. (54) and (55):

$$v_{ij} = r_{ij} \times w_j \tag{54}$$

$$V_{ij} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{m1} & v_{m2} & \cdots & v_{mn} \end{bmatrix} \tag{55}$$

5) Finding the positive ideal solution (A^+) by determining the maximum value based on equation (56) and the negative ideal solution (A^-) by determining the minimum value based on equation (57). V_j^+ is the positive ideal value for the jth index and V_j^- is the negative ideal value for the jth index.

$$A^+ = (V_1^+, V_2^+, V_3^+, \dots, V_n^+) \qquad \{(max_i\, v_{ij} | j \in J), (min_i v_{ij} | j \in J')\} = v_j^+ \tag{56}$$

$$A^- = (V_1^-, V_2^-, V_3^-, \dots, V_n^-) \qquad \{(min_i\, v_{ij} | j \in J), (max_i v_{ij} | j \in J')\} = v_j^- \tag{57}$$

6) Calculation of the distance between the positive ideal and the negative ideal based on the Euclidean method. S_i^+ is the distance of the i-th option from the positive ideal that is calculated based on Eq. (58). S_i^- is the distance of the i-th option from the negative ideal that is calculated based on Eq. (59).

7) Calculation of the relative closeness of the ideal solution based on the index C_i based on Eq. (60) which is defined to combine the values of S_i^+ and S_i^- and compare the options:

Table 2. User of desired comfort preferences.

Type	Appliances	Preferred temperature	Duration (h)	Preferred interval [,]	Rated power(kW)	Required energy(kW)	Coefficient resenting importance
NIA	WM	-	2	[8,23]	1.8	-	0.1
	DW1	-	1	[13,16]	1.2	-	0.2
	DW2	-	1	[17,23]	1.2	-	0.3
IA	PP	-	-	[2,15]	0.8	1.6	0.1
	EV1	-	-	[6,24]	2.2	6.5	0.2
	EV2	-	-	[11,23]	1.7	5	0.3
TCA	AC	27	-	-	2	-	1
	WH	55	-	-	3	-	2
HA	Boiler	-	-	-	2	-	-

$$S_i^+ = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^+)^2} \tag{58}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n}(v_{ij} - v_j^-)^2} \tag{59}$$

$$C_i = \frac{S_i^-}{S_i^+ + S_i^-} \tag{60}$$

8) Ranking options based on their relative proximity to the ideal solution and the similarity index C_i .This value varies between 0 and 1. The closer the similarity index value to one, the more superior the option is which means that the option is ranked better.

III. CASE STUDY

A. DESCRIPTION OF DESIGNED TEST SYSTEM

A home with various appliances connected to HEMS is considered as a test system. The costumers register their information that reflects their preferences for comfort and desired lifestyle. Some of the registered information which makes the user comfortable and other information about the equipment are given in Table 2. Other specifications related to thermostat-controlled appliances, which include water heaters and air conditioners, are listed in Tables 3 and 4, respectively. The specifications related to the electrical storage are given in Table 5 and the specifications related to the thermal storage are given in Table 6.The price of energy exchanged with the grid is shown in figure (4). The outside temperature is shown in Fig. 5. The amount of hot water consumption of the smart home is demonstrated in Fig. 6. Also, there are uncontrollable loads in the smart home, such as lighting loads, as shown in Fig. 7.

B. RESULT

In order to evaluate the effect of the proposed method on the operation cost and customer's welfare, three case studies have been designed. The designed case studies are as follows:

- Case Study 1: In this case, $\varpi_1 = 1$ and $\varpi_2 = 0$ are considered, which means that the objective function is to reduce the operation cost without considering customer satisfaction. In this case, the well-being of consumers is not important in using smart home devices, and the only goal is to optimize the cost of payment. This case is considered as a base case in this study.
- Case study 2: In this case, $\varpi_1 = 0.5$ and $\varpi_2 = 0.5$ are considered. In this case, the proposed model simultaneously considers both cost payment and customer satisfaction. Also, both objectives have the same preferences from the smart home point of view.
- Case Study 3: In this case, weight coefficients are changed from 0 to 1 and the best solution is selected by TOPSIS. Actually, this case study considers different preferences between objective functions.

Table 3. Specifications related to the water heater model

$T_{c,i}$ (°C)	$T_{0_{WH}}$ (°C)	M (kg)	θ_i^{dn} (°C)	θ_i^{up} (°C)	T_{cold} (°C)	C_w (J/kg/C)	λ (kWh/J)
55	49	140	45	68	30	4200	1/3600000

Table 4. Specifications related to the Air conditioning model

$T_{0_{AC}}$(°C)	$T_{c,i}$(°C)	θ_i^{dn} (°C)	θ_i^{up} (°C)	γ	η
23	27	23	29	-9.5	0.9

Table 5. Specifications related to the electric storage model

S_{ESS}^{max} (kWh)	S_{ESS}^{min} (kWh)	S_{ESS}^{ini} (kWh)	S_{ESS} (kWh)	R_{ESS}^{c} (kWh)	R_{ESS}^{d} (kW)	η_{ESS}^{c}	η_{ESS}^{d}
7.20	0.80	2.80	8	1.80	1.80	0.95	0.95

Table 6. Specifications related to the Thermal storage model

S_{TES}^{max} (kWh)	S_{TES}^{min} (kWh)	S_{TES}^{ini} (kWh)	S_{TES} (kWh)	R_{TES}^{c} (kWh)	R_{TES}^{d} (kWh)	η_{TES}^{c}	η_{TES}^{d}
10	0	2.50	11	1	1	0.95	0.95

(1) Case study 1:

The main goal of this case study is payment cost minimization. The simulation results show that the consumer payment cost is 79 cents and the dissatisfaction value is 199. Smaller dissatisfaction shows the higher the residents' welfare level. For the proposed model, 10 scenarios have been considered due to uncertainty in power generation by renewable sources. The energy transactions with the grid, encompassing both import and export, are depicted in Fig. 8. It can be observe that the purchasing electricity from the grid during off-peak hours is more than other hours because the electricity prices are low. Therefore, the customers prefer to purchase electricity from the grid. However, they reduce the purchasing electricity from peak hours because of high electricity prices. The charge and discharge power diagram of the electric storage and the charging level of the storage for case study 1 is shown in Fig. 9. Figure 9 shows that the electric storage is charged during off-peak times because the electricity prices are low. They discharged the stored energy during peak hours with high prices to reduce the total cost of the system. The charging and discharging capability of energy storage systems can effectively mitigate the uncertainty associated with renewable energy generation by balancing supply and demand in real time.The power consumption of home appliances is shown in Fig. 10 for case study 1. The types of equipment used in the smart home include washing machine, dishwasher 1 and 2, pool pump, electric vehicle 1 and 2, air conditioner, water heater, and boiler. The amount of energy consumption of each device is given along with the preset hour of its use. It can be observed that the proposed model attempts to achieve the lowest electricity bill by buying electricity from the grid during low

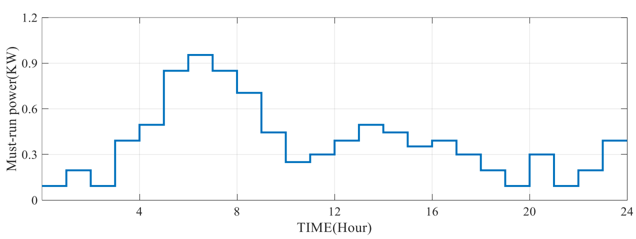


Fig. 7. Must-run power over a day.

price periods and charging the ESUs. The household appliances keep the lowest level of comfort in advance. According to Fig. 10, it can be observed that during the early hours of the day, the boiler is actively utilized due to the relatively low gas prices at that time. This operational strategy reflects the cost-effective use of available energy resources, whereby the system prioritizes gas-based heating when it is economically favorable. By doing so, the model minimizes energy expenses while ensuring that the thermal demands of the household are adequately met. This also allows the thermal storage unit to be charged in preparation for later hours when gas prices are higher. According to Fig. 11, during the same early hours when gas prices are relatively low, the thermal storage system is simultaneously charged while fulfilling the immediate thermal load requirements. This operational strategy reflects an effective energy management approach, wherein the boiler is utilized to generate heat at a lower cost, and the surplus thermal energy is stored for later

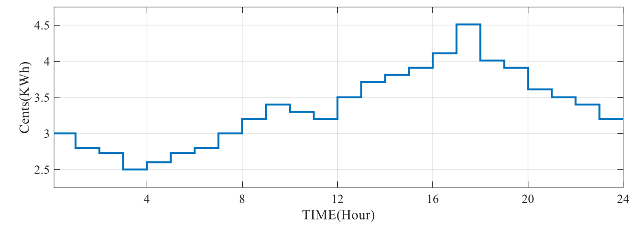


Fig 4. Information on the price of buying and selling electricity from the grid.

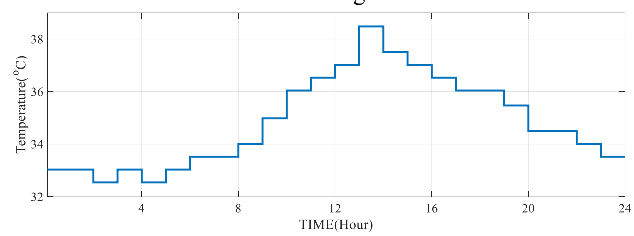


Fig. 5. Outdoor temperature over a day.

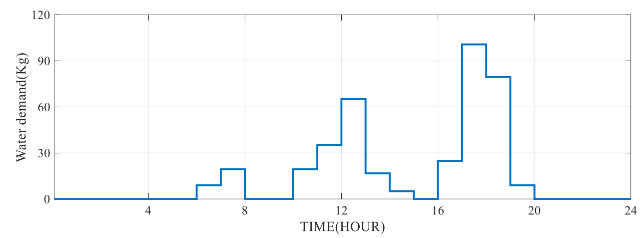


Fig. 6. Water demand over a day.

use. As gas prices increase during subsequent periods, the stored thermal energy is discharged to meet the heating demand, thereby avoiding the need for expensive fuel consumption during peak price hours. This illustrates the model’s capability to shift thermal energy usage from high-cost to low-cost periods, enhancing both economic efficiency and system flexibility. Such proactive utilization of thermal storage not only reduces operational costs but also contributes to the resilience and sustainability of the smart home energy system.

(2) Case study 2:

In this case, both the consumer's welfare and payment costs are considered with the same importance. The simulation results show that the payment cost by the consumer is 93 cents, and the amount of dissatisfaction is 38. Compared to case study 1, the level of consumer dissatisfaction has been significantly reduced. In this case study dissatisfaction as the objective function, the proposed model decreases it from 199 cents to 93 cents. The energy trading to the electricity grid in case study 2 is shown by Fig. 12. During off-peak periods, characterized by lower electricity prices, consumers are more likely to increase their energy usage or shift flexible loads to these hours in order to minimize their overall electricity costs. In contrast, during peak periods with higher prices, consumption is reduced or deferred. The proposed model effectively captures this behavior by optimizing the load scheduling to align with dynamic pricing signals, thereby improving economic efficiency while maintaining user satisfaction.

The consumed power of home appliances in case study 2 is shown in Fig. 13. Compared to case study 1, it is clear that some items such as DW2 and PP have been transferred to the previous times. The charging and discharging power diagram of the electric storage unit for case study 2 is shown in Fig. 14. Energy storage units are charged during off-peak hours when electricity prices are low, and subsequently discharged during peak demand periods to support the household's energy needs. This strategic operation not only enhances load flexibility but also contributes to cost reduction by leveraging the price differential between off-peak and peak hours. As shown in Fig. 15, the thermal storage unit is simultaneously charged when the boiler is in operation to meet the thermal loads. When the gas prices increase, the stored thermal energy is discharged to supply heating needs, thereby reducing the overall operational cost of the smart home.

(3) Case study 3:

In this case, we apply all of the available weight coefficients between zero and 1 and calculate 100 different solutions. This case shows the sensitivity of payment costs and dissatisfaction with the weight coefficients. Among these solutions, the best option is determined by the TOPSIS method. According to the TOPSIS ranking method, it can be seen that the optimal solution for ϖ_1 is $0.87 - 0.94$ and Also, for the weighting factorϖ_2 is $0.13 - 0.06$. In this case, the amount of consumer dissatisfaction is 40 and the payment cost is 80 cents. The TOPSIS method selected this solution because it offers an optimal balance between operation cost and consumer dissatisfaction. A slight increase in dissatisfaction can lead to substantial cost savings, making the trade-off efficient. Besides, system constraints tend to favor cost reduction without significantly compromising comfort. The chosen solution also ranks best by minimizing the distance to the ideal point and maximizing the distance from the worst, aligning with TOPSIS criteria.The amount of energy trading with the electricity grid in case study 3 is presented in Fig. 16. Due to the lower electricity prices during off-peak hours, consumers increase their energy transactions with the electricity grid. This increased activity is driven by the economic incentive to take advantage of cheaper electricity costs. Also, during peak hours 14-22, when electricity prices are substantially higher than at other times, substantially, the volume of energy transactions decreases to reduce the energy costs. The planning power of household appliances is shown in Fig. 17.

Electric vehicles are charged during off-peak periods to prevent increased operational costs associated with high electricity prices. Besides, dishwashers are scheduled to operate during low-price off-peak hours, reflecting a deliberate strategy to minimize energy expenses. Also, Figs. 18 and 19 demonstrate the charge and discharge power

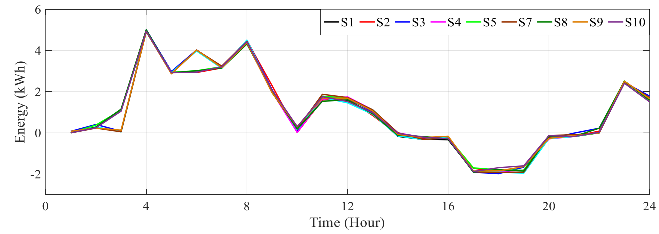


Fig. 8. Hourly purchase of energy from the grid or sale of energy to the grid for case 1.

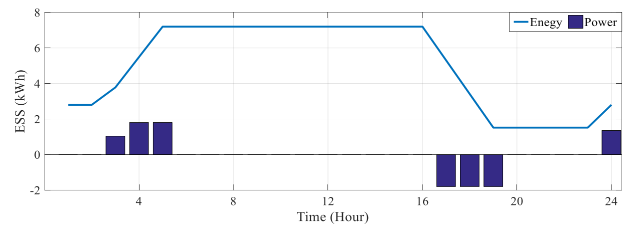


Fig. 9. Energy change of ESUs in one day for case 1.

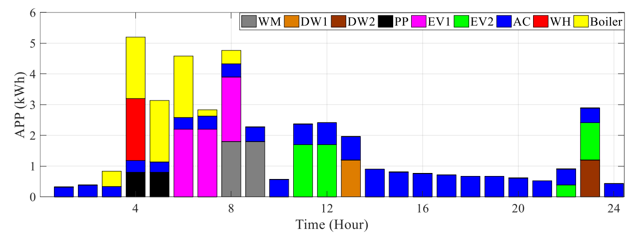


Fig. 10. Scheduling of household appliances for case 1.

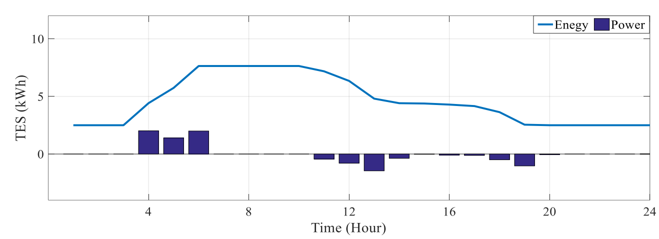


Fig. 11. Energy change of thermal storage in one day for case 1.

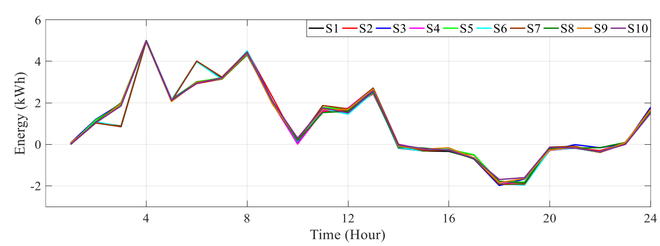


Fig. 12. Hourly purchase of energy from the grid or sale of energy to the grid for case 2.

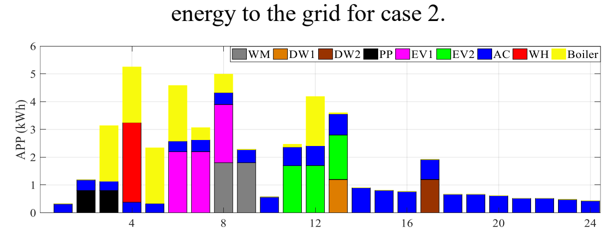


Fig. 13. Operation of smart home appliances for case 2.

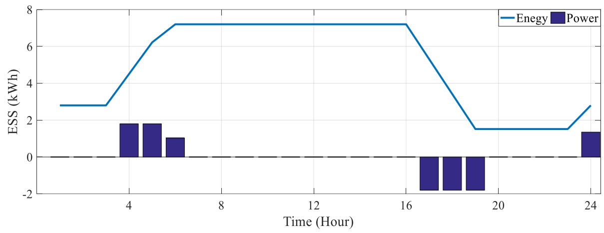


Fig. 14. Energy change of ESUs in one day for case 2.

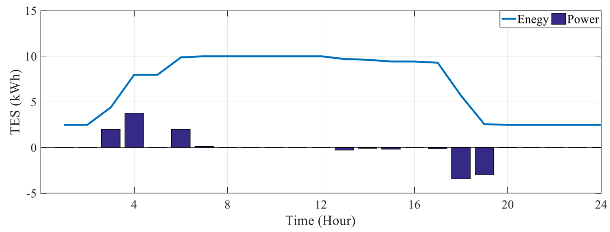


Fig. 15. Energy change of thermal storage in one day for case 2.

diagram of the electric and thermal storage devices, respectively. The results of Figs. 18 and 19 show that the energy storage devices are charged during low electricity prices and discharged during peak hours when prices are high. These charging and discharging performances enable the cost-saving opportunity for customers to reduce total cost.

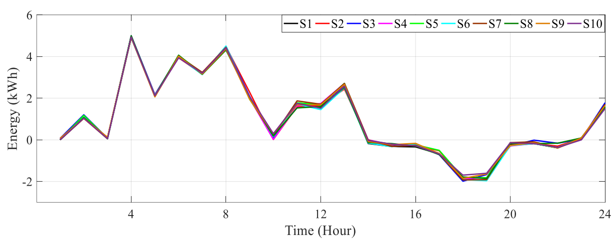


Fig. 16. Hourly purchase of energy from the grid or sale of energy to the grid for case 3.

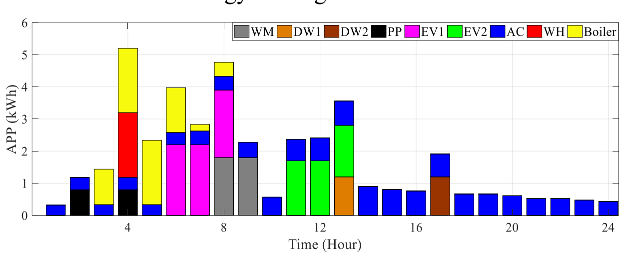


Fig. 17. Schedule of household appliances for case 3.

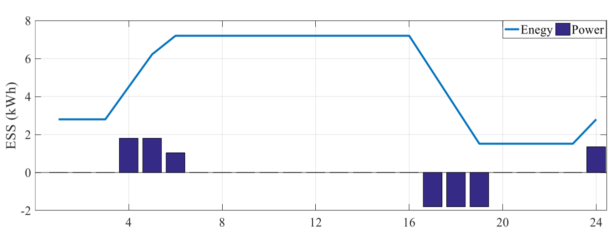


Fig. 18. Energy change of ESUs in one day for case 3.

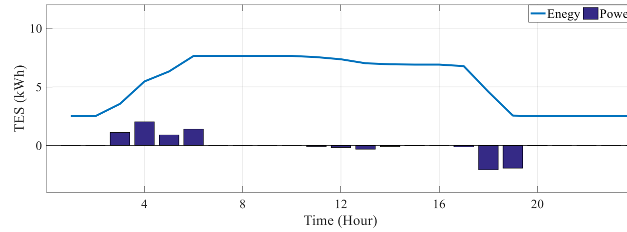


Fig. 19. Energy change of thermal storage in one day for case 3.

V. CONCLUSION

In this paper, a new price-based home energy management is proposed which tries to reduce costs and increase consumer satisfaction. In this model, several case studies have been discussed and all types of smart home appliances with various characteristics has been considered in the proposed framework. Also, with a developed satisfaction model, the proposed HEMS can provide several flexible solutions with different levels of user satisfaction for residents. Different case studies have been considered to evaluate the impact of weight coefficients on smart home performance. To use the model, a balance must be found between the cost and the welfare level, which is expressed as the optimal response. In optimal response Compared to case study 1, by paying one cent more, the customer's well-being increases significantly, and also compared to case study 2, by paying 13 cents less, he can get the appropriate amount of well-being. In future work, the peer-to-peer energy trading among several HEMS will be studied.

REFERENCES

[1] E. Shirazi, A. Zakariazadeh, S. Jadid, Optimal joint scheduling of electrical and thermal appliances in a smart home environment, Energy Conversion and Management, vol.106. December 2015.

[2] Jafarpour, Pourya, Mehrdad Setayesh Nazar, Miadreza Shafie-khah, and Joao PS Catalao. "Resiliency assessment of the distribution system considering smart homes equipped with electrical energy storage, distributed generation and plug-in hybrid electric vehicles." Journal of Energy Storage, Vol. 55, 2022, 105516.

[3] Wang, Huilong, Yongbao Chen, Jing Kang, Zhikun Ding, and Han Zhu. "An XGBoost-Based predictive control strategy for HVAC systems in providing day-ahead demand response." Building and Environment, Vol. 238, 2023, 110350.

[4] Babaei, Masoud, Ahmadreza Abazari, Mohammad Mahdi Soleymani, Mohsen Ghafouri, S. M. Muyeen, and Mohammad TH Beheshti. "A data-mining based optimal demand response program for smart home with energy storages and electric vehicles." Journal of Energy Storage, Vol. 36, 2021, 102407.

[5] O. Erdinc, N. G. Paterakis, T. D. P. Mendes, A. G. Bakirtzis, and J. P. S. Catalão, "Smart household operation considering bi-directional EV and ESS utilization by real-time pricing-based DR," IEEE Trans. Smart Grid, vol. 6, May 2015.

[6] P. Afzali, M. Rashidinejad, A. Abdollahi and A. Bakhshai, "Risk-Constrained Bidding Strategy for Demand Response, Green Energy Resources, and Plug-In Electric Vehicle in a Flexible Smart Grid," in IEEE Systems Journal, vol. 15, March 2021.

[7] Mir Jalal Vahid Pakdel, Farnaz Sohrabi, Behnam Mohammadi-Ivatloo,Multi-objective optimization of energy and water management in networked hubs considering transactive energy,Journal of Cleaner Production,Volume 266,2020.

[8] N. G. Paterakis, A. Taşcıkaraoğlu, O. Erdinç, A. G. Bakirtzis, and J. P. S. Catalão, "Assessment of Demand-Response-Driven Load Pattern Elasticity Using a Combined Approach for Smart Households," IEEE Transactions on Industrial Informatics, vol. 12, 2016.

[9] Paramvir Singh, Sandeep Dhundhara, Yajvender Pal Verma, Nisha Tayal,Optimal battery utilization for energy management and load scheduling in smart residence underdemandresponse scheme,Sustainable Energy, Grids and Networks,Volume 26,2021.

[10] X. Lu et al., "Optimal Bidding Strategy of Demand Response Aggregator Based On Customers' Responsiveness Behaviors Modeling Under Different Incentives," in IEEE Transactions on Industry Applications, vol. 57, no. 4, pp. 3329-3340, July-Aug. 2021.

[11] G.R. Aghajani, H.A. Shayanfar, H. Shayeghi,Demand side management in a smart micro-grid in the presence of renewable generation and demand response,Energy,Volume 126,2017.

[12] Ubaid ur Rehman, Kamran Yaqoob, Muhammad Adil Khan,Optimal power management framework for smart homes using electric vehicles and energy storage,International Journal of Electrical Power & Energy Systems,Vol 134,2022.

[13] K. Kwon, S. Lee and S. Kim, "AI-Based Home Energy Management System Considering Energy Efficiency and Resident Satisfaction," in IEEE Internet of Things Journal, vol. 9, no. 2, pp. 1608-1621, 15 Jan.15, 2022.

[14] W. Hou, K. Zhang, A. Chen and T. Liu, "A Real Time Energy Management Strategy for Smart Home Considering the Uncertainty," 2020 IEEE 9th International Power Electronics and Motion Control Conference (IPEMC2020-ECCE Asia), 2020.

[15] T. Li, Y. Xiao and L. Song, "Integrating Future Smart Home Operation Platform With Demand Side Management via Deep Reinforcement Learning," in IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 921-933, June 2021.

[16] C. Wang, F. Liu, J. Wang, F. Qiu, W. Wei, S. Mei, and S. Lei, "Robust risk-constrained unit commitment with large-scale wind generation: An adjustable uncertainty set approach," IEEE Trans. Power Syst. vol. 32, Jan. 2017.

[17] A. Can Duman, Hamza Salih Erden, Ömer Gönül, Önder Güler,"A home energy management system with an integrated smart thermostat for demand response in smart grids"Sustainable Cities and Society, Vol. 65, 2021.

[18] L. Yu, T. Jiang, and Y. Zhou, "Online energy management for a sustainable smart home with an HVAC load and random occupancy," IEEE Trans. Smart Grid, vol. 10, 2019.

[19] Paterakis, N.G. et al. Optimal household appliances scheduling under day-ahead pricing and load-shaping demand response strategies. IEEE Transactions onIndustrial Informatics, 2015.

[20] Y. Deng, Y. Zhang, F. Luo and G. Ranzi, "Many-Objective HEMS Based on Multi-Scale Occupant Satisfaction Modelling and Second-Life BESS Utilization," in IEEE Transactions on Sustainable Energy, vol. 13, April 2022.

[21] H. Nguyen, H. Nguyen and. D. Le "Energy Management for Households with Solar Assisted Thermal Load Considering Renewable Energy and Price Uncertainty," Smart Grid, IEEE Transactions on, vol. 6, 2015.

[22] A. Can Duman, Hamza Salih Erden, Ömer Gönül, Önder Güler, A home energy management system with an integrated smart thermostat for demand response in smart grids, Sustainable Cities and Society, Volume 65,2021.

[23] O. ALIÇ and Ü. B. FİLİK, "Optimal Management of Household Loads Based on Consumer Elasticity and Appliance Diversity," 2020 International Conference on Smart Energy Systems and Technologies (SEST), 2020.

[24] Khalid, A. et al. Towards dynamic coordination among home appliances usingmulti-objective energy optimization for demand side management in smart buildings. Ieee Access, 2018.

[25] A. Sangswang and M. Konghirun, "Optimal Strategies in Home Energy Management System Integrating Solar Power, Energy Storage, and Vehicle-to-Grid for Grid Support and Energy Efficiency," in IEEE Transactions on Industry Applications, vol. 56, Oct. 2020.

[26] B. Liang, W. Liu, L. Sun, Z. He and B. Hou, "Economic MPC-Based Smart Home Scheduling With Comprehensive Load Types, Real-Time Tariffs, and Intermittent DERs," in IEEE Access, vol. 8,2020.

[27] L. Yu, T. Jiang and Y. Zou, "Online Energy Management for a Sustainable Smart Home With an HVAC Load and Random Occupancy," in IEEE Transactions on Smart Grid, vol. 10, March 2019.

[28] M. Yousefi, A. Hajizadeh, M. N. Soltani and B. Hredzak, "Predictive Home Energy Management System With Photovoltaic Array, Heat Pump, and Plug-In Electric Vehicle," in IEEE Transactions on Industrial Informatics, vol. 17, Jan. 2021.

[29] Mahdi Azimian, Vahid Amir, Saeid Javadi, Economic and Environmental Policy Analysis for Emission-Neutral Multi-Carrier Microgrid Deployment, Applied Energy,Volume 277,2020.

[30] P. Munankarmi, H. Wu, A. Pratt, M. Lunacek, S. P. Balamurugan and P. Spitsen, "Home Energy Management System for Price-Responsive Operation of Consumer Technologies Under an Export Rate," in IEEE Access, vol. 10, 2022.

[31] A. Forootani, M. Rastegar and M. Jooshaki, "An Advanced Satisfaction-Based Home Energy Management System Using Deep Reinforcement Learning," in IEEE Access, vol. 10, 2022.

[32] K. Li, P. Zhang, G. Li, F. Wang, Z. Mi and H. Chen, "Day-Ahead Optimal Joint Scheduling Model of Electric and Natural Gas Appliances for Home Integrated Energy Management," in IEEE Access, vol. 7, 2019.

[33] H. R. Gholinejad, J. Adabi and M. Marzband, "Hierarchical Energy Management System for Home-Energy-Hubs Considering Plug-In Electric Vehicles," in IEEE Transactions on Industry Applications, vol. 58, Oct. 2022.

[34] S.A. Mansouri, A. Ahmarinejad, M. Ansarian, M.S. Javadi, J.P.S. Catalao,Stochastic planning and operation of energy hubs considering demand response programs using Benders decomposition approach,International Journal of Electrical Power & Energy Systems,Volume 120,2020.

[35] M. Haghighi, M. R. Masoudi, H. Karimi and H. R. Sezavar, "Stochastic Framework for Smart Home Energy Management: Integrating Demand Response and Renewable Resources Using Improved ABC Algorithm," 2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG) ", IEEE, 2025.

[36] M. R. Masoudi, M. Haghighi and H. R. Sezavar, "Optimization of Energy Hub Performance Using the Improved Particle Swarm Optimization (IPSO) Algorithm for Integrated Energy Systems," 2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG) ", IEEE, 2025.

[37] H. Karimi, H. R. Sezavar, S. Hasanzadeh, M. Haghighi and E. Heydarian-Forushani, "Identification of Effective Factors on Formation Generator Coherent Groups for Islanding Detection for Renewable Energy Integration," 2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG) ", IEEE, 2025.

[38] Mohammadi, N., Rashidinejad, M., Abdollahi, A., & Afzali, P. (2022). A new risk-based decentralized model for peer-to-peer energy trading among smart homes. International Journal of Industrial Electronics, Control and Optimization, 5(3), 231–240.

[39] Riki, A, Oukati Sadegh, M, and Narouei, O. “Flexibility-constrained energy management of smart energy hubs considering peer to peer transactive energy and demand response program”, International Journal of Industrial Electronics Control and Optimization, 8(1), pp. 67-82 (2025).

Mehran Haghighi received the B.Sc. degree in electrical engineering from Shahid RajaeeTeacher Training University, Tehran, Iran, in 2019, and the M.Sc. degree from the Department of Electrical Engineering, Iran University of Science and Technology, in 2022. He is currently Phd Candidate of power system in Shiraz University. His research interests include power system optimization, Smart grid and energy management.

Hamid Karimi received his B.Sc. and M.Sc. degrees in Electrical Engineering from Shahid Beheshti University (SBU) and Iran University of Science and Technology (IUST), Tehran, Iran, in 2015 and 2017, respectively. He obtained his Ph.D. in Electrical Engineering from Iran University of Science and Technology (IUST) in 2022. His research interests include power system operation optimization, multi-objective optimization, and game theory. He is currently an Assistant Professor at the Faculty of Electrical and Computer Engineering, Qom University of Technology (QUT), Iran.

Shahram Jadid received the Ph.D. degree from the Indian Institute of Technology, Mumbai, India, in 1991. He is currently a Professor with the Department of Electrical Engineering, Iran University of Science and Technology, Tehran, where he is also a member of the Center of Excellence for Power System Automation and Operation (CEPSAO). His research interests include smart grids, power system operation and restructuring, and load and energy management.

IECO
International Journal Of
Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

A Guaranteed Monotonically Convergent Optimal Iterative Learning Control for Multiple-Input Multiple-Output Continuous-Time Linear Time-Varying Systems

Naser Taghvamanesh | Ali Madady

Department of Electrical Engineering, Tafresh University, Tafresh 39518-79611, Iran.
Corresponding author's email: madadi@tafreshu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 08-February-2025
Received in revised form: 26-July-2025
Accepted: 16-August-2025
Published online: 23-Sep-2026

Keywords:
Iterative learning control,
Linear time-varying systems,
Linear quadratic tracking,
Monotonic convergence,
Trajectory tracking.

ABSTRACT

This paper introduces a novel optimal iterative learning control scheme for continuous-time systems with multiple-inputs and multiple-outputs and linear time-varying dynamics. While iterative learning control has been extensively studied in the discrete-time domain, the development of optimal iterative learning control for continuous-time systems remains limited due to the lack of lifted-formulations and associated mathematical challenges. The proposed method transforms the original optimal iterative learning control problem into a linear quadratic tracking-like problem, enabling the derivation of an explicit close-loop control law that ensures both tracking performance and control effort minimization. Unlike many existing approaches that rely on learning algorithms involving derivative terms, which are often sensitive to measurement noise, the proposed design avoids such terms and remains computationally efficient. Moreover, the monotonic convergence of the tracking error and the associated cost function are proved by rigorous mathematical analysis. Theoretical results are supported by four comprehensive simulation examples, including comparisons with several existing iterative learning control methods. Quantitative evaluations confirm that the proposed optimal scheme significantly outperforms previous techniques in terms of convergence speed and error reduction rate. This contribution offers a new framework for the optimal control of continuous-time systems with multiple inputs and outputs in repetitive tasks and provides a foundation for future extensions to constrained, nonlinear, or partially measurable systems.

I. Introduction

Iterative learning control (ILC) is an effective method for managing systems that perform repetitive tasks over a finite time interval [1, 2]. The key feature of ILC is the utilization of information from previous iterations to enhance the performance of the system in subsequent executions [3, 4]. This iterative process allows the system to progressively improve from iteration to iteration, which makes ILC particularly valuable in applications requiring high precision and repeatability, such as robotics, manufacturing, and biomedical fields. Since its introduction [5], ILC has seen a variety of design approaches and has been successfully applied across a wide range of practical scenarios [6, 7, 8].

Notably, various approaches to ILC have been proposed to address different challenges in control systems. For example, in [9], an adaptive repetitive controller is introduced to reject periodic disturbances with an unknown period. In [10], two ILC methods leveraging neural iterative learning identifiers are proposed for a class of unknown time-varying nonlinear systems. The design of ILC for continuous-time linear systems is reformulated as a two-dimensional stabilization problem in [11]. Subsequently, a dynamic output feedback control strategy is developed to effectively resolve this challenge. A data-driven ILC method, integrating predictive control and point-to-point ILC, is proposed in [12] for a class of unknown, repetitive, non-affine nonlinear single-input single-output (SISO) systems. The issue of quantized ILC, which utilizes encoding and decoding mechanisms for networked control systems (NCSs) with bandwidth-constrained communication links, is addressed in [13]. Additionally, a piecewise-type ILC law is designed for linear, discrete, time-varying systems with different initial time points using segment compensation techniques in [14]. Lastly, an error-tracking-driven adaptive back-stepping ILC is proposed in [15] for a type of nonlinear parametric strict-feedback systems performing repetitive tasks over finite intervals.

On the other hand, various optimization-based control strategies, such as genetic algorithms (GA) and particle swarm optimization (PSO), have been widely used to design optimal controllers for complex systems. For instance, a novel approach for the optimal control of robotic manipulators using PSO is presented in [16]. However, these methods fundamentally differ from ILC. While GA and PSO are global optimization techniques that explore the entire parameter space to find an optimal solution, ILC improves control performance through repetitive task execution, refining the control signal based on errors from previous iterations. This makes ILC particularly well-suited for systems performing repetitive tasks, where data from past executions can be utilized to enhance performance over time.

This unique characteristic of ILC has given rise to a category that combines optimization techniques with ILC for controlling systems that perform repetitive tasks. Since an optimal ILC can significantly improve both convergence speed and overall system performance [17], there has been growing interest among researchers in developing optimization-based ILC methods. Below, some of the key studies in this area are outlined.

For instance, in [18], a model-based ILC method is introduced, generalizing ILC using a quadratic performance criterion for time-varying linear constrained systems affected by both deterministic and stochastic disturbances. In [19], a parameter optimization-based ILC approach is presented as a new paradigm for solving the ILC problem in discrete-time linear time-invariant (LTI) systems. A recursive optimal algorithm minimizing the input error covariance matrix is proposed in [20] to determine the optimal forgetting and learning gain matrices for a P-type ILC in linear discrete-time-varying systems with arbitrary relative degrees.

In [21], a feed-forward parameter-optimal high-order ILC algorithm is proposed, noted for its ease of implementation. The extension of the classical proportional-integral-derivative (PID) controller to linear repetitive discrete-time systems is addressed in [22], where an optimal design method is introduced for determining the PID learning gains. Mishra et al. [23] further developed optimization-based ILC schemes for constrained discrete-time LTI systems. In [24], two paradigms of norm-optimal ILC and parameter-optimal ILC are explored for multivariable LTI systems.

Further studies include [25], which established an extended PID-type ILC with four independent learning gains, and [26], which proposed a non-lifted norm-optimal ILC method for linear time-varying (LTV) systems. In [27] and [28], rational basis functions and a novel optimal N-parametric ILC design for discrete-time LTI systems are presented to improve extrapolation properties and design flexibility.

To enhance the convergence rate, [29] introduced an optimal approach for online tuning of PID-type ILC. A robust optimal ILC scheme for updating PID gains in linear systems affected by initial state errors and disturbances is proposed in [30]. The high-performance tracking problem for discrete-time SISO and LTI systems with unknown dynamics is tackled in [31], where two parameter-optimal ILC approaches are introduced, one based on a data-driven algorithm and the other using reinforcement learning.

Zhuang et al. [32] proposed an optimal ILC algorithm for LTI MIMO discrete-time systems with non-uniform trial lengths under input constraints. In [33] and [34], norm-optimal ILC in integer-valued control and optimal ILC for constrained systems with bounded uncertainties are examined. Furthermore, in [35] and [36], an optimal feedforward ILC for discrete-time LTI, SISO systems is developed, combining performance cost and task flexibility. This method outperforms standard norm-optimal ILC in both flexibility and performance.

Finally, in [37] and [38], gradient-like and alternating projection-based optimal ILC designs are presented for LTI continuous-time SISO and MIMO systems, respectively. Meanwhile, in [39] and [40], data-driven tracking control for LTI discrete-time ILC systems and computationally efficient iterative optimization frameworks for nonlinear batch processes are proposed, respectively.

As observed in the aforementioned studies, the majority of optimization-based ILCs have been developed for discrete-time systems. This is primarily due to the fact that, in discrete systems, time is represented as an integer-valued variable, and thus the number of time points per iteration is finite, denoted by M. This enables a compact formulation, known as the lifted-form [17, 24] (or super-vector form [41, 42]), which simplifies the system in the repetition domain by eliminating the time variable and utilizing the Markov parameters of the system. The lifted-form approach reduces the optimal ILC design for discrete-time systems to a parametric and static optimization problem [22, 28, 42], making it solvable.

Despite these advances in discrete-time systems, designing optimal ILCs for continuous-time systems remains an under-explored area. The key challenge stems from the absence of a compact lifted representation in continuous time. Since time is a continuous variable, the infinite-dimensional nature of the signal space makes it difficult to formulate the learning problem as a static optimization problem. As such, only a limited number of studies have addressed optimal ILC in continuous-time settings, and most of them are restricted to scalar systems or assume time-invariant dynamics. One possible approach to overcome this challenge could be to discretize the system and then apply the lifted-form method. However, to achieve sufficient accuracy in the discretization, a very small step size would be required. This would lead to significantly large dimensions for the vectors and matrices in the resulting lifted-form, thereby causing a substantial increase in the computational volume and complexity.

This paper addresses this gap by proposing an optimal ILC framework for continuous-time systems, specifically for MIMO linear time-varying (LTV) systems. To the best of the authors' knowledge, this approach has not been explored in previous research. The closest results to those presented in this paper can be found in Chapter 9 of [42] for time-invariant systems. However, there are fundamental differences, which are detailed in Section 3. In this work, we propose an alternative approach to prove convergence, avoiding the use of complex mathematical concepts such as the properties of self-adjoint operators in Hilbert spaces.

To present our approach more clearly, we propose a novel optimal iterative learning control (OPILC) method for continuous-time systems with MIMO dynamics and time-varying characteristics. This method reformulates the ILC problem into a linear quadratic tracking-like (LQT) problem, facilitating the derivation of an explicit closed-loop control law that guarantees both accurate tracking and minimized control effort. This approach is especially advantageous for applications where monotonic convergence of the tracking error is of paramount importance. In contrast to previously developed continuous-time ILC methods, which often rely on derivative-based learning algorithms [1], [2], [51], that may be susceptible to measurement noise, the proposed method eliminates these terms, thereby ensuring both computational efficiency and robustness. We provide theoretical guarantees for monotonic convergence through a rigorous mathematical analysis, which is validated by extensive simulation results. For a comprehensive comparison, the proposed method is analyzed in relation to existing ILC techniques, such as those presented in [14], [48], and [51], with key distinctions highlighted in terms of convergence speed and error reduction rate.

The main contributions of the proposed optimal ILC in this paper are summarized as follows:

A. Direct Addressing of Continuous-Time Dynamics:

It works directly with continuous-time system dynamics, eliminating the need to discretize the system and apply the lifted-form method. As a result, the computational workload remains low and efficient.

B. Penalizing Error and Input Variation:

To ensure a fair and comprehensive evaluation, our method considers both tracking accuracy and control smoothness. Specifically, the weighting matrices in the cost function, which penalize error and input variation, are adjusted to balance the trade-off between aggressive learning and conservative input usage. This highlights the flexibility of the proposed optimal ILC in addressing performance specifications and physical constraints.

C. Independence from Derivative-Based Learning:

The proposed approach does not rely on derivative-based learning algorithms (D-type learning). While many previously presented continuous-time ILCs are based on D-type learning, which involves differentiating tracking errors or system outputs, such algorithms are highly sensitive to noise. This sensitivity becomes a significant issue in real-world applications, where measurements are inherently noisy.

D. Applicability to Time-Varying Systems:

Unlike the method in [42], which is limited to time-invariant systems, our formulation is designed to handle general time-varying systems.

E. Guaranteed Monotonic Convergence:

We prove that the proposed optimal ILC ensures the monotonic convergence of both the tracking error and the cost function, utilizing only elementary mathematical tools, such as the concepts and properties of the transition matrix for homogeneous linear differential equations. In contrast, the method used in [42] to prove convergence, even for the time-invariant case, involves many mathematical premises, including the theory of self-adjoint operators in Hilbert spaces and the properties of these operators.

F.Computational Efficiency and Ease of Implementation:

In the proposed method, two types of learning gains exist. The first type is a symmetric $n \times n$ matrix, where n represents the order of the controlled system. This matrix is obtained prior to the first iteration by solving a Riccati matrix differential equation, using the system coefficient matrices and the cost function weighting matrices. This learning gain is then used throughout all iterations. The second type of learning gain is a vector with n components, which needs to be determined by solving a linear vector differential equation, using the tracking error obtained from the previous iteration. This calculation is done before the start of each new iteration and when the system is being reset. Thus, all learning gains are determined either before the first iteration or during the intermediate phases between iterations. In other words, no algebraic or differential equations need to be solved during the iterations themselves. This characteristic significantly enhances computational efficiency and simplifies the practical implementation of the proposed method.

G. Superior Performance:

Extensive simulation results, including comparisons with the ILC methods in [14], [48], and [51], show that the proposed optimal ILC achieves faster convergence and higher tracking accuracy, while maintaining an acceptable computational cost.

The rest of this paper is organized as follows. Section II formulates the OPILC problem for continuous-time LTV systems and defines the mathematical assumptions. Section III derives the optimal control input using LQT theory and presents the resulting OPILC structure. Section IV rigorously proves the monotonic convergence of the tracking error and cost function. Section V provides detailed simulation results and performance comparisons with existing ILC methods. Section VI concludes the paper and discusses future extensions such as incorporating input/state constraints, uncertainty, or nonlinear dynamics.

II. Problem Setup

Consider the continuous-time MIMO LTV system that performs a repetitive operation:

$$\begin{cases} \dot{x}_i(t) = A(t)x_i(t) + B(t)u_i(t) + \eta_x(t) \\ y_i(t) = C(t)x_i(t) + \eta_y(t) \\ 0 \le t \le t_f i = 0,1,\dots \end{cases}, \tag{1}$$

where t denotes the time variable, i is the operation or iteration index, $x \in \mathbb{R}^n$, $u \in \mathbb{R}^p$, and $y \in \mathbb{R}^q$ are the system sate, input, and output vectors, respectively. $\eta_x \in \mathbb{R}^n$ and $\eta_y \in \mathbb{R}^q$ are the unknown but bounded deterministic disturbances, or the effects of the un-modeled dynamics of the system, which impact the state and output vectors of the system, respectively. t_f is a constant and specified, indicating the time duration of iterations. $A(t)$, $B(t)$, and $C(t)$ are the real-valued coefficients with appropriate dimensions. Moreover, the superimposed dot denotes the time derivative.

Consider system (1) and make the following reasonable assumptions:

A1) Every entry of $B(t)$ is a continuously differentiable function in the interval $t \in (0, t_f)$.

A2) All entries of $A(t)$, $B(t)$, $C(t)$, and $\dot{B}(t)$ are bounded and continuous functions in the interval $t \in [0, t_f]$.

A3) $C(t)B(t)$ has full row rank in the interval $t \in [0, t_f]$, that is:

$$rank(C(t)B(t)) = q \forall t \in [0, t_f]. \tag{2}$$

A4) The initial condition $x_i(0)$ remains constant in all iterations, i.e.

$$x_i(0) = x_0 i = 0,1,... \tag{3}$$

Remark 1. Assumptions A1 and A2 have been considered in order to guarantee the existence and uniqueness of the solution to the homogeneous differential equations that will be encountered later.

Remark 2. Assumption A3 is a standard assumption in ILC design, which guarantees the existence of learning gains [14, 29, 32, 34, 47]. It is clear that a necessary condition for assumption A3 to hold is that the number of system inputs must not be less than the number of outputs ($p \geq q$). Since the goal is to independently control all system outputs through appropriate selection of the inputs, the number of inputs must be greater than or at least equal to the number of outputs.

Remark 3. ILC systems usually have the same initial states x_0 for each working cycle [34, 42, 46, 47]. Thus, assumption A4 is not a restriction. This assumption is required for the perfect convergence of any type of ILC. Clearly, if the initial conditions of the system change randomly, it will be impossible to achieve perfect convergence because the controller encounters new conditions in each iteration, and therefore, learning and using past experiences to improve the performance becomes ineffective.

Let us define the problem of optimal iterative learning control (OPILC) as follows:

Consider system (1) and suppose that operates in a repetitive mode in the finite continuous time interval $t \in [0, t_f]$ and assumptions A1-A4 hold. A desired output trajectory $y_d(t) \in \mathbb{R}^q$ is given. The error between the system output and $y_d(t)$ at iteration i is defined as follows:

$$e_i(t) = y_d(t) - y_i(t) i = 0,1,2,... \tag{4}$$

Suppose that we are at the start of the current iteration i, and all the system signals in the previous iteration $i-1$, such as $u_{i-1}(t), x_{i-1}(t)$ and $e_{i-1}(t)$, are known and stored in the controller memory. Determine the system input in the current iteration i, that is $u_i(t)$, so that the system output approaches $y_d(t)$ by minimizing the following quadratic cost function:

$$J_i = \frac{1}{2}\{\|e_i(.)\|_Q^2 + \|v_i(.)\|_R^2\} i = 1,2,... \tag{5}$$

where

$$v_i(t) = u_i(t) - u_{i-1}(t), \tag{6}$$

and

$$\|e_i(.)\|_Q = \left[\int_0^{t_f} e_i^T(t)Q(t)e_i(t)dt\right]^{\frac{1}{2}},$$

$$\|v_i(.)\|_R = \left[\int_0^{t_f} v_i^T(t)R(t)v_i(t)dt\right]^{\frac{1}{2}}, \tag{7}$$

and T denotes the transpose, $Q(t) \in \mathbb{R}^{q \times q}$ and $R(t) \in \mathbb{R}^{p \times p}$ are the weight symmetric and positive definite matrices whose entries are the continuous functions of $t \in [0, t_f]$.

The physical interpretation of the given cost function (5) is that we want to bring the output of the system close to the desired output trajectory, and at the same time, the input of the system does not differ much from the previous iteration. The relative degrees of these two opposing goals is absorbed in matrices $Q(t)$ and $R(t)$.

III. Problem Solution

Let us define the vectors $z_i(t)$ and $w_i(t)$ as follows:

$$\begin{cases} z_i(t) = x_i(t) - x_{i-1}(t) \\ w_i(t) = y_i(t) - y_{i-1}(t). \\ 0 \leq t \leq t_f i = 1,2,... \end{cases} \tag{8}$$

Differentiating $z_i(t)$ with respect to t yields:

$$\begin{cases} \dot{z}_i(t) = \dot{x}_i(t) - \dot{x}_{i-1}(t) \\ w_i(t) = y_i(t) - y_{i-1}(t). \\ 0 \leq t \leq t_f i = 1,2,... \end{cases}$$

By substituting $\dot{x}_{i-1}(t)$, $\dot{x}_i(t)$, $y_{i-1}(t)$ and $y_i(t)$ from (1) into the above equations, one obtains:

$$\begin{cases} \dot{z}_i(t) = A(t)x_i(t) + B(t)u_i(t) + \eta_x(t) - A(t)x_{i-1}(t) \\ \qquad -B(t)u_{i-1}(t) - \eta_x(t) \\ \qquad = A(t)(x_i(t) - x_{i-1}(t)) + B(t)(u_i(t) - u_{i-1}(t)) \\ w_i(t) = y_i(t) - y_{i-1}(t) = C(t)x_i(t) + \eta_y(t) \\ \qquad -C(t)x_{i-1}(t) - \eta_y(t) \\ \qquad = C(t)(x_i(t) - x_{i-1}(t)) \end{cases}.$$

Inserting $x_i(t) - x_{i-1}(t)$ and $u_i(t) - u_{i-1}(t)$ from (8) and (6), respectively, into the above equations, yields:

$$\begin{cases} \dot{z}_i(t) = A(t)z_i(t) + B(t)v_i(t) \\ w_i(t) = C(t)z_i(t) \end{cases}.$$

$$0 \leq t \leq t_f, i = 1,2,... \tag{9}$$

On the other hand, substituting $y_i(t)$ from (8) into (4) gives:

$e_i(t) = y_d(t) - y_{i-1}(t) - w_i(t),$
that is:

$$e_i(t) = r_i(t) - w_i(t) i = 1,2,\dots \tag{10}$$

where

$$r_i(t) = e_{i-1}(t). \tag{11}$$

Substituting $e_i(t)$ from (10) into (5) yields:

$$J_i = \frac{1}{2}\int_0^{t_f} \{[r_i(t) - w_i(t)]^T Q(t)[r_i(t) - w_i(t)] + v_i^T(t)R(t)v_i(t)\} dt.$$
$$i = 1,2,\dots \tag{12}$$

Now, let us consider relation (9). This relation represents a hypothetical linear time-varying system, where $z_i(t)$, $v_i(t)$, and $w_i(t)$ are its state, input, and output vectors, respectively. $r_i(t)$ is the given desired output for this hypothetical system, and minimizing cost function (12) for this hypothetical system implies addressing a problem similar to the linear quadratic tracking (LQT) problem.

A review of the LQT problem and its solution procedure is provided in the Appendix. Based on this solution, and utilizing equation (A-3), the proposed OPILC problem is solved as follows:

$$v_i(t) = -R^{-1}(t)B^T(t)[K(t)z_i(t) - g_i(t)].$$
$$0 \le t \le t_f, i = 1,2,\dots \tag{13}$$

and the optimum value of cost function (12) is obtained as (see (A-8)):

$$J_i^* = \frac{1}{2} z_i^T(0)K(0)z_i(0) - z_i^T(0)g_i(0) + \phi_i(0).$$
$$i = 1,2,\dots \tag{14}$$

where $K(t) \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix that is the solution of the following Riccati matrix differential equation (see (A-4)):

$$\dot{K}(t) = -K(t)A(t) - A^T(t)K(t) + K(t)B(t)R^{-1}(t)B^T(t)K(t) - C^T(t)Q(t)C(t). \tag{15}$$

and $g_i(t) \in \mathbb{R}^n$ is a vector that is the solution of the following linear vector differential equation (see (A-6)):

$$\dot{g}_i(t) = -[A(t) - B(t)R^{-1}(t)B^T(t)K(t)]^T g_i(t) - C^T(t)Q(t)e_{i-1}(t).$$
$$i = 1,2,\dots \tag{16}$$

The third term in the right-hand side of (14), i.e., $\phi_i(0)$, is obtained from the backward integration of the following scalar differential equation (see (A-9)):

$$\dot{\phi}_i(t) = -\frac{1}{2} e_{i-1}^T(t)Q(t)e_{i-1}(t) + \frac{1}{2} g_i^T(t)B(t)R^{-1}(t)B^T(t)g_i(t).$$
$$i = 1,2,\dots \tag{17}$$

By comparing (12) with (A-2), it is observed that the matrix F in the cost function (12) is zero. Therefore, from (A-5), (A-7), and (A-10), it follows that the terminal conditions of the differential equations (15), (16), and (17) are all zero, that is

$$K(t_f) = 0, g_i(t_f) = 0, \phi_i(t_f) = 0. \tag{18}$$

Using (6) and (13), the input to system (1) at the i -th iteration is expressed in terms of its input at the previous iteration and other available information as follows

$$u_i(t) = u_{i-1}(t) - R^{-1}(t)B^T(t)\{K(t)[x_i(t) - x_{i-1}(t)] - g_i(t)\}.$$
$$0 \le t \le t_f, i = 1,2,\dots \tag{19}$$

Remark 4. Equation (15) is independent of the iterations. Therefore, it is solved only once prior to the first iteration, and the resulting gain $K(t)$ is used in all subsequent iterations. It should be noted that although matrix $K(t)$ has n^2 components, its symmetry implies that Riccati equation (15) represents a system of $\frac{n(n+1)}{2}$ first-order differential equations instead of n^2 equations.

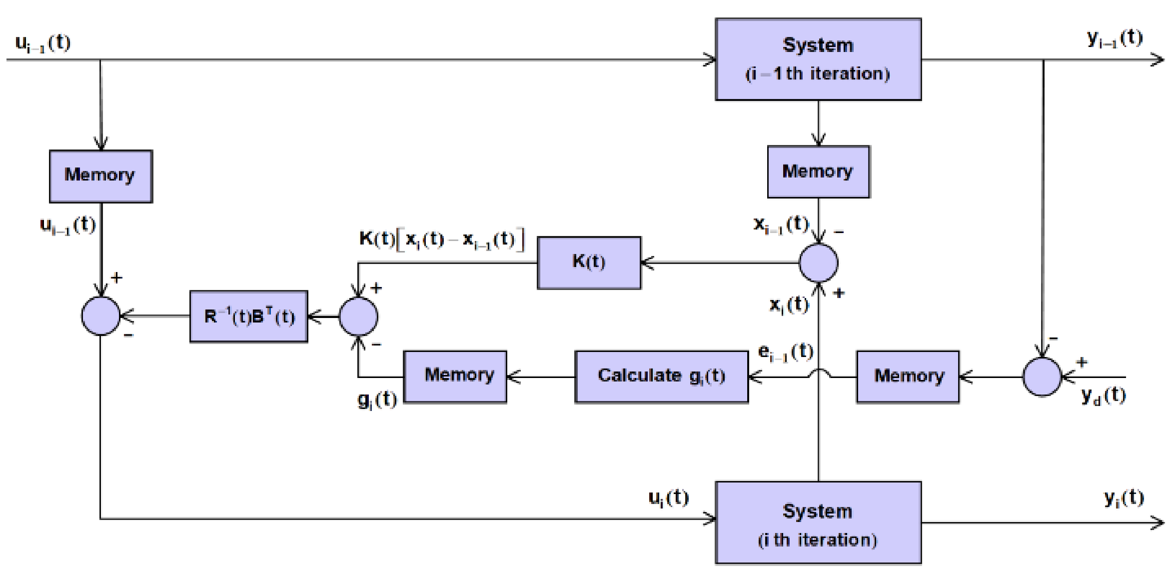


Fig. 1. The block diagram of the proposed OPILC.

Remark 5. The existence of $e_{i-1}(t)$ in equation (16) makes it an iteration-dependent equation. Therefore, it is necessary to re-compute $g_i(t)$ for each iteration. Given $e_{i-1}(t)$, one can compute $g_i(t)$ off-line in the time between iteration i and $i-1$, when the system is being reset. After this time, $e_{i-1}(t)$ can be removed from the memory.

Considering (19), the obtained OPILC can be illustrated as Fig. 1

As mentioned in Section I, in [42] Algorithm 9.5, one can find results similar to those of (15), (16), and (19) for the time-invariant case. The main differences between these two approaches can be summarized as follows:

1) The results given in Algorithm 9.5 of [42] are based on time-invariant self-adjoint operators acting on a Hilbert space. These assumptions ensure the applicability of classical spectral theory and operator analysis. However, such results are not directly extendable to time-varying operators, whose domains, boundedness, or self-adjointness properties may evolve over time. Consequently, the results of Algorithm 9.5 in [42] cannot be applied directly to time-varying systems. In contrast, our approach is fundamentally different from that of [42] and has been specifically developed from the outset to handle time-varying systems.

2) The method of proving convergence in [42], even for the time-invariant case, involves many mathematical premises, including the theory of self-adjoint operators in Hilbert spaces and the properties of these operators (see Theorems 9.1, 9.2, and 9.3 in [42]). However, our scheme to prove convergence is completely different from the method used in [42] and only relations (14), (16), and (17) and the concept of the transition matrix of the homogeneous linear differential equation are used in it (see Theorems 1, 2, and 3 in Section IV).

IV. Convergence Analysis

One of the desired features in the ILC design is to achieve convergence, and particularly monotonic convergence. Convergence implies that the tracking error converges to zero as the number of iterations increases. That is:

$$\lim_{i\to\infty} e_i(t) = 0 \forall t \in (0, t_f], \tag{20}$$

where $e_i(t)$ is the tracking error defined in (4).

The concept of monotonic convergence is that the norm of the tracking error in each iteration is smaller than the previous iteration. It means to have:

$$\begin{cases} \|e_i(.)\| < \|e_{i-1}(.)\| \text{if} \|e_{i-1}(.)\| \neq 0 \\ \|e_i(.)\| = \|e_{i-1}(.)\| = 0 \text{if} \|e_{i-1}(.)\| = 0 \end{cases} \forall i = 1,2,\dots \tag{21}$$

where $\|e_i(.)\|$ is a chosen norm for $e_i(t)$.

Monotonic convergence is a stronger condition than general convergence, as the latter can be deduced from the former. However, the reverse does not hold in the general case.

For the convergence analysis of the proposed OPILC, the following theorem is required.

Theorem 1. For any $i = 1,2,\dots$ the following three statements are equivalent:

$$e_{i-1}(t) = 0 \forall t \in (0, t_f]. \tag{22}$$

$$B^T(t) g_i(t) = 0 \forall t \in [0, t_f]. \tag{23}$$

$$v_i(t) = 0 \forall t \in [0, t_f]. \tag{24}$$

Proof. (22)⇔ (23): Suppose (22) holds. Then, substituting $e_{i-1}(t) = 0$ into (16) yields:

$$\dot{g}_i(t) = F_1(t) g_i(t), \tag{25}$$

where

$$F_1(t) = -[A(t) - B(t) R^{-1}(t) B^T(t) K(t)]^T. \tag{26}$$

Since all of the entries of $F_1(t)$ are bounded and continuous functions in the interval $t \in [0, t_f]$, for any given terminal condition $g_i(t_f)$, homogenous equation (25) has a unique solution in the interval $t \in [0, t_f]$, and this solution is obtained as follows:

$$g_i(t) = \Phi_1(t, t_f) g_i(t_f) \forall t \in [0, t_f], \tag{27}$$

where $\Phi_1(.,.)$is the state transition matrix of equation (25).

By applying the terminal condition $g_i(t_f) = 0$, as given in (18), to (27), we obtain

$$g_i(t) = 0 \forall t \in [0, t_f]. \tag{28}$$

Left-multiplying both sides of (28) by $B^T(t)$ yields relation (23).

Conversely, assume that equation (23) holds. Then, using (16), we obtain the following:

$$\dot{g}_i(t) = -A^T(t) g_i(t) - C^T(t) Q(t) e_{i-1}(t). \tag{29}$$

Left-multiplying both sides of (29) by $B^T(t)$ gives:

$$B^T(t)\dot{g}_i(t) = -B^T(t) A^T(t) g_i(t) - B^T(t) C^T(t) Q(t) e_{i-1}(t). \tag{30}$$

We have:

$$\frac{d}{dt}[\underbrace{B^T(t) g_i(t)}_{0}] = \dot{B}^T(t) g_i(t) + B^T(t)\dot{g}_i(t) = 0. \tag{31}$$

Thus,

$$B^T(t)\dot{g}_i(t) = -\dot{B}^T(t) g_i(t). \tag{32}$$

Combining (32) with (30) yields the following result:

$$[C(t)B(t)]^T Q(t) e_{i-1}(t) = \left[\dot{B}^T(t) - B^T(t) A^T(t)\right] g_i(t). \tag{33}$$

By left-multiplying both sides of the above equation by $C(t)B(t)$, the following relation is obtained:

$$C(t)B(t)[C(t)B(t)]^T Q(t) e_{i-1}(t) = C(t)B(t)\left[\dot{B}^T(t) - B^T(t) A^T(t)\right] g_i(t). \tag{34}$$

Since, according to Assumption A3, $C(t)B(t)$ has full row rank, it follows that $C(t)B(t)[C(t)B(t)]^T$ is invertible.

Multiplying both sides of equation (34) from the left by $Q^{-1}(t)(C(t)B(t)[C(t)B(t)]^T)^{-1}$ yields:

$$e_{i-1}(t) = Q^{-1}(t)(C(t)B(t)[C(t)B(t)]^T)^{-1} \times C(t)B(t)\big[\dot{B}^T(t) - B^T(t)A^T(t)\big]g_i(t). \quad (35)$$

Substituting $e_{i-1}(t)$ from (35) into (29) yields:

$$\dot{g}_i(t) = F_2(t)g_i(t). \quad (36)$$

where

$$F_2(t) = -\{A(t) + [\dot{B}(t) - A(t)B(t)](C(t)B(t))^T \times (C(t)B(t)[C(t)B(t)]^T)^{-1}C(t)\}^T. \quad (37)$$

According to Assumption A2, all entries of $A(t), B(t), C(t)$ and $\dot{B}(t)$ are continuous and bounded functions in the interval $t \in [0, t_f]$. It then follows from (37) that all entries of $F_2(t)$ are also continuous and bounded. As a result, for any given terminal condition $g_i(t_f)$, equation (36) has a unique solution in the interval $t \in [0, t_f]$, and this solution is obtained as follows:

$$g_i(t) = \Phi_2(t, t_f)g_i(t_f)\forall t \in [0, t_f], \quad (38)$$

where $\Phi_2(.,.)$is the state transition matrix of equation (36).

Applying the terminal condition $g_i(t_f) = 0$, as given in (18), to (38) gives (28). Then substituting $g_i(t)$ from (28) into (35) yields (22).

(23)⇔ (24): Suppose (23) holds. Then, substituting $B^T(t)g_i(t) = 0$ into (13) yields:

$$v_i(t) = -R^{-1}(t)B^T(t)K(t)z_i(t). \quad (39)$$

By substituting $v_i(t)$ from (39) into (9), the following homogeneous equation is obtained:

$$\dot{z}_i(t) = F_3(t)z_i(t), \quad (40)$$

where

$$F_3(t) = A(t) - B(t)R^{-1}(t)B^T(t)K(t). \quad (41)$$

All of the entries of $F_3(t)$ are continuous and bounded functions in the interval $t \in [0, t_f]$. Hence, for any given initial condition $z_i(0)$, equation (40) has a unique solution in the interval $t \in [0, t_f]$, as follows:

$$z_i(t) = \Phi_3(t, 0)z_i(0)\forall t \in [0, t_f], \quad (42)$$

where $\Phi_3(.,.)$is the state transition matrix of equation (40).

From (3) together with the definition of $z_i(t)$ in (8), it follows that:

$$z_i(0) = 0 i = 1,2,... \quad (43)$$

Substituting $z_i(0)$ from (43) into (42) yields the following result:

$$z_i(t) = 0\forall t \in [0, t_f]. \quad (44)$$

By substituting $z_i(t)$ from (44) into (39), one can obtain (24).

Conversely, if (24) holds, then from (9) and (43), we can obtain (44). Finally, substituting $v_i(t)$ and $z_i(t)$ from (24) and (44), respectively, into (13) yields (23).

Now, we are in a position that we can analyze the convergence properties of the proposed OPILC.

Theorem 2. The proposed OPILC guarantees monotonic convergence in the sense of norm $\|e_i(.)\|_Q$ defined in (7).

Proof. Substituting $z_i(0)$ from (43) into (14) yields that the optimum value of cost function (12) is as follows:

$$J_i^* = \phi_i(0) i = 1,2, ... \quad (45)$$

Integrating both sides of equation (17) over the interval $[0, t_f]$ gives:

$$\phi_i(t_f) - \phi_i(0) = -\frac{1}{2}\|e_{i-1}(.)\|_Q^2 + \frac{1}{2}\int_0^{t_f} g_i^T(t)B(t)R^{-1}(t)B^T(t)g_i(t)dt.$$

By inserting $\phi_i(t_f)$ and $\phi_i(0)$ from (18) and (45), respectively, into the above equations, the following expression is obtained:

$$J_i^* = \frac{1}{2}\|e_{i-1}(.)\|_Q^2 - \frac{1}{2}\int_0^{t_f} g_i^T(t)B(t)R^{-1}(t)B^T(t)g_i(t)dt. \quad (46)$$

On the other hand, based on (7), (10) and (12), J_i^* has the following form:

$$J_i^* = \frac{1}{2}\{\|e_i(.)\|_Q^2 + \|v_i(.)\|_R^2\}. \quad (47)$$

From (46) and (47) one gets:

$$\frac{1}{2}\|e_i(.)\|_Q^2 = \frac{1}{2}\|e_{i-1}(.)\|_Q^2 - \frac{1}{2}\int_0^{t_f} \{v_i^T(t)R(t)v_i(t) + g_i^T(t)B(t)R^{-1}(t)B^T(t)g_i(t)\}dt. \quad (48)$$

Since $R(t)$ and, consequently, $R^{-1}(t)$ are symmetric positive definite matrices for all $t \in [0, t_f]$, equation (48) yields the following result:

$$\|e_i(.)\|_Q \leq \|e_{i-1}(.)\|_Q i = 1,2,... \quad (49)$$

and (49) holds as an equality if and only if one has:

$$v_i(t) = 0\forall t \in [0, t_f], \quad (50)$$

and:

$$B^T(t)g_i(t) = 0\forall t \in [0, t_f]. \quad (51)$$

Taking Theorem 1 into account, each of the relations (50) and (51) is equivalent to (22).

Therefore, the proposed OPILC is convergent, and this convergence is monotonic in the sense of norm $\|e_i(.)\|_Q$.

Theorem 3. The proposed OPILC guarantees monotonic convergence in the sense of cost function (5).

Proof. By adding and subtracting $\frac{1}{2}\|v_{i-1}(.)\|_R^2$ to the right-hand side of (46), one obtains:

$$J_i^* = J_{i-1}^* - \frac{1}{2}\int_0^{t_f} \{v_{i-1}^T(t)R(t)v_{i-1}(t) + g_i^T(t)B(t)R^{-1}(t)B^T(t)g_i(t)\}dt. \quad i = 2,3,... \quad (52)$$

Since $R(t)$, and therefore, $R^{-1}(t)$ are symmetric positive definite for all $t \in [0, t_f]$, (52) yields:

$$J_i^* \leq J_{i-1}^* i = 2,3,\ldots \tag{53}$$

and (53) holds as an equality if and only if we have:

$$v_{i-1}(t) = 0, B^T(t) g_i(t) = 0.$$
$$\forall t \in [0, t_f], i = 2,3,\ldots \tag{54}$$

According to Theorem 1, (54) is equivalent to:

$$e_{i-2}(t) = e_{i-1}(t) = 0 \forall t \in (0, t_f], i = 2,3,\ldots \tag{55}$$

Consequently, the proposed OPILC is convergent, and this convergence is monotonic in the sense of cost function (5).

V. Illustrative Examples and Comparison with Other ILCs

In this section, four simulation examples are provided to demonstrate the applicability and performance of the proposed OPILC.

All simulations were performed on a computer with a 12th Gen Intel® Core™ i3-12100 CPU @ 3.30 GHz, 16.0 GB of RAM, running Windows 10 Pro (version 22H2). MATLAB R2016b was used for implementation and numerical computations.

Example 1. Consider a second-order repetitive system in the form of (1), where the coefficient matrices, time duration of iterations, and disturbances $\eta_x(t)$, $\eta_y(t)$ are given as follows:

$$A(t) = \begin{bmatrix} -t & 1 \\ 2\,sin(\pi t) & -\text{Ln}(t+1) \end{bmatrix}, B(t) = \begin{bmatrix} e^t \\ sinh(t) \end{bmatrix},$$
$$C(t) = [0.5 \quad t^2], t_f = 1,$$
$$\eta_x(t) = \begin{bmatrix} e^{sin(4\pi t)} \\ t\,cos(2\pi t) \end{bmatrix}, \eta_y(t) = 0.05(t - 0.5).$$

To perform the simulation, the initial condition of the system and its input at the initial iteration $i = 0$ are set to be zero. That is:

$$x_i(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{for} i = 0,1,\ldots, u_0(t) = 0 \text{for} 0 \leq t \leq 1.$$

The desired output trajectory $y_d(t)$ is chosen as:

$$y_d(t) = sin(60t - 92t^2 + 48t^3 - 8t^4) \text{ for} 0 \leq t \leq 1.$$

In order to compare the performance of the method proposed in this paper with the approach presented in [48], the simulation of this example is conducted in two cases. In these cases, the method proposed in this paper and the approach presented in [48] are used, respectively.

Case 1. Using the OPILC proposed in this paper

The symmetric positive definite matrices $Q(t)$ and $R(t)$ in cost function (5), are chosen as follows:
$Q(t) = 10^4, R(t) = 1.$

The time response of the system output, along with the desired output trajectory, is depicted in Fig. 2. This figure indicates that the system output successfully converges to the desired output trajectory by increasing the number of iterations.

The values of $\frac{1}{2}\|e_i(.)\|_Q^2$ and cost function J_i versus the number of iterations are given in Fig. 3 (in the log scale). This figure demonstrates that the convergence is monotonic with respect to both $\|e_i(.)\|_Q$ and J_i, which are consistent with Theorems 2 and 3.

A key aspect of iterative learning control is how the control input evolves over successive trials. Figure 4 illustrates the progression of the input profiles across iterations. It can be observed that the input converges to a particular signal that was not known a priori and only emerged as a result of the learning process.

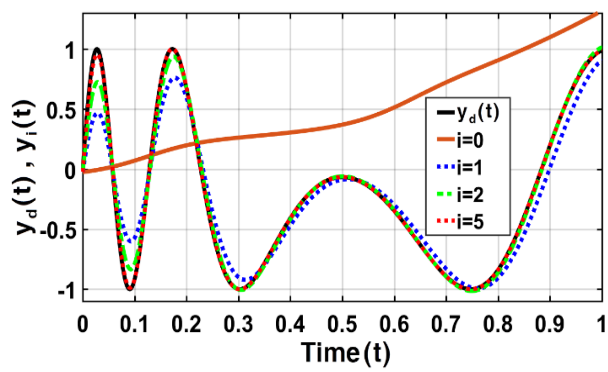


Fig. 2. The output trajectories $y_d(t)$ and $y_i(t) (i = 0,1,2,5)$ for Example 1 (Case 1).

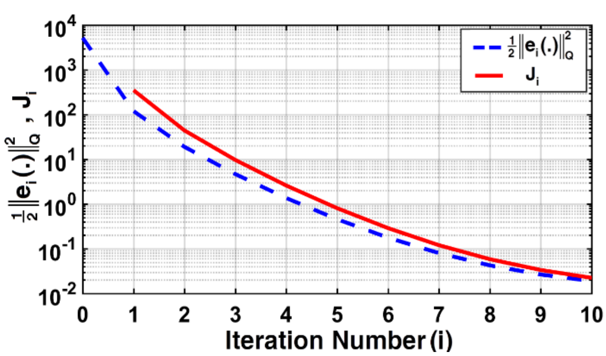


Fig. 3. The monotonic convergence of $\|e_i(.)\|_Q$ and J_i with respect to i for Example 1 (Case 1).

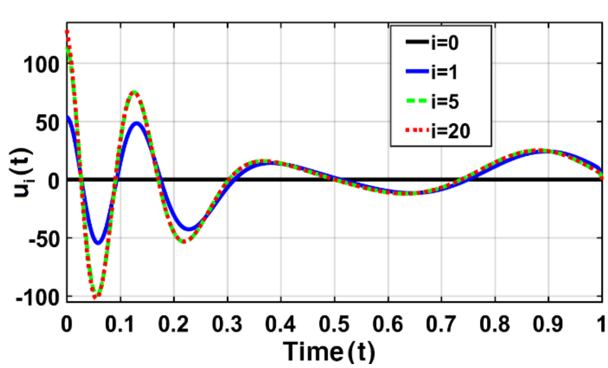


Fig. 4. The system input at iterations $i = 0,1,5,20$ for Example 1 (Case 1).

Case 2. Using the ILC presented in [48]

First, we provide a brief review the method given in [48]. In [48], the system under control is discrete-time, MIMO, and LTV, which is represented by:

$$\begin{cases} x_i(k+1) = A_D(k)x_i(k) + B_D(k)u_i(k) \\ y_i(k) = C_D(k)x_i(k) \\ k = 0,1,\ldots,N i = 0,1,\ldots \end{cases} . \quad (56)$$

and the ILC problem is defined as follows:
Given System (56) with boundary conditions $x_i(0) = x_0$ $(i \geq 0)$ and $u_0(k)(0 \leq k \leq N-1)$, and reference output trajectory $y_d(k)(1 \leq k \leq N)$, find an appropriate control input $u_i(k)$ $(0 \leq k \leq N-1$ and $i \geq 1)$ such that the system output follows the reference trajectory.

To solve this problem, the following ILC is presented by [48]:

$$\begin{matrix} u_{i+1}(k) = u_i(k) + F(k+1)e_i(k+1) \\ k = 0,1,\ldots,N-1; i = 0,1,\ldots \end{matrix} . \quad (57)$$

where $F(k+1)$ is a time-varying learning gain matrix with appropriate dimensions, and $e_i(k+1)$ shows the tracking error, which is:
$e_i(k+1) = y_d(k+1) - y_i(k+1)$.

Using two-dimensional (2-D) system theory, it has been shown in [48] that if the gain matrix $F(k)$ is chosen such that the following relation holds:

$$\rho\{I - C_D(k)B_D(k-1)F(k)\} < 1, \\ k = 1,2,\ldots,N. \quad (58)$$

where I is the identity matrix and $\rho\{.\}$ represents the spectral radius of the matrix.

Then ILC (57) guarantees:

$$\lim_{i\to\infty} e_i(k) = 0 \text{for} k = 1,2,\ldots,N. \quad (59)$$

Moreover, it has been shown that there exists a gain matrix $F(k)$ that yields (58) only if matrix $C_D(k)B_D(k-1)$ has full-row rank, and in this case, condition (58) can be achieved by calculating $F(k)$ as follows [48]:

$$F(k) = (C_D(k)B_D(k-1))^T\{C_D(k)B_D(k-1) \\ \times (C_D(k)B_D(k-1))^T\}^{-1}(I - \Phi(k)), \\ \text{for} k = 1,2,\ldots,N. \quad (60)$$

where $\Phi(k)$ is an arbitrary matrix with appropriate dimensions that satisfies the condition $\rho\{\Phi(k)\} < 1$ for $k = 1,2,\ldots,N$.

When system (56) is SISO, then $C_D(k)B_D(k-1)$, $\Phi(k)$, and $F(k)$ are scalars, and consequently, formula (60) simplifies to:

$$F(k) = \frac{1-\Phi(k)}{C_D(k)B_D(k-1)} \text{for} k = 1,2,\ldots,N. \quad (61)$$

and condition $\rho\{\Phi(k)\} < 1$ is reduced to $|\Phi(k)| < 1$ for $k = 1,2,\ldots,N$.

It should be noted that the above ILC is given for discrete-time systems, whereas this study focuses on continuous-time system. For this reason, first, the system given in Example 1 is discretized by choosing the sampling time $\Delta t = 0.001$. Therefore, we have:

$$N = \frac{t_f}{\Delta t} = 10^3.$$

In addition, since the system is SISO, (61) is used to calculate the learning gain $F(k)$. To have the maximum possible convergence rate, $\Phi(k)$ is chosen as $\Phi(k) = 0.9975$ (for $k = 1,2,\ldots,10^3$) through trial and error.

The trajectories obtained for the system output from the simulation in some iterations are shown in Fig. 5. Although according to this figure, the output of the system is closer to the desired trajectory with the increase in the number of iterations, by comparing this figure with Fig. 2, it is clearly inferred that the convergence rate in Fig. 5 is much slower than in Fig. 2. To evaluate this more precisely, let us use the following total square error of tracking:

$$EE(i) = \sum_{k=0}^{N} e_i^2(k) = \sum_{k=0}^{1000} e_i^2(k). \quad (62)$$

Figure 6 shows the total squared error of tracking (in the log scale) for both cases 1 and 2. It can be observed that the convergence rate of the OPILC proposed in this paper is much faster than that in the method presented in [48].

To enable a more comprehensive comparison of the performance of the OPILC proposed in this paper with other ILCs, we employ quantitative performance metrics including error reduction rate, convergence speed, and computational complexity. For this purpose, let $I_{n\%}$ (for $n = 50,10,1$) denote the minimum number of iterations required for the error to become less than or equal to n percent of its initial value (i.e., the error at iteration zero), as a measure of the error reduction rate. Similarly, let $I_{10^{-3}}$ represent the minimum number of iterations required for the absolute value of error to reach or fall below 10^{-3}, serving as a measure of convergence speed. To assess computational complexity, T_1 is used to denote the time required to execute a single iteration on the hardware platform specified at the beginning of this section.

The results obtained from the comparison of Cases 1 and 2 for Example 1 are presented in Table I. It should be noted that the term "error" in this table refers to the total square error as defined in equation (62). According to this table, the error reduction rate and convergence speed of the proposed method are significantly higher than those of the ILC method presented in [48]. Specifically, the error is reduced by 90% after just one iteration using the proposed method, whereas the ILC in [48] requires 710 iterations to achieve the same reduction. Moreover, the convergence speed of the proposed method is over 75 times faster than that of the ILC in [48]. Although the proposed scheme requires approximately 3.8225 times more computational time per iteration compared to the ILC method presented in [48], as will be clarified in Remark 6, this increase does not introduce any practical impediments.

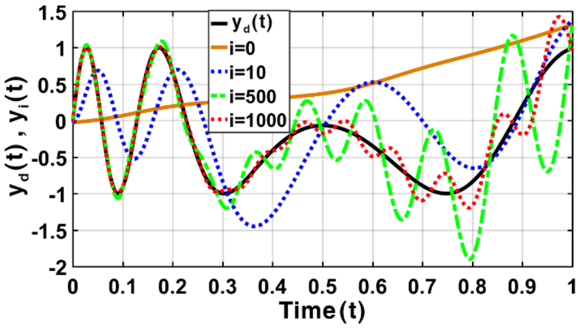


Fig. 5. The output trajectories $y_d(t)$ and $y_i(t)$ $(i = 0,10,500,1000)$ for Example 1 (Case 2).

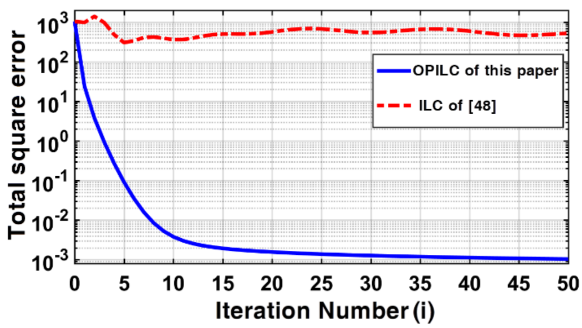


Fig. 6. The total square errors of tracking with respect to i for Example 1 (comparison of Cases 1 and 2).

TABLE I. Comparison of Cases 1 and 2 for Example 1

Method		ILC of [48]	OPILC proposed in this paper
Error Reduction Rate	$I_{50\%}$	285	1
	$I_{10\%}$	710	1
	$I_{1\%}$	1220	2
Convergence Speed	$I_{10^{-3}}$	4150	55
Computational Complexity	T_1 (sec.)	0.012535	0.047915

Example 2. The rotation dynamics of a rigid body in space can be described by Euler equations in the principal coordinates as follows [49, 50]:

$$\begin{cases} J_x\dot{\omega}_x = (J_y - J_z)\omega_y\omega_z + u_x \\ J_y\dot{\omega}_y = (J_z - J_x)\omega_z\omega_x + u_y. \\ J_z\dot{\omega}_z = (J_x - J_y)\omega_x\omega_y + u_z \end{cases} \quad (63)$$

where ω_x, ω_y, and ω_z are the components of the angular velocity vector ω along the principal axes; u_x, u_y, and u_z are the torque inputs applied about the principal axes; and J_x, J_y, and J_z are the principal moments of inertia.

Assuming that the body is axial-symmetric, i.e., $J_x = J_y = J$, rotational dynamics along the x and y axes are considered, under the assumption that $\omega_z(t)$ is a known function of time. This yields an LTV system of the form

$$\begin{cases} \dot{x} = \underbrace{\begin{bmatrix} 0 & \Omega(t) \\ -\Omega(t) & 0 \end{bmatrix}}_{A(t)} x + \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{B(t)} u \\ y = \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{C(t)} x \end{cases}. \quad (64)$$

where

$$x = \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix}, y = \begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix}, u = \frac{1}{J}\begin{bmatrix} u_x \\ u_y \end{bmatrix}, \Omega(t) = \frac{(J - J_z)}{J}\omega_z(t). \quad (65)$$

Let us consider model (64). This is a two input-two output (TITO) system. It is desired to control this system iteratively in a given time interval $[0, t_f]$.

For simulations, we choose $t_f = 2$ and $\Omega(t) = t\sin(2\pi t)$. Moreover, the desired output trajectories are selected as follows:

$$y_d^{(1)}(t) = \tanh[20\cos(t)]\sin(t^2),$$
$$y_d^{(2)}(t) = 10te^{-2t}\cos(\pi\sqrt{t}). \quad (66)$$

The initial conditions and the initial input of the system (the input of the system in iteration $i = 0$) are all chosen to be zero.

In order to compare the performance of the OPILC proposed in this paper with the methods presented in [51] and [14], the simulation of this example is conducted in three cases. In these cases, the procedure proposed in this paper and the ILCs presented in [51] and [14] are used, respectively.

Case 1. Using the OPILC proposed in this paper

The symmetric positive definite matrices $Q(t)$ and $R(t)$ in cost function (5) are chosen as follows:

$Q(t) = 10^4 I, R(t) = I.$

The trajectories obtained for the system outputs from the simulation at some iterations, along with the desired outputs trajectories, are shown in Figs. 7a and 7b. These figures indicate that by increasing the iteration number, the outputs quickly converge to the given desired outputs trajectories, such that at iteration $i = 1$, the system outputs exactly follow the desired outputs trajectories.

Furthermore, Fig. 8 shows $\frac{1}{2}\|e_i(.)\|_Q^2$ and J_i (in the log scale) versus the number of iterations. According to this figure, both $\|e_i(.)\|_Q$ and J_i converge to zero with a monotonic convergence rate, which confirms Theorems 2 and 3.

An essential feature of ILC lies in the evolution of the control input over repeated trials. As shown in Figs. 9a and 9b, the system inputs profiles gradually adjust through iterations and eventually converge to two consistent patterns that are not predefined but arise from the learning dynamics.

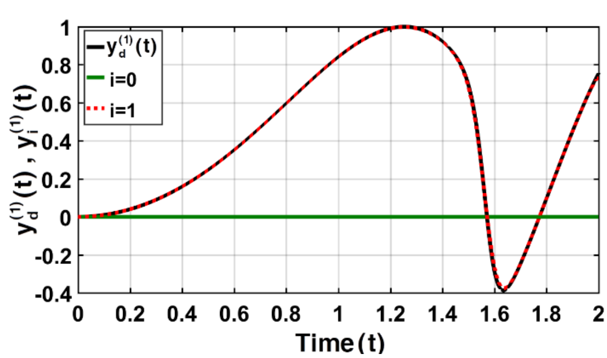


Fig. 7a. The output trajectories $y_d^{(1)}(t)$ and $y_i^{(1)}(t) (i = 0,1)$ for Example 2 (Case 1).

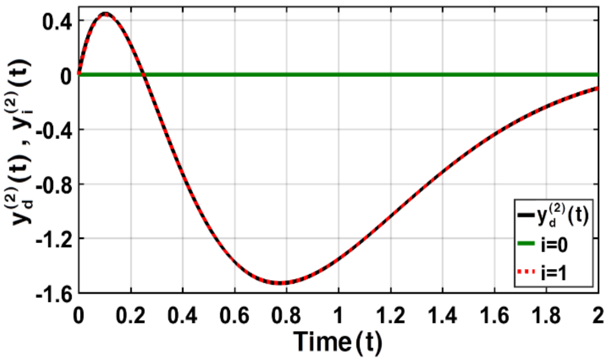


Fig. 7b. The output trajectories $y_d^{(2)}(t)$ and $y_i^{(2)}(t)(i = 0,1)$ for Example 2 (Case 1).

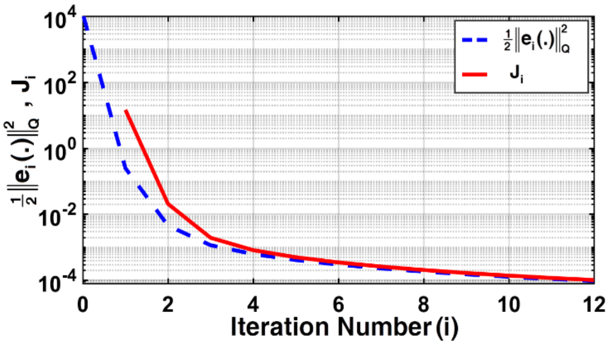


Fig. 8. The monotonic convergence of $\|e_i(.)\|_Q$ and J_i with respect to i for Example 2 (Case 1).

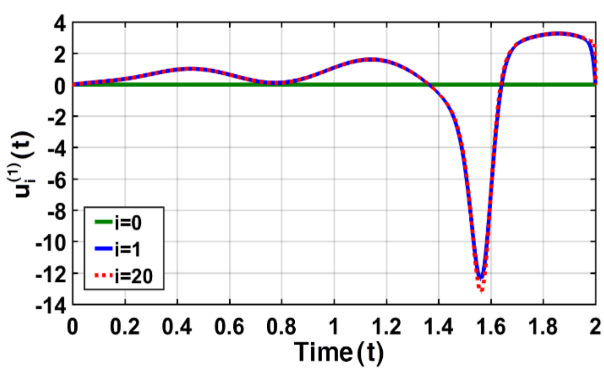


Fig. 9a. The system input $u_i^{(1)}(t)$ at iterations $i = 0,1,20$ for Example 2 (Case 1).

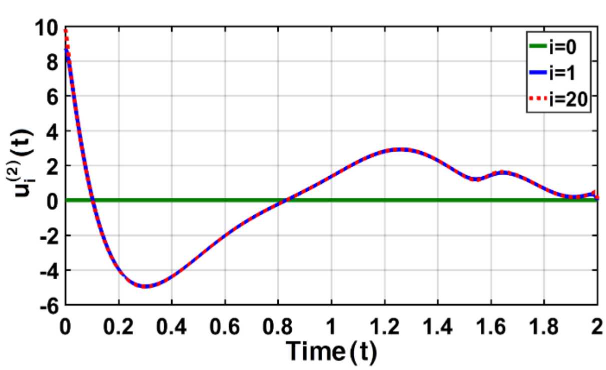


Fig. 9b. The system input $u_i^{(2)}(t)$ at iterations $i = 0,1,20$ for Example 2 (Case 1).

Case 2. Using the ILC presented in [51]

The iterative learning tracking control problem for a class of MIMO nonlinear continuous-time LTV systems is studied in [51]. The system at the i-th iteration is described as equation (67):

$$\begin{cases} \dot{x}_i(t) = f(t, x_i(t)) + B(t)u_i(t) \\ y_i(t) = C(t)x_i(t) \\ x_i(0) = x_0 \\ \quad 0 \le t \le t_f i = 0,1,\dots \end{cases} . \tag{67}$$

where $B(t) \in \mathbb{R}^{n\times p}$ and $C(t) \in \mathbb{R}^{q\times n}$ are time-varying coefficients, and $f(.,.)$ is a nonlinear function that is piecewise continuous in $t \in [0, t_f]$ and globally Lipschitz in $x \in \mathbb{R}^n$, i.e.,

$$\exists 0 < \gamma < \infty: \|f(t, x_1) - f(t, x_2)\| \le \gamma\|x_1 - x_2\|, \\ \forall x_1, x_2 \in \mathbb{R}^n, \forall t \in [0, t_f]. \tag{68}$$

γ is the Lipschitz constant.

The ILC presented in [51] for system (67) is a PD-type one and has the following form:

$$u_{i+1}(t) = u_i(t) + K_P(t)e_{i+1}(t) + \lambda(t)K_D(t)\dot{e}_{i+1}(t). \tag{69}$$

where $e_{i+1}(t)$ is the tracking error at iteration $i + 1$, which is defined as (4); $K_P(t)$ and $K_D(t)$ are the proportional and derivative gains, respectively, with appropriate dimensions; and $\lambda(t) \ge 1$ is an exponentially increasing function with respect to t such as e^{Mt} $(M > 0)$.

As proved in [51], if a unique control input $u_d(t)$ exists for system (67) such that it generates the desired output trajectory $y_d(t)$, i.e., $u_d(t)$ and $y_d(t)$ satisfy the following equations:

$$\begin{cases} \dot{x}_d(t) = f(t, x_d(t)) + B(t)u_d(t) \\ y_d(t) = C(t)x_d(t)) \\ x_d(0) = x_0 \\ \quad 0 \le t \le t_f \end{cases} . \tag{70}$$

where $x_d(t)$ is the state corresponding to $y_d(t)$.

Then for any initial input $u_0(t)(0 \le t \le t_f)$ and initial state $x_i(0) = x_0(i = 0,1,\dots)$, we have:
$\lim_{i\to\infty} e_i(t) = 0 \forall t \in [0, t_f]$.

If the following condition holds:

$$\rho\{[I + \lambda(t)K_D(t)C(t)B(t)]^{-1}\} < 1 \forall t \in [0, t_f]. \tag{71}$$

where I denotes the identity matrix and $\rho\{.\}$ is the spectral radius of the matrix.

It is clear that if system (67) is linear similar to (1), that is if $f(.,.)$ is as follows:

$$f(t, x_i(t)) = A(t)x_i(t), A(t) \in \mathbb{R}^{n\times n}. \tag{72}$$

and all entries of $A(t)$ are the piecewise continuous and bonded functions in the interval $t \in [0, t_f]$, then condition (68) is satisfied. Consequently, let us try to use ILC (69) for linear system (64).

For any given differentiable desired output trajectory $y_d(t) = \begin{bmatrix} y_d^{(1)}(t) & y_d^{(2)}(t) \end{bmatrix}^T$, such as the one given in (66), a unique control input $u_d(t) = \dot{y}_d(t) - A(t)y_d(t)$ is obtained for system (64) such that it generates $y_d(t)$. Hence, ILC (69) is applicable to system (64).

For system (64), the following gains of ILC (69) satisfy the convergence condition (71):

$$K_P(t) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}, K_D(t) = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}, \lambda(t) = e^{0.8t}.$$

The desired output trajectories and the obtained trajectories for the system outputs are given in Fig. 10a and 10b for some iterations. Although these figures show that as the number of iterations increases, the system outputs tend towards the desired output trajectories, comparing these figures with Figs. 7a and 7b indicates that the convergence rate in Figs. 7a and 7b is much faster than in Figs. 10a and 10b. For a more accurate comparison, the following performance index, i.e., the 2-norm of the tracking error, is considered:

$$\|e_i(.)\|_2 = \left[\int_0^{t_f} e_i^T(t) e_i(t) dt\right]^{\frac{1}{2}}. \tag{73}$$

The obtained results for $\|e_i(.)\|_2$ in both cases 1 and 2 are illustrated in Fig. 11. This figure clearly demonstrates that the tracking performance of the proposed OPILC is superior to that of the method in [51].

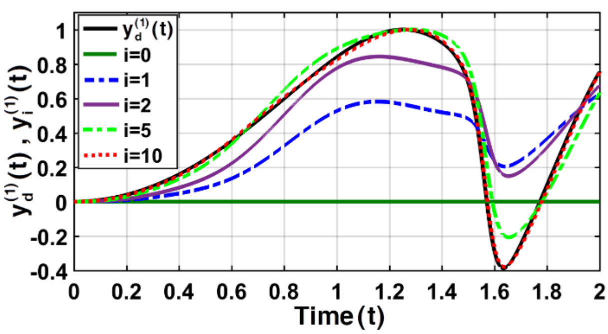


Fig. 10a. The output trajectories $y_d^{(1)}(t)$ and $y_i^{(1)}(t)$ $(i = 0,1,2,5,10)$ for Example 2 (Case 2).

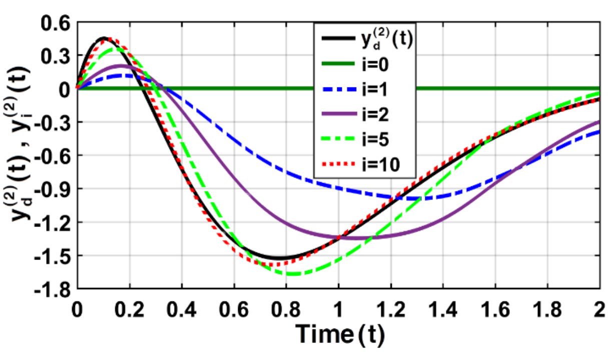


Fig. 10b. The output trajectories $y_d^{(2)}(t)$ and $y_i^{(2)}(t)$ $(i = 0,1,2,5,10)$ for Example 2 (Case 2).

Quantitative performance metrics for comparing Cases 1 and 2 in Example 2 are provided in Table II. In this table, the term "error" refers to the 2-norm of the tracking error as defined in equation (73). As observed, the proposed method exhibits a substantially higher error reduction rate and faster convergence speed compared to the ILC method introduced

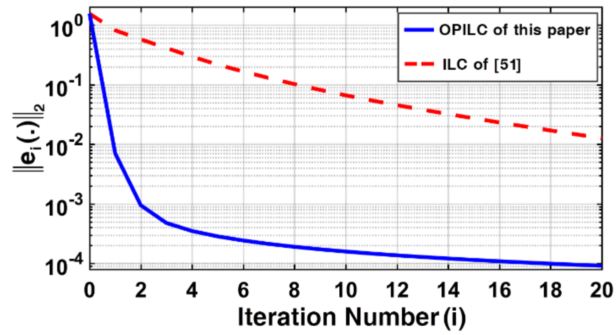


Fig. 11. The 2-norm of the tracking error $e_i(t)$ with respect to i for Example 2 (comparison of Cases 1 and 2).

in [51]. Remarkably, a single iteration of the proposed method results in a 99% reduction in error, whereas the ILC approach in [51] requires 19 iterations to achieve a similar level of improvement. Furthermore, the convergence speed of the proposed scheme is approximately 19 times greater than that of the ILC in [51]. Despite the fact that the per-iteration execution time is nearly 4.1 times greater than that of the ILC approach in [51], as elaborated in Remark 6, such a rise in computational effort does not translate into any significant practical constraints.

Case 3. Using the ILC presented in [14]

Recently, an ILC approach has been presented in [14] for linear discrete-time, MIMO, and LTV systems described by (56), where the system has the same number of inputs and outputs and different initial time points. However, in this paper, we focus on systems that have fixed initial time points. For such systems, the structure of the ILC given in [14] is in the form of relation (57), but the method of determining the gain matrix $F(k)$ is different from equation (58). In [14], it has been shown that if $F(k)$ is chosen to make:

TABLE II. Comparison of Cases 1 and 2 for Example 2

Method		ILC of [51]	OPILC proposed in this paper
Error Reduction Rate	$I_{50\%}$	2	1
	$I_{10\%}$	7	1
	$I_{1\%}$	19	1
Convergence Speed	$I_{10^{-3}}$	38	2
Computational Complexity	T_1 (sec.)	0.078188	0.319841

$$\sup_{k \in \{1,2,\dots,N\}} \|I - F(k) C_D(k) B_D(k-1)\| < 1, \tag{74}$$

then (59) holds.

Let us apply the above ILC to the system explained in Example 2. Since this ILC is given for discrete-time systems and in Example 2, the system is continuous-time, the given system is first discretized by choosing the sampling time $\Delta t = 2 \times 10^{-4}$. Therefore, we have:

$$N = \frac{t_f}{\Delta t} = 10^4.$$

The learning gain $F(k)$ is chosen as follows:

$$F(k) = \begin{bmatrix} 1 & -0.5 \\ -0.25 & 2 \end{bmatrix} \text{for} k = 1,2,\dots, 10^4.$$

It is easy to show that this learning gain satisfies convergent condition (74).

As a result, Figs. 12a and 12b show the actual tracking performance of the system outputs $y_i^{(k)}(t)$ to the desired trajectories $y_d^{(k)}(t)$ at the iterations $i = 0,10,100,500$ for $k = 1,2$, respectively. As seen in these figures, with the increase of the iteration number, the system outputs converge to the desired trajectories. However, comparing these figures with Figs. 7a and 7b demonstrates that the convergence rate in Figs. 7a and 7b is much faster than in Figs. 12a and 12b. Let us use the following index to evaluate this fact more precisely:

$$EE(i) = \sum_{k=0}^{N} e_i^T(k)\, e_i(k) = \sum_{k=0}^{10^4} e_i^T(k) e_i(k). \qquad (75)$$

This index is an extension of (62) to the MIMO case. Hence, it can be stated that it is the total square error of tracking.

Figure 13 shows the profiles of $EE(i)$ versus iterations for both cases 1 and 3. It is observed that the system with the OPILC proposed in this paper has a much better convergence performance in speed and precision than the one with the ILC presented in [14].

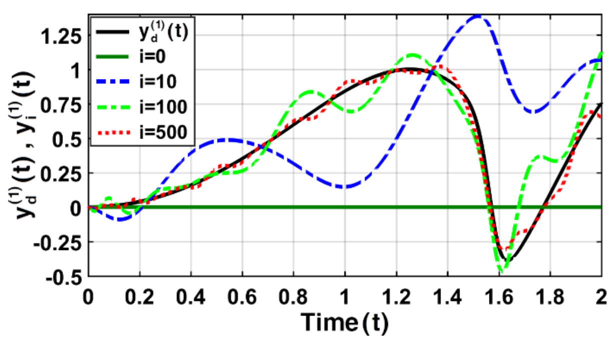


Fig. 12a. The output trajectories $y_d^{(1)}(t)$ and $y_i^{(1)}(t)$ $(i = 0,10,100,500)$ for Example 2 (Case 3).

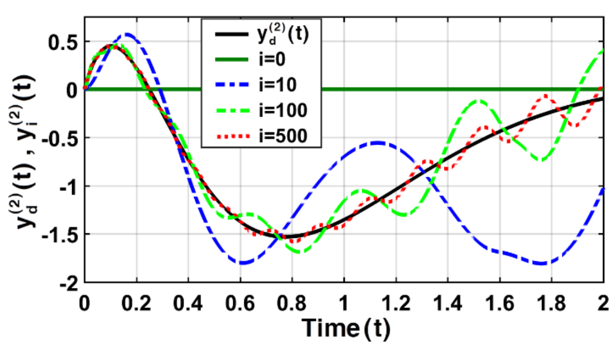


Fig. 12b. The output trajectories $y_d^{(2)}(t)$ and $y_i^{(2)}(t)$ $(i = 0,10,100,500)$ for Example 2 (Case 3).

Table III reports the quantitative performance metrics for comparing Cases 1 and 3 in Example 2. In this table, the error is defined as the total square error of tracking, as formulated in equation (75). The results clearly demonstrate that the proposed method offers a dramatic improvement over the ILC approach introduced in [14], both in terms of error reduction rate and convergence speed. Specifically, the proposed algorithm achieves a 99% reduction in tracking error after only a single iteration, while the ILC method in [14] requires 390 iterations to reach a comparable level of accuracy. Moreover, the convergence speed of the proposed strategy is approximately 4062 times faster than that of the referenced ILC. Although the iterative execution time is approximately 1.8 times longer relative to the ILC framework described in [14], as will be discussed in greater detail in Remark 6, this increased computational burden does not undermine the practicality or operational efficiency of the proposed method.

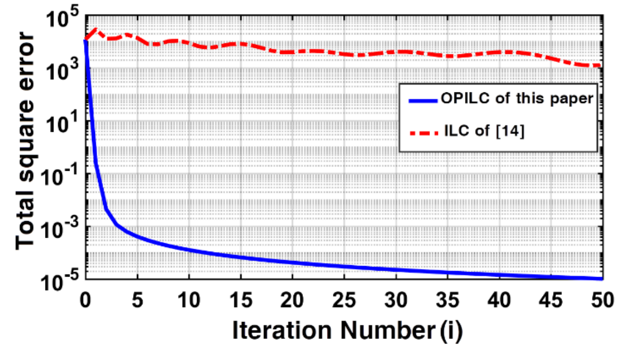


Fig. 13. The total square errors of tracking with respect to i for Example 2 (comparison of Cases 1 and 3)

Example 3. This example, which considers a practical MIMO process, is taken from [52]. The schematic diagram of a quadruple-tank process is shown in Fig. 14. In this process, the target is to control the level in the lower two tanks with two pumps. The process inputs are $v_{\text{in}1}$ and $v_{\text{in}2}$ (input voltages to the pumps) and the outputs are $v_{\text{out}1}$ and $v_{\text{out}2}$ (voltages from lower two tanks level measurement devices).

The voltage applied to Pump j is $v_{\text{in}j}$ and the corresponding flow is $k_j v_{\text{in}j}$ $(j = 1,2)$. The parameters $\gamma_1,\ \gamma_2 \in (0,1)$ are determined from how the valves are set prior to an experiment. The flow to Tank 1 is $\gamma_1 k_1 v_{\text{in}1}$ and the flow to Tank 4 is $(1-\gamma_1)k_1 v_{\text{in}1}$ and similarly for Tank 2 and Tank 3.

The linearized model of this process around a stationary point is obtained as follows [52]:

$$\begin{cases} \dot{x} = \underbrace{\begin{bmatrix} -\frac{1}{T_1} & 0 & \frac{A_3}{A_1T_3} & 0 \\ 0 & -\frac{1}{T_2} & 0 & \frac{A_4}{A_2T_4} \\ 0 & 0 & -\frac{1}{T_3} & 0 \\ 0 & 0 & 0 & -\frac{1}{T_4} \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} \frac{\gamma_1 k_1}{A_1} & 0 \\ 0 & \frac{\gamma_2 k_2}{A_2} \\ 0 & \frac{(1-\gamma_2)k_2}{A_3} \\ \frac{(1-\gamma_1)k_1}{A_4} & 0 \end{bmatrix}}_{B} u \\ y = \underbrace{\begin{bmatrix} k_c & 0 & 0 & 0 \\ 0 & k_c & 0 & 0 \end{bmatrix}}_{C} x \end{cases}. \qquad (76)$$

and A_j is the cross-section of Tank j; $a_j$is cross-section of the outlet hole and h_j is water level of Tank $j (j = 1,2,3,4)$. The acceleration due to gravity is denoted by g, and the superscript 0 indicates the stationary point.

TABLE III. Comparison of Cases 1 and 3 for Example 2

Method		ILC of [14]	OPILC proposed in this paper
Error Reduction Rate	$I_{50\%}$	17	1
	$I_{10\%}$	57	1
	$I_{1\%}$	390	1
Convergence Speed	$I_{10^{-3}}$	16250	4
Computational Complexity	T_1 (sec.)	0.178568	0.319841

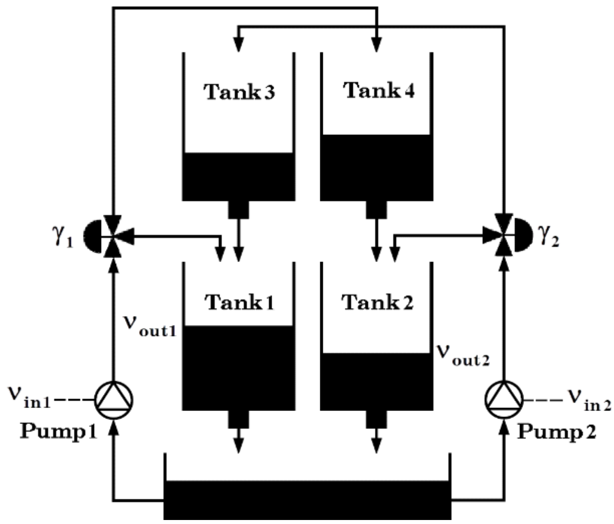


Fig. 14. Schematic diagram of the quadruple-tank process.

In the state-space model (76), the variables and parameters are defined as follows:

$$x=\begin{bmatrix}x_1 & x_2 & x_3 & x_4\end{bmatrix}^T=\begin{bmatrix}h_1 & h_2 & h_3 & h_4\end{bmatrix}^T-\begin{bmatrix}h_1^0 & h_2^0 & h_3^0 & h_4^0\end{bmatrix}^T,$$
$$u=\begin{bmatrix}u_1 & u_2\end{bmatrix}^T=\begin{bmatrix}\mathrm{v}_{\text{in1}} & \mathrm{v}_{\text{in2}}\end{bmatrix}^T-\begin{bmatrix}\mathrm{v}_{\text{in1}}^0 & \mathrm{v}_{\text{in2}}^0\end{bmatrix}^T,$$
$$y=\begin{bmatrix}y_1 & y_2\end{bmatrix}^T=\begin{bmatrix}\mathrm{v}_{\text{out1}} & \mathrm{v}_{\text{out2}}\end{bmatrix}^T-\begin{bmatrix}\mathrm{v}_{\text{out1}}^0 & \mathrm{v}_{\text{out2}}^0\end{bmatrix}^T,$$
$$T_j=\frac{A_j}{a_j}\sqrt{\frac{2h_j^0}{g}} \qquad j=1,2,3,4. \tag{77}$$

The linear model (76) is a fourth-order system that has two inputs and two outputs. We wish to control this system iteratively in a finite time internal $[0,t_f]$.

For simulation the parameters and stationary point of the process are chosen as [52], which are follows:

$A_1=A_3=28\text{cm}^2, A_2=A_4=32\text{cm}^2,$
$a_1=a_3=0.071\text{cm}^2, a_2=a_4=0.057\text{cm}^2,$
$k_c=0.5\frac{V}{\text{cm}}, k_1=3.33\frac{\text{cm}^3}{\text{Vs}}, k_2=3.35\frac{\text{cm}^3}{\text{Vs}},$
$\gamma_1=0.7, \gamma_2=0.6, g=981\frac{\text{cm}}{s^2},$
$h_1^0=12.4\text{cm}, h_2^0=12.7\text{cm},$
$h_3^0=1.8\text{cm}, h_4^0=1.4\text{cm},$
$v_{\text{in1}}^0=v_{\text{in2}}^0=3V.$

The time duration of iterations is selected ten seconds ($t_f=10sec$). The desired output trajectories, and the state and output disturbances are chosen as follows:

$$y_{1d}(t)=\begin{cases}\frac{1}{6}t & 0\le t\le 3\\ 0.5 & 3\le t\le 7\\ \frac{1}{6}(10-t) & 7\le t\le 10\end{cases},$$
$$y_{2d}(t)=0.01\sinh\left(5\sin(0.2\pi t)\right).$$

$$\eta_x(t)=\begin{bmatrix}cos(t)\,tanh(0.75t)\\ 0.5\,sin(t)\\ 0.25\,cos(\pi t)\\ -\text{Ln}(t+1)\end{bmatrix},\eta_y(t)=\begin{bmatrix}0.01t^2\\ 0.5(1-e^{sin(0.25\pi t)})\end{bmatrix}.$$

The system initial conditions and the initial input all selected to be zero.

To compare the performance of the proposed method with the approach in [14], simulations of this example are conducted under two scenarios, applying the method proposed here and the approach of [14], respectively.

Case 1. Using the OPILC proposed in this paper

The symmetric positive definite matrices $Q(t)$ and $R(t)$ in the cost function (5) are chosen as follows:

$Q(t)=10^4I, R(t)=I.$

The simulated output trajectories of the system at several iterations, along with the desired trajectories, are shown in Figs. 15a and 15b. These plots demonstrate that as the iteration index increases, the system outputs quickly converge to the desired trajectories. In fact, at iteration $i=2$, a perfect match is observed between the actual and desired outputs.

Fig. 16 further illustrates the evolution of $\frac{1}{2}\|e_i(.)\|_Q^2$ and J_i over iterations. Both measures are seen to decrease steadily and approach zero, which supports the theoretical findings established in Theorems 2 and 3.

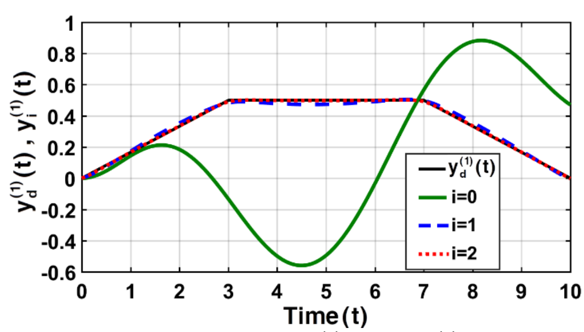


Fig. 15a. The output trajectories $y_d^{(1)}(t)$ and $y_i^{(1)}(t) (i=0,1,2)$ for Example 3 (Case 1).

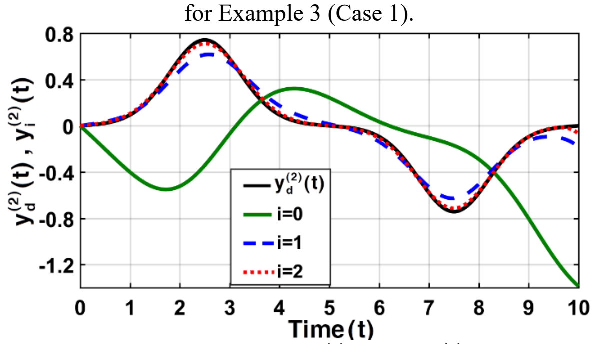


Fig. 15b. The output trajectories $y_d^{(2)}(t)$ and $y_i^{(2)}(t) (i=0,1,2)$ for Example 3 (Case 1).

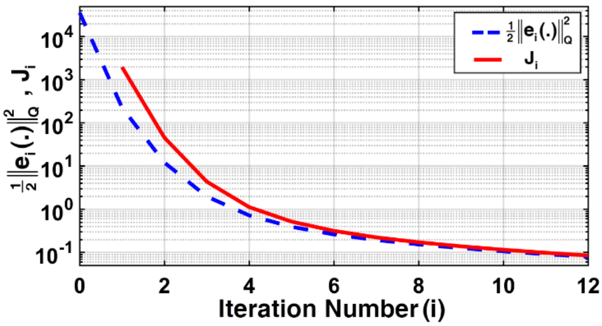


Fig. 16. The monotonic convergence of $\|e_i(.)\|_Q$ and J_i with respect to i for Example 3 (Case 1).

The system input profiles, shown in Figs. 17a and 17b, change gradually with each iteration and ultimately settle into two consistent forms. These patterns are not predetermined, but instead result from the underlying learning mechanism.

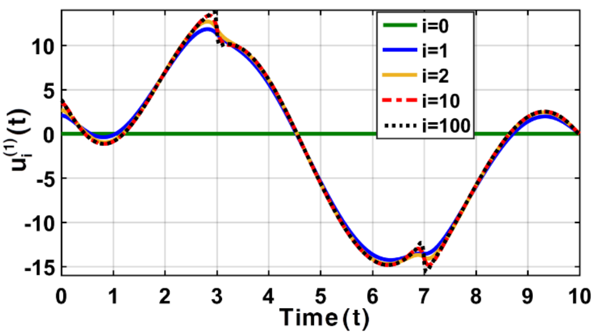


Fig. 17a. The system input $u_i^{(1)}(t)$ at iterations $i = 0,1,2,10,100$ for Example 3 (Case 1).

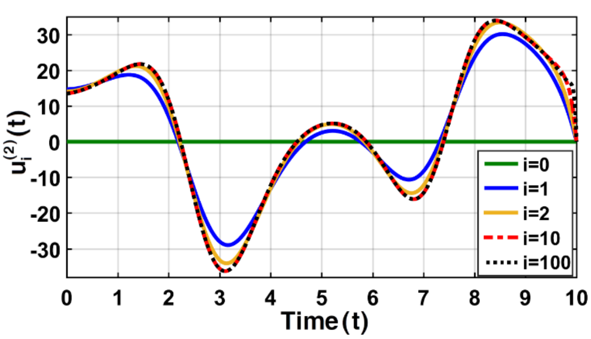


Fig. 17b. The system input $u_i^{(2)}(t)$ at iterations $i = 0,1,2,10,100$ for Example 3 (Case 1).

Case 2. Using the ILC presented in [14]

The ILC proposed in [14], which is discussed in Case 3 of Example 2, is now applied to the system in Example 3. Since the method in [14] is designed for discrete-time systems, and the system in Example 3 is continuous-time, we first discretize the system using a sampling time of $\Delta t = 0.01$. Then, we have:

$$N = \frac{t_f}{\Delta t} = 10^3.$$

The following learning gain is chosen:

$$F(k) = \begin{bmatrix} 10 & -5 \\ 0.8 & 20 \end{bmatrix} \text{for} k = 1,2,\ldots,10^3.$$

It can be readily shown that this learning gain satisfies the convergence condition (74).

Figures 18a and 18b illustrate the tracking performance of the system outputs $y_i^{(k)}(t)$ relative to the desired trajectories $y_d^{(k)}(t)$ at iterations $i = 0,50,100$, for $k = 1,2$, respectively. As evident from these figures, the tracking accuracy improves progressively with increasing iteration number, leading to convergence toward the reference signals. Nevertheless, a comparative analysis with Figs. 15a and 15b reveals that the convergence rate in Case 1 is considerably higher. To provide a more rigorous assessment, the total squared tracking error $EE(i)$, as defined in equation (75), is employed. Figure 19 presents the evolution of $EE(i)$ across iterations for both Case 1 and Case 2. The results confirm that the OPILC scheme proposed in this study offers significantly enhanced convergence performance, both in terms of speed and accuracy, compared to the ILC approach reported in [14].

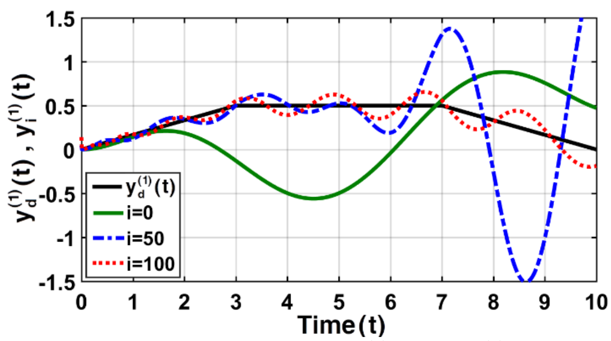


Fig. 18a. The output trajectories $y_d^{(1)}(t)$ and $y_i^{(1)}(t)(i = 0,50,100)$ for Example 3 (Case 2).

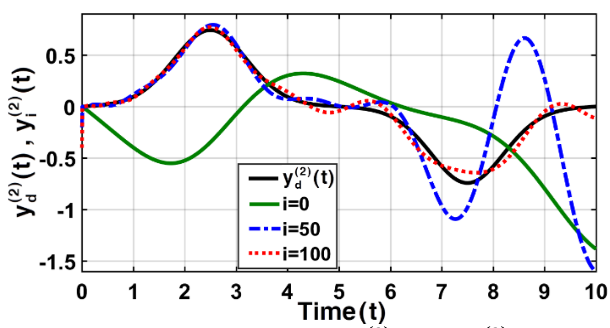


Fig. 18b. The output trajectories $y_d^{(2)}(t)$ and $y_i^{(2)}(t)(i = 0,50,100)$ for Example 3 (Case 2).

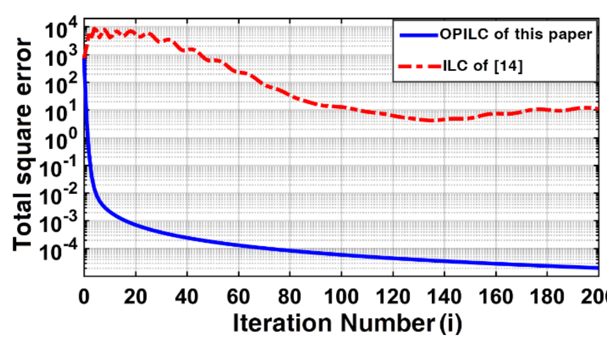


Fig. 19. The total square errors of tracking with respect to i for Example 3 (comparison of Cases 1 and 2)

Table IV presents the quantitative performance measures comparing Cases 1 and 2 for Example 3. In this table, the error metric refers to the total square error of tracking, as defined in equation (75). As the data indicate, the proposed method also demonstrates substantially superior error reduction rate and convergence speed in Example 3 compared to the ILC method reported in [14]. Specifically, a single iteration of the proposed algorithm leads to a 99% reduction in tracking error, whereas the ILC approach in [14] requires 232 iterations to achieve a similar reduction. Additionally, the convergence speed of the proposed method in this example exceeds that of the referenced ILC by a factor of over 83. While the method requires a per-iteration runtime approximately 1.7 times that of the ILC technique referenced in [14], this elevated computational demand, as will be further addressed in Remark 6, does not pose any meaningful limitations in practical applications.

TABLE IV. Comparison of Cases 1 and 2 for Example 3

Method		ILC of [14]	OPILC proposed in this paper
Error Reduction Rate	$I_{50\%}$	56	1
	$I_{10\%}$	71	1
	$I_{1\%}$	232	1
Convergence Speed	$I_{10^{-3}}$	1330	16
Computational Complexity	T_1 (sec.)	0.034856	0.058784

Example 4. A relatively complex fifth-order fully time-varying MIMO system is given in the form of (1) with the following coefficients:

$$A(t) = 0.4\begin{bmatrix} cos(2\pi t) & t & 2t-4 & e^{-2t} & 0.01(t^2-1) \\ sin(\pi t^2) & e^{-t} & 2\,sin(3\pi t) & 0.1\,cosh(t) & 0.2\,sin(0.5\pi\sqrt{t}) \\ (t)^{e^{-0.5t}} & t\,cos(4\pi t) & \mathrm{Ln}(t+0.01) & 0.1t & -t^2 \\ -20\,tan^{-1}(4t) & sin(cosh(t)) & -\dfrac{te^{-0.1t}}{t+10} & 3\,cos^{-1}(0.02t) & 4\,cos(2\pi t^3) \\ 25-t & -\mathrm{Ln}(1.1+sin(20t)) & (t+1)^{sin^{-1}(0.01t)} & e^{cos(8t)} & \dfrac{t}{t+25} \end{bmatrix},$$

$$B(t) = 0.02\begin{bmatrix} 2+te^{-t} & t-cos(t) & \sqrt[3]{sin(\pi t)} \\ t^2 & \mathrm{Ln}(0.1+t) & te^{-t}cos(\pi\sqrt{t}) \\ sin(4t) & \dfrac{3t^2}{t^2+4} & 7e^{sinh^{-1}(0.25t)} \\ \sqrt{\dfrac{2}{t+1}} & e^{2t} & tanh[sin(t^2)] \\ -t & \sqrt{10-t} & \dfrac{e^{2t}}{t+3} \end{bmatrix},$$

$$C(t) = 0.1\begin{bmatrix} \dfrac{1}{1+t^{25-t}} & t^{t+1} \\ (sin^{-1}(e^{-0.02t}))^3 & cos(t^2) \\ \sqrt[5]{t} & \dfrac{sinh^{-1}(\mathrm{Ln}(1+t))}{e^t+7} \\ e^{sin(\pi t)} & 0.5t-3 \\ -\dfrac{1}{tan^{-1}(t)+0.1} & |t-2.5| \end{bmatrix}^T.$$

We aim to control this complex system iteratively in the time interval [0,5].

Let the desired trajectories be defined as:

$$\begin{cases} y_d^{(1)}(t) = 1.25t^2 - 0.05t^4 \\ y_d^{(2)}(t) = t\,tan^{-1}(2.5-t) \end{cases} \quad 0 \le t \le 5.$$

To perform the simulation, the initial conditions of the system and its input at the initial iteration $i = 0$ are all assumed to be zero. The disturbances $\eta_x(t)$ and $\eta_y(t)$ are chosen as follows:

$$\eta_x(t) = 0.1\begin{bmatrix} te^{-0.5t} \\ cosh(sin(t)) \\ t(20-t) \\ -5\,cos(0.5t) \\ tan^{-1}(t) \end{bmatrix}, \eta_y(t) = 0.5\begin{bmatrix} sin(\mathrm{Ln}(t+1)) \\ \mathrm{Ln}(sin(t)+2) \end{bmatrix}.$$

This example presents a particularly challenging scenario, under which previously reported ILC methods do not achieve convergence, to the best of our knowledge. Let us now apply the proposed OPILC to this system.

The symmetric positive definite matrices $Q(t)$ and $R(t)$ as follows:

$Q(t) = 10I, R(t) = I.$

Here, due to the limitation in the number of pages, only the results obtained for $\frac{1}{2}\|e_i(.)\|_Q^2$ and cost function J_i from the simulation are given in Fig. 20 (in the log scale on both axes). This figure proves that the learning convergence is monotonic from the point of view of both $\|e_i(.)\|_Q$ and J_i, consistent with Theorems 2 and 3.

To study the effect of the weighing matrices $Q(t)$ and $R(t)$ in cost function (5) on the performance of the proposed OPILC, the simulation of Example 4 is repeated using $Q(t) = \lambda I$ and $R(t) = I$, for several different values of the positive scalar λ. As a result, the obtained profiles for the 2-norm of the tracking error $e_i(t)$ and the input variation $v_i(t)$, which are defined as (73), are given in Figs. 21 and 22 in the log scales, respectively. These figures clearly demonstrate that with the increase of λ, the convergence rate of $e_i(t)$ increases, while that of $v_i(t)$ decreases. To comprehend and explain this phenomenon, it can be stated that by choosing large or small weighing matrices $Q(t)$ and $R(t)$ relative to each other, the desired importance can be assigned to each term $\|e_i(.)\|_Q^2$ and $\|v_i(.)\|_R^2$ in cost function (5). That is, if the minimization of the tracking error has a higher significance, then $Q(t)$ must be chosen larger than $R(t)$. However, if the prevention of sharp and fast variation of the system input is of interest, then $Q(t)$ must be chosen to be small relative to $R(t)$.

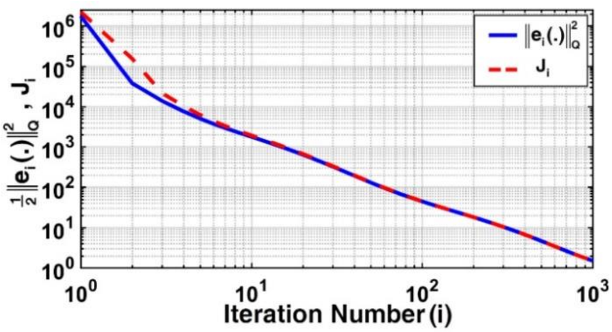


Fig. 20. The monotonic convergence of $\|e_i(.)\|_Q$ and J_i with respect to i for Example 4.

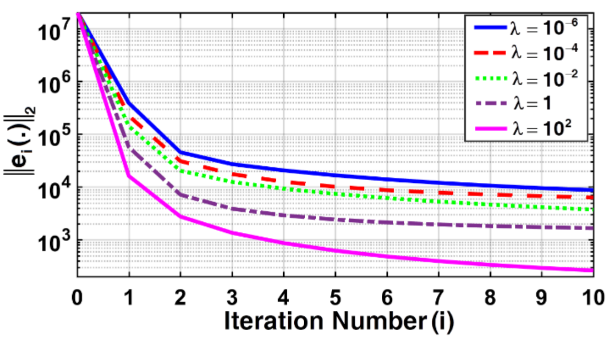


Fig. 21. The profiles of the 2-norm of the tracking error $e_i(t)$ for several different values of λ for Example 4.

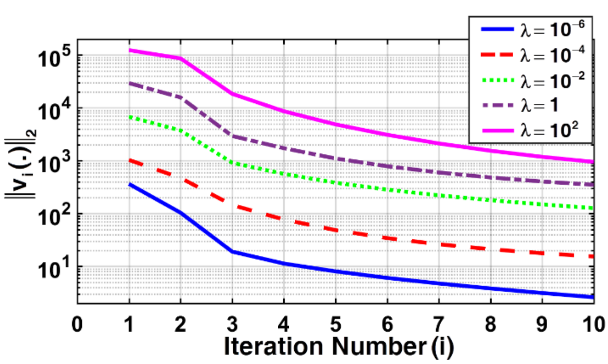


Fig. 22. The profiles of the 2-norm of the input variation $v_i(t)$ for several different values of λ for Example 4.

Remark 6. The proposed OPILC has been compared with several existing ILC approaches. Due to space limitations, only three of these comparisons are presented in Examples 1 through 3. The results clearly demonstrate that the proposed OPILC achieves significantly higher error reduction rate and faster convergence speed than previous methods. Furthermore, the convergence behavior is monotonic, an important characteristic for any ILC algorithm, which has also been theoretically proven in Theorems 2 and 3. However, as shown in Tables I to IV, the proposed method exhibits higher computational complexity than some existing ILC approaches, leading to a longer computation time for generating the control input in each iteration. To identify the source of this increased computation time, let us analyze the structure of the proposed OPILC. As noted in Remark 4, the matrix Riccati differential equation (15) is solved only once, prior to the first iteration, and the resulting gain matrix $K(t)$ is used throughout all subsequent iterations. Therefore, the time required to solve equation (15) does not contribute to the per-iteration computational burden. The main source of increased computation lies in solving the linear vector differential equation (16) to compute $g_i(t)$, which must be carried out before each iteration, as discussed in Remark 5. Nevertheless, in practical industrial repetitive systems (such as robotic manipulators) a reset period typically exists between two successive iterations. This period is usually long enough to accommodate the solution of equation (16), making its computational burden practically negligible. Thus, the additional computational effort does not pose any serious challenge or limitation in real-world applications.

VI. Conclusion

This paper introduced a novel optimal ILC approach for MIMO continuous-time LTV systems. The proposed method transforms the ILC problem into a linear quadratic tracking-like (LQT) problem. Subsequently, LQT theory is employed to derive an explicit closed-loop learning control law that guarantees both accurate tracking and minimized control effort. In contrast to many existing ILC methods for continuous-time systems that depend on derivative-based learning algorithms, which are sensitive to measurement noise, the proposed design avoids these terms, mitigating noise amplification.

Through rigorous mathematical analysis, it has been proven that the proposed OPILC guarantees the monotonic convergence of both the tracking error and the associated cost function, a property that is essential for any high-performance ILC. The effectiveness of the method is demonstrated through four comprehensive simulation examples, which show that the proposed scheme achieves faster convergence and higher tracking accuracy compared to existing ILC methods. Additionally, the method demonstrates significant computational efficiency by determining certain learning gains prior to the first iteration and updating others during system reset periods.

In summary, the proposed OPILC framework offers a powerful solution for iterative learning control in continuous-time systems, providing a foundation for further advancements and extensions to more complex systems. Extending the approach to discrete-time LTV systems is straightforward. However, the formulation and solution of the problem for each of the following cases require further investigation:

1. Output Measurement: The current approach assumes full access to the state vector. Extending this method to cases where only output measurements are available is an important future step.

2. System Constraints: Incorporating input and state constraints into the OPILC framework is crucial for practical applications where physical limitations must be considered.

3. Iteration-Varying Disturbances: The method assumes iteration-invariant disturbances, but many real-world systems experience iteration-varying disturbances. Accounting for this in the optimization framework could improve robustness.

4. System Uncertainties: Future work could explore handling system uncertainties and unknown parameters within the OPILC framework, enhancing its robustness in practical applications.

5. Nonlinear Systems: The proposed method is tailored for linear systems. Extending it to nonlinear systems remains a valuable research direction, as many real-world systems exhibit nonlinear dynamics.

REFERENCES

[1] H.-S. Ahn, Y. Chen, and K. L. Moore, "Iterative learning control: Brief survey and categorization," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 37, pp. 1099-1121, 2007, doi: 10.1109/TSMCC.2007.905759.

[2] J.-X. Xu, "A survey on iterative learning control for nonlinear systems," International Journal of Control, vol. 84, pp. 1275-1294, 2011, doi: 10.1080/00207179.2011.574236.

[3] A. Madady, "A self-tuning iterative learning controller for time variant systems," Asian Journal of Control, vol. 10, pp. 666-677, 2008, doi: 10.1002/asjc.67.

[4] D. Shen and Y. Wang, "Survey on stochastic iterative learning control," Journal of Process Control, vol. 24, pp. 64-77, 2014, doi: 10.1016/j.jprocont.2014.04.013.

[5] S. Arimoto, S. Kawamura, and F. Miyazaki, "Bettering operation of robots by learning," Journal of Robotic systems, vol. 1, pp. 123-140, 1984, doi: 10.1002/rob.4620010203.

[6] D. Shen, "Iterative learning control with incomplete information: A survey," IEEE/CAA Journal of Automatica Sinica, vol. 5, pp. 885-901, 2018, doi: 10.1109/JAS.2018.7511123.

[7] F. Afsharnia, A. Madady, and M. B. Menhaj, "Adaptive iterative learning control for unknown linear time-varying continuous systems," Transactions of the Institute of Measurement and Control, vol. 42, pp. 981-996, 2020, doi: 10.1177/0142331219880110.

[8] J. Zhou, C. T. Freeman, and W. Holderbaum, "Multiple-model iterative learning control with application to stroke rehabilitation," Control Engineering Practice, vol. 154, p. 106134, 2025, doi: 10.1016/j.conengprac.2024.106134.

[9] H. Farokhi Moghadam, N. Vasegh, and S. M. Seyed Moosavi, "Adaptive repetitive control for periodic disturbance rejection with unknown period," International Journal of Industrial Electronics Control and Optimization, vol. 3, pp. 187-194, 2020, doi: 10.22111/ieco.2019.31017.1197.

[10] F. Afsharnia, A. Madady, and M. B. Menhaj, "Neural iterative learning identifier-based iterative learning controller for time-varying nonlinear systems," Asian Journal of Control, vol. 24, pp. 1998-2012, 2022, doi: 10.1002/asjc.2617.

[11] R. Movahed Asl and A. Madady, "Stabilization of two-dimensional mixed continuous-discrete-time systems via dynamic output feedback with application to iterative learning control design," Transactions of the Institute of Measurement and Control, vol. 44, pp. 172-187, 2022, doi: 10.1177/01423312211022205.

[12] X. Zhang and Z. Hou, "Data-driven predictive point-to-point iterative learning control," Neurocomputing, vol. 518, pp. 431-439, 2023, doi: 10.1016/ j.neucom.2022.11.014.

[13] Y. Tao, H. Tao, Z. Zhuang, V. Stojanovic, and W. Paszke, "Quantized iterative learning control of communication-constrained systems with encoding and decoding mechanism," Transactions of the Institute of Measurement and Control, vol. 46, pp. 1943-1954, 2024, doi: 10.1177/01423312231225782.

[14] Y. Kong, X.-D. Li, and X. Li, "Iterative learning control approach for linear discrete time-varying systems with different initial time points," Journal of the Franklin Institute, vol. 361, p. 106702, 2024, doi: 10.1016/j.jfranklin.2024.106702.

[15] J. Liu, Y. Liu, R. Chi, and Z. Hou, "Error tracking-driven adaptive backstepping ILC of nonlinear systems with initial shifts," Journal of the Franklin Institute, vol. 362, p. 107455, 2025, doi: 10.1016/j.jfranklin.2024.107455.

[16] S-M Hashem Zadeh, S. Khorashadizadeh, M-M Fateh, and M. Hadadzarif, "Optimal sliding mode control of a robot manipulator under uncertainty using PSO," Nonlinear Dynamics, vol. 84, pp. 2227–2239, 2016, doi: 10.1007/s11071-016-2641-4.

[17] B. Dai, J. Gong, C. Li, and H. Ning, "Iterative learning control realized using an iteration-varying forgetting factor based on optimal gains," Transactions of the Institute of Measurement and Control, vol. 43, pp. 2334-2344, 2021, doi: 10.1177/0142331221996507.

[18] H. Jay, "Model based iterative learning control with a quadratic criterion for linear time-varying systems," Automatica, vol. 36, pp. 641-657, 2000, doi: 10.1016/S0005-1098(99)00194-6.

[19] D. H. Owens and K. Feng, "Parameter optimization in iterative learning control," International Journal of Control, vol. 76, pp. 1059-1069, 2003, doi: 10.1080/0020717031000121410.

[20] S. S. Saab, "Optimal selection of the forgetting matrix into an iterative learning control algorithm," IEEE Transactions on Automatic Control, vol. 50, pp. 2039-2043, 2005, doi: 10.1109/TAC.2005.860232.

[21] J. Hätönen, D. H. Owens, and K. Feng, "Basis functions and parameter optimisation in high-order iterative learning control," Automatica, vol. 42, pp. 287-294, 2006, doi: 10.1016/j.automatica.2005.05.025.

[22] A. Madady, "PID type iterative learning control with optimal gains," International Journal of Control, Automation, and Systems, vol. 6, pp. 194-203, 2008.

[23] S. Mishra, U. Topcu, and M. Tomizuka, "Optimization-based constrained iterative learning control," IEEE Transactions on Control Systems Technology, vol. 19, pp. 1613-1621, 2010, doi: 10.1109/TCST.2010.2083663.

[24] D. H. Owens, "Multivariable norm optimal and parameter optimal iterative learning control: a unified formulation," International Journal of Control, vol. 85, pp. 1010-1025, 2012, doi: 10.1080/00207179.2012.673136.

[25] A. Madady, "An extended PID type iterative learning control," International Journal of Control, Automation and Systems, vol. 11, pp. 470-481, 2013, doi: 10.1007/s12555-012-0350-4.

[26] H. Sun and A. G. Alleyne, "A computationally efficient norm optimal iterative learning control approach for LTV

systems," Automatica, vol. 50, pp. 141-148, 2014, doi: 10.1016/j.automatica.2013.09.009.
[27] J. Van Zundert, J. Bolder, and T. Oomen, "Optimality and flexibility in iterative learning control for varying tasks," Automatica, vol. 67, pp. 295-302, 2016, doi: 10.1016/j.automatica.2016.01.026.
[28] A. Madady, H.-R. Reza-Alikhani, and S. Zamiri, "Optimal N-parametric type iterative learning control," International Journal of Control, Automation and Systems, vol. 16, pp. 2187-2202, 2018, doi: 10.1007/s12555-017-0259-z.
[29] F. Memon and C. Shao, "An optimal approach to online tuning method for PID type iterative learning control," International Journal of Control, Automation and Systems, vol. 18, pp. 1926-1935, 2020, doi: 10.1007/s12555-018-0840-0.
[30] F. Memon and C. Shao, "Robust optimal PID type ILC for linear batch process," International Journal of Control, Automation and Systems, vol. 19, pp. 777-787, 2021, doi: 10.1007/s12555-019-1033-1.
[31] Y. Zhang, B. Chu, and Z. Shu, "Parameter optimal iterative learning control design: from model-based, data-driven to reinforcement learning," IFAC-PapersOnLine, vol. 55, pp. 494-499, 2022, doi: 10.1016/ j.ifacol.2022.07.360.
[32] Z. Zhuang, H. Tao, Y. Chen, V. Stojanovic, and W. Paszke, "An optimal iterative learning control approach for linear systems with nonuniform trial lengths under input constraints," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 53, pp. 3461-3473, 2022, doi: 10.1109/TSMC.2022.3225381.
[33] F. Arnold and R. King, "Norm-optimal iterative learning control in an integer-valued control domain," International Journal of Control, vol. 96, pp. 170-181, 2023, doi: 10.1080/00207179.2021.1983213.
[34] Y. Zhou, K. Gao, X. Tang, H. Hu, D. Li, and F. Gao, "Conic input mapping design of constrained optimal iterative learning controller for uncertain systems," IEEE Transactions on Cybernetics, vol. 53, pp. 1843-1855, 2022, doi: 10.1109/TCYB.2022.3155754.
[35] K. Tsurumoto, W. Ohnishi, T. Koseki, M. van Haren, and T. Oomen, "Combined time-domain optimization design for task-flexible and high performance ILC," in 2024 European Control Conference (ECC), 2024, pp. 1190-1195, doi: 10.23919/ECC64448.2024.10590964.
[36] M. van Haren, K. Tsurumoto, M. Mae, L. Blanken, W. Ohnishi, and T. Oomen, "A frequency-domain approach for enhanced performance and task flexibility in finite-time ILC," European Journal of Control, p. 101033, 2024, doi: 10.1016/j.ejcon.2024.101033.
[37] E. Rafajłowicz and W. Rafajłowicz, "Iterative learning in optimal control of linear dynamic processes," International Journal of Control, vol. 91, pp. 1522-1540, 2018, doi: 10.1080/00207179.2017.1320810.
[38] Z. Zhuang, H. Tao, Y. Chen, T. Oomen, W. Paszke, and E. Rogers, "Optimal iterative learning control design for continuous-time systems with nonidentical trial lengths using alternating projections between multiple sets," Journal of the Franklin Institute, vol. 360, pp. 3825-3848, 2023, doi: 10.1016/j.jfranklin.2023.02.006.
[39] Y. Wu and D. Meng, "Data informativity for tracking control of learning systems: Test and design conditions," Automatica, vol. 171, p. 111885, 2025, doi: 10.1016/j.automatica.2024.111885.
[40] K. Gao, J. Lu, Y. Zhou, and F. Gao, "A computationally efficient policy optimization scheme in feedback iterative learning control for nonlinear batch process," Computers & Chemical Engineering, p. 109005, 2025, doi: 10.1016/j.compchemeng.2025.109005.
[41] A. Madady and H. R. Reza-Alikhani, "A guaranteed monotonically convergent iterative learning control," Asian Journal of Control, vol. 14, pp. 1299-1316, 2012, doi: 10.1002/asjc.397.
[42] D. H. Owens, Iterative learning control: An optimization paradigm. London: Springer, 2016, ISBN: 978-1-4471-6772-3.
[43] R. Dosthosseini, N. Banisaeid, and A. Dashti, "Dynamic modelling and optimal control of a rotor of active magnetic bearings," International Journal of Industrial Electronics Control and Optimization, vol. 2, pp. 59-70, 2019, doi: 10.22111/IECO.2018.25595.1053.
[44] P. L. Falb and M. Athans, Optimal control: An introduction to the theory and its applications: McGraw-Hill, 2000, ISBN: 0070024138, 9780070024137.
[45] F. L. Lewis, D. Vrabie, and V. L. Syrmos, Optimal control: John Wiley & Sons, 2012, ISBN-10. 0470633492 ; ISBN-13. 978-0470633496.
[46] P. D. Nguyen and N. H. Nguyen, "An intelligent parameter determination approach in iterative learning control," European Journal of Control, vol. 61, pp. 91-100, 2021, doi: 10.1016/j.ejcon.2021.06.001.
[47] X. Ruan and Y. Liu, "Monotone convergence rate of norm-optimal-gain-arguable iterative learning control for LDTI systems," Asian Journal of Control, vol. 24, pp. 920-941, 2022, doi: 10.1002/asjc.2498.
[48] X.-D. Li, J. K. Ho, and T. W. Chow, "Iterative learning control for linear time-variant discrete systems based on 2-D system theory," IEE Proceedings-Control Theory and Applications, vol. 152, pp. 13-18, 2005, doi: 10.1049/ip-cta:20041125.
[49] M. J. Sidi, Spacecraft dynamics and control: a practical engineering approach vol. 7: Cambridge university press, 1997, ISBN-10: 0521787807.
[50] M. Tranninger, R. Seeber, J. G. Rueda-Escobedo, and M. Horn, "Strong detectability and observers for linear time-varying systems," Systems & Control Letters, vol. 170, p. 105398, 2022, doi: 10.1016/j.sysconle.2022.105398.
[51] R. Ma and G. Zhang, "Iterative learning tracking control for a class of MIMO nonlinear time-varying systems," International Journal of Modelling, Identification and Control, vol. 27, pp. 271-278, 2017, doi: 10.1504/IJMIC.2017.084721.
[52] K. H. Johansson, "The quadruple-tank process: A multivariable laboratory process with an adjustable zero," IEEE Transactions on control systems technology, vol. 8, pp. 456-465, 2000, doi: 10.1109/87.845876.

Appendix

Let us consider a linear time-varying system whose state $x(t) \in \mathbb{R}^n$, input $u(t) \in \mathbb{R}^p$, and output $y(t) \in \mathbb{R}^q$ are related by the system of equations:

$$\begin{cases} \dot{x}(t) = A(t)x(t) + B(t)u(t) \\ y(t) = C(t)x(t) \end{cases} . \tag{A-1}$$

For system (A-1), the linear quadratic tracking (LQT) problem is defined as follows [44, 45]:

The desired output $\mathrm{r(t)} \in \mathbb{R}^{\mathrm{q}}$ is given, while $\mathrm{u(t)}$ is unconstrained in magnitude. Determine $\mathrm{u(t)}$ such that the following quadratic cost function is minimized:

$$J = \frac{1}{2}\left[r(t_f) - y(t_f)\right]^T F\left[r(t_f) - y(t_f)\right]$$
$$+\frac{1}{2}\int_0^{t_f}\{[r(t) - y(t)]^T Q(t)[r(t) - y(t)] + u^T(t)R(t)u(t)\}\,dt. \quad \text{(A-2)}$$

where the terminal time t_f is specified, $Q(t) \in \mathbb{R}^{q\times q}$ and $R(t) \in \mathbb{R}^{p\times p}$ are the positive definite matrices, and $F \in \mathbb{R}^{q\times q}$ is a constant positive semidefinite matrix.

The calculus of variations has been employed to demonstrate that the above optimal control problem admits a unique solution, expressed analytically as follows (see Chapter 9 of [44] or Chapter 12 of [45]):

$$u(t) = -R^{-1}(t)B^T(t)[K(t)x(t) - g(t)]. \quad \text{(A-3)}$$

where $K(t) \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix, obtained from the solution of the following Riccati-type matrix differential equation:

$$\dot{K}(t) = -K(t)A(t) - A^T(t)K(t)$$
$$+K(t)B(t)R^{-1}(t)B^T(t)K(t)$$
$$-C^T(t)Q(t)C(t), \quad \text{(A-4)}$$

with the terminal condition

$$K(t_f) = C^T(t_f)FC(t_f). \quad \text{(A-5)}$$

The vector $g(t) \in \mathbb{R}^n$ is the solution of the linear vector differential equation:

$$\dot{g}(t) = -[A(t) - B(t)R^{-1}(t)B^T(t)K(t)]^T g(t)$$
$$-C^T(t)Q(t)r(t), \quad \text{(A-6)}$$

with the terminal condition

$$g(t_f) = C^T(t_f)Fr(t_f). \quad \text{(A-7)}$$

It has also been shown that applying input (A-3) to system (A-1) yields the minimum of the cost function (A-2), as follows

$$J^* = \frac{1}{2}x^T(0)K(0)x(0) - x^T(0)g(0) + \phi(0). \quad \text{(A-8)}$$

where $\phi(0)$ is obtained from the backward integration of the following scalar differential equation (see Chapter 9 of [44] or Chapter 12 of [45]):

$$\dot{\phi}(t) = -\frac{1}{2}r^T(t)Q(t)r(t) + \frac{1}{2}g^T(t)B(t)R^{-1}(t)B^T(t)g(t), \quad \text{(A-9)}$$

with the terminal condition:

$$\phi(t_f) = \frac{1}{2}r^T(t_f)Fr(t_f). \quad \text{(A-10)}$$

Naser Taghvamanesh was born in 1988 in Boroujerd, Iran. He received his B.Sc., M.Sc. degrees in electrical engineering from Shahid Sattari University of Science and Technology and Tafresh University in 2005 and 2014, respectively. He is currently pursuing the Ph.D. degree in the Control Department at the Faculty of Electrical Engineering, Tafresh University. His research interests are learning control systems, convex optimization, and optimal control.

Ali Madady was born in Aqdash, Saveh, Iran in 1966. He received his B.Sc., M.Sc. and Ph.D. degrees in electrical and electronic engineering from Amir-Kabir University of Technology (Tehran Polytechnic), Iran, in 1989, 1993 and 2001, respectively. From 1991 to 1996, he was a part-time research member of Iranian Research Organization for Science and Technology (IROST). Also from 1994 to 2001, he was a part-time lecturer of Shamsipour Institute of Technology (SIT), Tehran, Iran. In 2001, he joined Tafresh University, Tafresh, Iran, where he is currently an Associate Professor and originator and head of Control Engineering Group. His research interests include iterative learning, adaptive control, robust control and fractional-order systems.

IECO International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

5G-R Framework for High-Speed Rail Connectivity Using LSTM and Multi-Access Edge Computing

Shahpour Rahmani | Nasser Yazdani

School of Electrical and Computer Engineering, College of Engineering, University of Tehran, Tehran, Iran.[1,2]
Faculty of Mathematics, Statistics and Computer Science, University of Sistan and Baluchestan, Zahedan, Iran.[1]
Corresponding author's email: yazdani@ut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 29-July-2025
Received in revised form: 28-September-2025
Accepted: 13-October-2025
Published online: 23-Sep-2026

Keywords:
5G-R,
Handover Optimization,
LSTM,
Network Slicing,
MEC.

ABSTRACT

High-speed rail systems, operating at speeds up to 350 km/h, face significant challenges in delivering reliable network connectivity due to frequent handovers, signal degradation, and network congestion. This paper proposes the 5G-R framework, an optimized solution integrating beamforming, network slicing, railway-specific LSTM algorithms, and MEC to enhance connectivity performance. The novelty lies in the railway-tuned LSTM's integration with MEC and network slicing. This enables proactive handover prediction based on predictable rail trajectories and reduces latency by 70% while improving handover success by 8% over non-integrated systems. By leveraging real-time train data, such as speed and GPS location, the framework optimizes handover prediction and traffic management, achieving robust performance in diverse environments. Compared to 4G LTE and standard 5G-R, the 5G-R framework demonstrates significant improvements, including a 250 Mbps throughput, 15 ms latency, and 95% handover success rate. Network slicing optimizes resource allocation, reducing congestion by approximately 30% (PRB utilization, validated in simulations with 50/300 UEs), while MEC enables low-latency control for train systems. Field trials along the Beijing-Zhangjiakou railway (174 km, urban/suburban) and simulations validate the framework's adaptability across urban and rural routes. Designed for compatibility with FRMCS, the 5G-R framework lays a foundation for future advancements, including 6G and satellite communications. Future research should focus on optimizing performance in extreme environments and densely populated routes to support autonomous transport systems. This optimization-driven approach establishes a scalable model for next-generation rail communication systems.

I. Introduction

Seamless internet connectivity in high-speed trains is critical for meeting passenger demands for communication, entertainment, and productivity, as well as enabling operational efficiencies like real-time train monitoring and safety systems [1]. However, delivering reliable internet at speeds up to 350 km/h poses significant challenges. Conventional technologies, such as Long-Term Evolution (LTE) and Wi-Fi, face frequent handovers, signal degradation due to Doppler effects, and network congestion from dense user populations (50–300 passengers per train) [1], [2]. Network slicing addresses this congestion by reducing PRB utilization by approximately 30% in high-density scenarios. LTE, limited to ~10 Mbps in rail settings due to handover failures and signal loss, cannot support streaming or real-time applications [3]. Wi-Fi suffers from signal attenuation and limited coverage in rural or tunnel-heavy routes [4]. Environmental interference, such as heavy rain reducing throughput by 10–15%, and sparse rural infrastructure further disrupt connectivity [3], [5].

5G for Railways (5G-R) , a specialized 5G variant, addresses these challenges by leveraging Ultra-Reliable Low-Latency Communication (URLLC) for rapid handovers, Massive Machine-Type Communications (mMTC) for dense device support, and Enhanced Mobile Broadband (eMBB) for high data rates [6]. Despite its potential, 5G-R deployments face handover instability at extreme speeds and limited validation across diverse rail networks, such as mountainous or coastal routes [3], [7]. Prior studies, including the European Future Railway Mobile Communication System

(FRMCS) [7] and isolated approaches like handover protocols [3] or network slicing [8], lack a comprehensive railway-specific framework to simultaneously address handovers, congestion, and signal degradation. Recent advancements in AI-enabled mobility management for beyond 5G (B5G) networks further highlight the need for adaptive handover schemes in high-speed scenarios [39].

This study proposes a novel 5G-R optimization framework integrating beamforming (using Massive Multiple-Input Multiple-Output (MIMO) in sub-6 GHz bands [9]), network slicing, railway-tuned Long Short-Term Memory (LSTM)-based artificial intelligence (AI) for predictive traffic and handover management, Multi-Access Edge Computing (MEC), multi-connectivity, and dynamic spectrum allocation. The novelty lies in the holistic, railway-specific integration of these six components, which is non-trivial due to the need for synchronized real-time data flow (train Global Positioning System (GPS) feeding into LSTM for handover prediction, then MEC for low-latency execution, and network slicing for prioritized allocation)—an approach underexplored in prior isolated studies [3], [15]. This integration creates synergistic impacts, such as reducing overall system latency by 70% compared to non-integrated baselines by enabling proactive resource adjustments tailored to fixed rail trajectories, unlike generic 5G setups that treat mobility as vehicular or urban without rail-specific predictability [16].

The motivation for using LSTM lies in its superior capability to handle sequential, time-dependent data inherent to high-speed rail environments, where factors like train speed, GPS position, and user density evolve predictably over time but are subject to long-term dependencies (recurring patterns along fixed tracks). The railway-tuned LSTM improves over generic mobility prediction models (standard LSTM or Deep Q-Network (DQN) in non-rail contexts) by incorporating rail-specific features like predictable routes and speed profiles, achieving 8% higher handover accuracy (95% vs. 87%) through customized training on Beijing-Zhangjiakou data, reducing false positives in handover triggers by 15% compared to untuned models [15], [17]. Unlike simpler models such as Multi-Layer Perceptrons (MLPs), which struggle with temporal correlations and vanishing gradients in extended sequences, LSTM's gated architecture (including forget, input, and output gates) preserves relevant historical information, enabling precise forecasting of handover zones and traffic loads with minimal error propagation [16], [17]. This is particularly advantageous in rail scenarios, as it leverages the deterministic nature of train trajectories to anticipate disruptions, outperforming non-recurrent alternatives by 5-10% in prediction accuracy for mobility tasks [15]. Validated through Beijing-Zhangjiakou railway trials (174 km, urban/suburban) and MATLAB/NS-3 simulations, the framework achieves 250 Mbps throughput, 15 ms latency, and 95% handover success at 350 km/h in urban settings, outperforming LTE by 2400% and standard 5G (100 Mbps) by 150% in throughput. The railway-tuned LSTM improves handover success by 8% over standard 5G-R benchmarks (Standard 5G-R baseline derived from untuned/non-integrated systems [3], [15]). In rural scenarios, it maintains 220 Mbps and 92% success, ensuring scalability despite potential weather impacts [3]. A cost-benefit analysis projects 20–30% urban cost savings via phased deployments and premium services, enhancing feasibility for railway operators.

Contributions:

1. A holistic 5G-R framework tailored for high-speed rail connectivity.
2. Empirical validation through trials and simulations, with significant performance gains (key performance indicators (KPIs) labeled as trial- or simulation-based in results).
3. Sensitivity and cost-benefit analyses for robustness and deployment feasibility.
4. A foundation for sixth generation network (6G) and satellite-based railway communications.

This study targets railway operators and researchers exploring wireless solutions for high-mobility environments. By establishing 5G-R as a cornerstone for railway communications, it paves the way for innovations like 6G and satellite systems. The paper is organized as follows: Section II reviews related work, Section III details the methodology, Section IV presents performance results, Section V discusses implications, and Section VI concludes with future directions.

II. Related Work

A. Connectivity Challenges in High-Speed Trains

High-speed trains face significant connectivity challenges due to rapid mobility, dense user populations, and diverse network conditions [1], [2]. At speeds up to 350 km/h, Doppler effects degrade signal quality, causing up to 20% signal loss in high-frequency bands [4]. Frequent handovers disrupt service, with handover failure rates reaching 10–15% in dense urban settings [3]. Dense passenger populations (50–300 users per train) lead to network congestion, reducing throughput during peak hours [5]. Environmental factors, such as signal attenuation in tunnels or interference from heavy rain, further degrade performance, with rainfall potentially reducing throughput by 10–15% [3]. Sparse infrastructure in rural or mountainous regions creates coverage gaps, limiting connectivity for passengers and operational systems like train monitoring [5]. These challenges necessitate railway-specific wireless solutions capable of handling extreme mobility and dynamic conditions [3].

B. 5G for railway

5G-R, a railway-specific 5G variant, leverages URLLC for rapid handovers, mMTC for dense device support, and eMBB for high data rates [6]. The Beijing-Zhangjiakou trial, spanning 174 km, achieved ~200 Mbps throughput, 20–30 ms latency, and <1% packet loss at 350 km/h, significantly outperforming LTE's ~10 Mbps and 5–10% packet loss [3]. The European FRMCS validates 5G-R for train control, passenger services, and safety applications, integrating with legacy systems to ensure interoperability [10]. Positioning techniques using 5G-R enhance train localization, achieving sub-meter accuracy critical for collision avoidance [11]. However, handover instability in high-interference scenarios, such as urban canyons, and high infrastructure costs (~$500,000/km in urban areas) remain barriers, requiring optimized deployment strategies [3], [12].

C. Optimization Techniques

Several techniques enhance 5G-R performance in rail environments. Beamforming, using Massive MIMO (64x64 antennas) and mmWave, mitigates signal loss and Doppler effects, achieving 20–40% throughput gains by directing signals toward moving trains [9]. Network slicing creates virtual sub-networks for passenger internet, train control, and multimedia, reducing congestion by 20–40% in crowded carriages (e.g., 300 users) per 3GPP standards [8], [13]. Edge computing processes data at base stations, reducing latency to 10–20 ms for URLLC applications like real-time train control and predictive maintenance [14]. AI-driven methods optimize resource allocation and handover prediction [15], [16]. DQN achieves 90–92% handover success in generic 5G networks but overlooks railway-specific patterns like predictable train trajectories [15]. MLP models similarly achieve 90–92% success but lack rail-specific tuning [17]. Software-Defined Networking (SDN)-based approaches improve mobility management, with ~85% handover success in rail settings, though they struggle with real-time adaptability to train speeds [18]. Recent studies highlight antenna optimizations for 5G wireless communications. For instance, the authors [40] proposed a compact microstrip antenna with parasitic AMC for 4G/5G, achieving wide bandwidth (4.55–6.83 GHz), 34% size reduction, 8 dBi gain, and 90% efficiency. Advantages include compactness and high efficiency, suitable for mobile applications; disadvantages are lack of testing in high-mobility environments like rail and limited to sub-6 GHz bands, potentially restricting Doppler effect mitigation in high-speed trains. These techniques, while effective individually, lack comprehensive railway-specific integration.

D. AI Techniques for Handover and Traffic Prediction in Rail Systems

LSTM has been extensively used in rail communications for tasks requiring sequential data processing, such as handover prediction, traffic forecasting, and signal quality estimation, where it outperforms non-recurrent models due to its handling of temporal dependencies [16]. For example, in high-speed rail studies, LSTM has been applied to predict handover events using time-series inputs like train velocity and location, achieving accuracies of 92-95% in simulations and trials, as seen in Beijing-Shanghai rail analyses where it reduced handover failures by 10-15% compared to threshold-based methods [1], [16].

Specific similar works include LSTM-based proactive handover schemes for 5G high-speed railway networks, which predict handover events to reduce radio link failures, demonstrating superior performance in mobility management. Beyond rail, LSTM is widely adopted in broader mobility networks, such as vehicular communications for trajectory prediction and in 5G networks for resource allocation, with applications in over 50% of recent AI-driven handover papers due to its robustness to variable sequence lengths [15], [17]. Similar modules include Gated Recurrent Units (GRU), which streamline LSTM by combining forget and input gates into a single update gate, resulting in fewer parameters (typically 25-30% less) and faster training (20-30% quicker convergence) while maintaining comparable accuracy in rail scenarios—e.g., GRU has been used for traffic load prediction in FRMCS systems, offering 93% accuracy with lower computational overhead for edge devices [24], [25].

Other Recurrent Neural Network (RNN) variants, like vanilla RNNs, are simpler but prone to vanishing/exploding gradients, making them less suitable for long rail sequences (e.g., >100 timesteps), where they underperform LSTM by 15-20% in prediction tasks [26]. Bidirectional RNNs (Bi-RNNs) extend standard RNNs by processing sequences forward and backward, improving context awareness for bidirectional rail data (e.g., upcoming and past track conditions) but at double the computational cost, with applications in signal anomaly detection achieving 96% accuracy [27]. Attention and Transformer-based modules, which use self-attention mechanisms for parallel processing of sequences, have been explored for mobility predictions, offering up to 10-15% better accuracy in long-sequence tasks like traffic forecasting in 5G networks compared to LSTM, though they require more computational resources (e.g., for training on large datasets) and are less common in real-time rail due to inference overhead [28].

Temporal Convolutional Networks (TCN) employ causal convolutions for sequence modeling, providing faster inference (often 2-5x quicker than LSTM/GRU) and parallel training, suitable for real-time handover in 5G mobility, though they may underperform on irregular rail patterns requiring deep memory [29]. Convolutional LSTM (ConvLSTM) combines convolutions with LSTM for spatio-temporal data, enhancing predictions in rail environments with spatial elements (e.g., track layouts and multi-train interactions), achieving 5-10% higher accuracy in hybrid scenarios but at increased complexity [30]. Overall, while LSTM remains prevalent in rail due to its balance of accuracy and interpretability, GRUs are favored for resource-limited deployments, Transformers for large-scale data, TCN for speed, and ConvLSTM for spatial-temporal tasks; our framework uses LSTM for its proven efficacy in High-speed rail (HSR) handover, but future comparisons could integrate these alternatives.

E. Research Gaps

Significant gaps persist in 5G-R research for high-speed rail connectivity. Most studies rely on simulations or small-scale trials, with limited real-world validation in rural or mountainous regions where coverage is sparse [11], [19]. Holistic integration of beamforming, network slicing, AI, and edge computing tailored for extreme speeds (up to 350 km/h) remains underexplored [3]. Prior approaches, such as FRMCS [10] and SDN-based mobility management [18], focus on interoperability or generic handover protocols, achieving ~85% handover success but lack railway-specific optimizations. High infrastructure costs (~$500,000/km) and energy efficiency in dense deployments are rarely addressed, limiting scalability [12]. AI-based handover strategies, like DQN and MLP, achieve 90–92% success in generic contexts but fail to leverage railway-specific data [15], [17]. Our railway-tuned LSTM achieves 95% handover success, improving by 3-5% over these generic AI models and by 8% over standard 5G-R benchmarks (87%) [3], [15]. For example, by incorporating train speed and GPS data, our LSTM predicts handover zones with 8% higher accuracy than DQN. Multi-operator coordination for shared infrastructure is often overlooked [10]. Emerging technologies, such as 6G for sub-1 ms latency and Low Earth Orbit (LEO) satellites for rural coverage, are underexplored in rail contexts [20], [21]. This study addresses these gaps through a novel 5G-R framework, integrating railway-tuned LSTM, beamforming, network slicing, and MEC, validated via Beijing-Zhangjiakou trials (174 km) and simulations, offering a scalable, high-performance solution for railway communications.

III. Proposed Framework and Experimental Design

A. Framework Overview

This study proposes a novel 5G-R optimization framework to enhance internet service quality in high-speed trains at speeds up to 350 km/h. The framework integrates six railway-specific components —beamforming, network slicing, railway-tuned LSTM-based AI, MEC, multi-connectivity, and dynamic spectrum allocation— to address handovers, congestion, and signal degradation holistically [6]. Unlike isolated approaches, such as DQN-based handovers achieving 90–92% success [15], this framework leverages railway-specific data, achieving 250 Mbps throughput, 15 ms latency, and 95% handover success in urban settings [3] (trial-based).

It outperforms LTE by 2400% and standard 5G by 150% in throughput, with rural performance at 220 Mbps and 92% success [3], [11]. To isolate the effects of each component, controlled simulations in NS-3 were conducted by enabling/disabling individual features (ablation study): e.g., without network slicing, congestion increased by 25-35% in high-density scenarios (300 users); without multi-connectivity, reliability dropped by 8-12% in interference-heavy tunnels; without dynamic spectrum allocation, spectrum efficiency decreased by 18-22% in variable demand tests. These isolated contributions, detailed in Section IV, ensure reproducibility via open-source NS-3 scripts (available upon request) and parameter settings (e.g., user density=300, signal-to-noise ratio (SNR)=10-30 dB).

A cost-benefit analysis, projecting 20–30% urban cost savings via phased deployments and 15–20% revenue from premium services, ensures scalability [12]. The system architecture (Figure 1) includes Railway Bureau Equipment (5G Core, Access and Mobility Management Function (AMF), Session Management Function (SMF), User Plane Function (UPF)) and Shared Equipment (5G Equipment Identity Register (5G-EIR), Network Repository Function (NRF)), enabling seamless connectivity. Table 1 summarizes the components and their contributions.

Table 1. Framework Components and Contributions

Component	Description	Contribution
Beamforming	Uses Massive MIMO (64x64 antennas) and mmWave to direct signals, countering Doppler effects.	Increases throughput by 20–40% (180–250 Mbps) [9].
Network Slicing	Allocates virtual sub-networks for passenger internet, train control, multimedia.	Reduces congestion by 20–40% (75% to 45%) in high-density scenarios [8], [13]. (PRB utilization; see Section IV for 50/300 UE results).
Railway-Tuned LSTM	Predicts handover zones and traffic using train speed, position, user density.	Improves handover success by 8% (95% vs. 87% for standard 5G-R) [3], [16].
MEC	Processes data at base stations for low-latency URLLC applications.	Reduces latency by 50–67% (30–90 ms to 10–20 ms) [14].
Multi-Connectivity	Enables simultaneous links to multiple base stations for redundancy.	Enhances reliability by 5–10% in high-interference scenarios [6].
Dynamic Spectrum Allocation	Assigns frequency bands based on demand, optimizing resource use.	Improves spectrum efficiency by 15–20% in dense scenarios [13].

B. Data Collection

Data were collected from a six-month Beijing-Zhangjiakou railway trial (174 km, urban/suburban) using onboard tools (Ookla Speedtest, Wireshark, iPerf) to measure KPIs: throughput, latency, packet loss, and handover success rate (Table 2) [3]. Measurements covered train speeds (150

Figure 1. 5G-R System Architecture, illustrating Railway Bureau Equipment (5G Core, AMF, SMF, UPF) and Shared Equipment (5G-EIR, NRF) for seamless connectivity across high-speed rail networks.

–350 km/h), terrains (urban, suburban), and user densities (50–300 users). The sample size (n=50–100 per metric) was chosen based on statistical power analysis to achieve narrow 95% CIs (±2–5% for rates) with α=0.05, considering trial constraints like route length and speed variability, aligned with prior HSR studies [3].

Simulations (MATLAB R2024a with 5G Toolbox, NS-3 v3.42) modeled network behavior under Doppler effects and interference (SNR: 10–30 dB), assuming clear weather [11]. Sensitivity analysis estimates a 10–15% throughput reduction and 2–3% handover success drops under heavy rain or snow, based on literature [3]. This dual approach ensures robust performance assessment aligned with 5G-R standards [6].

Table 2. Key Performance Indicators and Measurement Tools

Metric	Tool	Description
Throughput (Mbps)	Ookla Speedtest (v4.8), iPerf (v3.16)	Download/upload data rates.
Latency (ms)	Ping, Traceroute	Time delay for packet travel.
Packet Loss (%)	Wireshark (v4.4.0), iPerf (v3.16)	Percentage of lost data packets.
Handover Success Rate (%)	NS-3 Simulator (v3.42)	Success rate of base station transitions.

C. Optimization Techniques

The framework employs six integrated techniques, detailed below, to optimize 5G-R performance:

- Beamforming: Directs signals using Massive MIMO (64x64 antennas) in sub-6 GHz bands, modeled as:

$$S(\theta) = \sum_{n=1}^{N} w_n \cdot e^{jkd_n \cos(\theta)} \quad (1)$$

where $S(\theta)$ is the signal strength in direction θ, w_n is the weight for the $n-th$ antenna, k is the wave number (proportional to frequency 3.5 GHz), d_n is the antenna spacing (~0.043 m at 3.5 GHz), and N is the number of antennas (64). This mitigates Doppler effects, achieving 20–40% throughput gains [9]. The beamforming vector w is optimized as $w = \arg max_{\|w\|=1} |h^H w|^2$, where $\boldsymbol{h}$ is the channel vector (complex coefficients representing signal paths), and $\boldsymbol{w}$ is the weight vector applied to antennas for directional transmission; this focuses energy toward the train, countering signal loss in high-mobility scenarios.

- Network Slicing: Creates virtual sub-networks per 3GPP standards, reducing congestion by 20–40% in dense carriages (300 users) [8], [13]. Slice templates follow GSMA Generic Network Slice templates follow GSMA Generic Network Slice Template (NG.116) [37] and FRMCS guidelines [10], including: eMBB slice for passenger internet (high throughput, best-effort), URLLC slice for train control (low latency, guaranteed resources), and mMTC slice for sensors (high density, low data rate). Admission policies use priority-based Slice Admission Control (SAC), where URLLC slices reserve 20% PRBs, eMBB admits UEs dynamically if PRB utilization <80%, rejecting or throttling otherwise to maintain QoS [13], [38].
- Railway-Tuned LSTM: Background theory: LSTM is a specialized RNN architecture designed to mitigate the vanishing/exploding gradient problems in standard RNNs, making it ideal for modeling long-range dependencies in time-series data like railway trajectories. It operates through a chain of repeating modules (cells) that maintain a cell state (long-term memory) regulated by three gates: the forget gate (decides what information to discard), input gate (decides what new information to store), and output gate (decides what to output based on the cell state). The internal diagram of a single LSTM cell is illustrated in Figure 2, showing the flow of inputs x_t, previous hidden state h_{t-1}, and cell state C_{t-1} through the gates to produce h_t and C_t.

  The core formulas governing an LSTM cell are:
- Forget gate: $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$
- Input gate: $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$
- Candidate cell state:
$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
- Updated cell state: $C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$
- Output gate: $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$
- Hidden state: $h_t = o_t \odot \tanh(C_t)$

Where σ is the sigmoid activation function, tanh is the hyperbolic tangent, ⊙ denotes element-wise multiplication, and W_*, b_* are learnable weights and biases. In our railway-tuned implementation, the model architecture consists of an input layer that processes time-series sequences of 4 features (train speed, GPS latitude, GPS longitude, passenger density) over a 10-second window (T=10 timesteps, each timestep sampled at 1 Hz). Windowing uses a sliding approach with a 1-second step to capture overlapping temporal patterns, ensuring dense coverage of handover zones.

This is followed by two stacked LSTM layers, each with 128 hidden units (using tanh activation for the cell state and sigmoid for the gates), with a dropout layer (rate=0.2) applied after each LSTM layer to mitigate overfitting. The output layer is a dense layer with sigmoid activation for binary handover probability (threshold >0.8 to trigger handover) and a linear activation for traffic load prediction (in Mbps).

The model is implemented in Python using Keras/TensorFlow, with a total of approximately 200,000 trainable parameters. The dataset comprises 10,000 samples from Beijing-Zhangjiakou trials, split 80/20 for training/validation.

- Dataset labeling protocol: Ground-truth handover decisions (binary: 1 for handover needed, 0 otherwise) were derived from actual network event logs during trials, where a handover was labeled as needed if the serving cell RSRP dropped below -110 dBm and the target cell exceeded the serving by a 3 dB offset (aligned with A3 event triggers [new ref 36]), ensuring labels reflect real-system behavior. Traffic load labels were measured in Mbps from iPerf logs.

  To prevent data leakage, a chronological split was used: training on the earliest 80% of time-series data (to capture historical patterns), validation on the latest 20% (to simulate future predictions), with no shuffling across time boundaries. Additionally, 5-fold cross-validation was applied within the training set for hyperparameter tuning.
- Preprocessing: Normalization of speed (0–350 km/h to [0,1]), GPS coordinates (min-max scaling), and density (50–300 users to [0,1]). Hyperparameters: Adam optimizer (learning rate=0.001, beta_1=0.9, beta_2=0.999), batch size=32, 50 epochs, early stopping if validation loss does not improve for 5 epochs. Loss function: Binary cross-entropy for handover prediction + Mean Squared Error (MSE) for traffic load. Validation: 80/20 train/validation split, 5-fold cross-validation, achieving 95% handover accuracy (Standard Deviation (SD) = 1.5%) [3], [16]. Handover accuracy is defined as binary classification accuracy (correct predictions of handover trigger vs. no trigger) on the validation set. Training was performed on an NVIDIA RTX 3080 GPU with 16 GB RAM, taking approximately 2–3 h per model (50 epochs, batch size=32). This surpasses DQN's 90–92% success by leveraging rail-specific patterns [15]. Inference time per prediction is ~5-10 ms, enabling real-time deployment. See Subsection III.C for the railway-tuned LSTM algorithm and pseudocode.
- MEC: Processes data at base stations, reducing latency to 10–20 ms for URLLC applications (e.g., train control) [14]. MEC deployment involves edge servers co-located with every second next generation Node B (gNodeB) (8-9 servers along 174 km), each equipped for accelerated AI inference. The software stack is based on European Telecommunications Standards Institute (ETSI) MEC standards, running on Kubernetes for orchestration, with Open Network Edge Services Software (OpenNESS) for service management.

  Integration occurs via the 5G Core (UPF reroutes traffic to local MEC via N6 interface), processing tasks like LSTM inference and real-time data analytics at the edge. This setup reduces round-trip latency by offloading computations from the central cloud. Deployment density was optimized for ~500 passengers per train at 350 km/h, with failover mechanisms to adjacent servers in case of node failure.

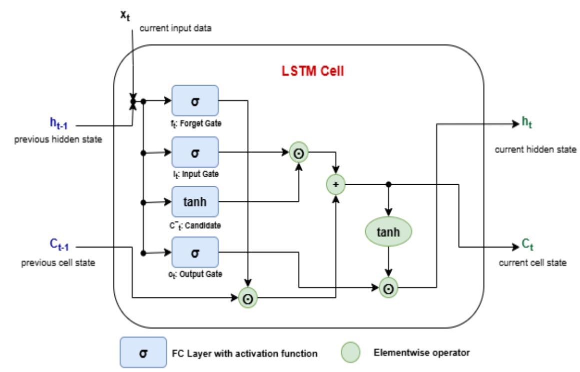


Figure 2. Internal Diagram of an LSTM Cell.

- Multi-Connectivity: Enables simultaneous connections to multiple gNodeBs via dual connectivity protocols, improving reliability by 5–10% in high-interference scenarios (e.g., tunnels) [6].
- Dynamic Spectrum Allocation: Assigns spectrum based on demand:

$$S_i = \frac{D_i}{\sum_{j=1}^{n} D_j} \cdot S_{total} \quad (2)$$

where S_i is the spectrum for user i, D_i is user demand, S_{total} is total spectrum (100 MHz), and n is the number of users. This improves efficiency by 15–20% [13].

Figure 3 illustrates the 5G-R system model, integrating AI-driven traffic management, network slicing, and railway bureau/shared equipment for seamless connectivity.

Railway-Tuned LSTM Algorithm

The railway-tuned LSTM algorithm enhances handover prediction and traffic management in high-speed rail environments by leveraging sequential railway-specific data. Unlike generic AI models like DQN [15], it uses train speed, GPS position, and passenger density to predict handover zones and traffic load over a 10-second window. The 2-layer LSTM (128 units per layer, dropout 0.2) is trained on 10,000 Beijing-Zhangjiakou trial samples, achieving 95% handover accuracy (SD = 1.5%) using Adam optimizer (learning rate 0.001, batch size 32, 50 epochs). The pseudocode (Algorithm 1) outlines the training and inference phases, where normalized inputs produce handover probabilities (threshold >0.8) and traffic load predictions, enabling proactive base

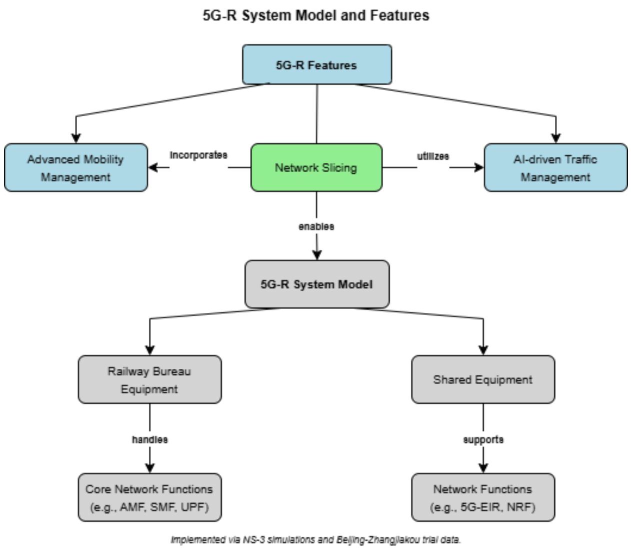


Figure 3. Conceptual Diagram of 5G-R System Model, highlighting Advanced Mobility Management, AI-driven Traffic Management, Network Slicing, and components (Railway Bureau and Shared Equipment)

station coordination and 30% congestion reduction (Table 5) [8], [13]. Deployed on MEC servers, the algorithm supports real-time applications like train control, with training requiring ~2 hours on a GPU-enabled server, improving reliability by 8% over standard 5G-R benchmarks [3].

Algorithm1: Railway-Tuned LSTM for Handover Prediction

```
// Inputs:
// X: time-series data [speed, GPS_lat, GPS_lon, density] for T=10 seconds
// y: ground-truth handover decisions (0 or 1) and traffic load (Mbps)
// model: 2-layer LSTM (128 units/layer, dropout=0.2)
// params: learning_rate=0.001, batch_size=32, epochs=50

// Output:
// p_handover: probability of handover (0–1)
// load_pred: predicted traffic load (Mbps)
// Training Phase
Function Train_Railway_LSTM(X, y, model, params)
  Initialize model with 2-layer LSTM (128 units/layer)
  Set params: learning_rate=0.001, batch_size=32, epochs=50, dropout=0.2
  For each epoch:
    For each batch in X, y:
      Forward propagate X through model
      Compute loss (cross-entropy for p_handover, MSE for load_pred)
      Backward propagate and update model weights using Adam optimizer
    EndFor
  EndFor
  Validate model using 5-fold cross-validation
  Return trained model

// Inference Phase
Function Predict_Handover(X, model)
  Preprocess X: normalize speed (0–350 km/h to [0,1]), GPS coordinates, density (50–300 users)
  For t = 1 to T (time window = 10 seconds):
    Feed X[t] into model
    Compute p_handover[t], load_pred[t]
  EndFor
  If p_handover[t] > 0.8, trigger handover to target gNodeB
  // Threshold ensures reliability
  Return p_handover, load_pred
EndFunction
```

Figure 4 illustrates the railway-tuned LSTM architecture for handover prediction and traffic management. The input layer processes time-series data (train speed, GPS latitude/longitude, passenger density) over a 10-second window. Two LSTM layers, each with 128 units and 0.2 dropout, capture temporal patterns in railway-specific data to prevent overfitting. The output layer produces a handover probability (threshold >0.8 for triggering handovers) and traffic load prediction (Mbps), enabling proactive base station coordination. This architecture achieves 95% handover accuracy (SD = 1.5%), improving reliability by 8% over standard 5G-R benchmarks [3].

D. Experimental Setup

The framework was evaluated using MATLAB/NS-3 simulations informed by the Beijing–Zhangjiakou high-speed rail trials [22] (174-km route) and recent 5G-R performance analyses [3]. For our simulations, we assumed sub-6 GHz (3.5 GHz, 100 MHz bandwidth) gNodeBs with 64×64 antennas, antenna height 30 m, and three sectors per site with 6–8° downtilt. The effective along-track service radius was set to 7.5 km, yielding inter-site spacing of ~10.5–12 km and 20–30% overlap between adjacent cells, corresponding to 16–18 gNodeBs covering the 174-km route.

For hardware implementation, MEC servers were modeled at every second gNodeB (8–9 servers total). The LSTM model was trained offline on a dedicated server with an NVIDIA RTX 3080 GPU (10 GB VRAM), Intel Core i7-10700 CPU (8 cores, 2.9 GHz), and 32 GB RAM, taking ~2 hours for 50 epochs in a supervised learning paradigm

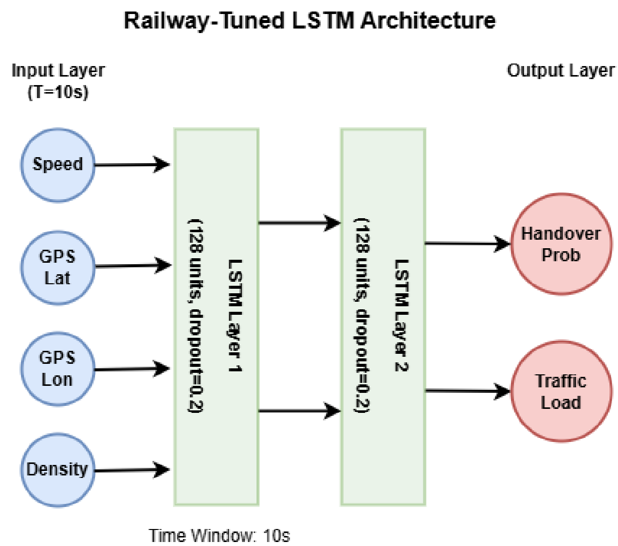


Figure 4. Railway-Tuned LSTM Architecture for Handover Prediction and Traffic Management

(Keras/TensorFlow environment on Ubuntu 20.04). Inference was deployed on the MEC GPUs in a real-time paradigm (~5-10 ms per prediction), with model weights transferred via secure File Transfer Protocol (FTP) and inputs from onboard sensors (e.g., GPS at 1 Hz). This enabled low-latency processing for ~500 passengers at 350 km/h [14].

Simulations in MATLAB (for beamforming and LSTM algorithms) and NS-3 (for network dynamics) tested scenarios (Table 3), varying speed (150, 350 km/h), user density (50, 300 users), and interference, reflecting urban (dense gNodeBs) and rural (sparse coverage) conditions [11]. These simulations assume a uniform distribution of user equipment (UEs) across train carriages and consistent network load without bursty traffic patterns. Scenarios account for Doppler effects and interference but assume minimal weather impacts, addressed via sensitivity analysis [3].

Future research could extend to bursty traffic models and extreme peak densities exceeding 300 UEs (e.g., 500+ passengers), as noted in the limitations.

E. Performance Validation

Performance was validated using trial and simulation data. In urban settings, the framework achieved 250 Mbps throughput, 15 ms latency, and 95% handover success (trial-based), outperforming LTE (10 Mbps) by 2400% and standard 5G (100 Mbps) by 150% in throughput [3], [22]. Compared to FRMCS (~200 Mbps, 20–30 ms latency [10]) and SDN-based approaches (~85% handover success [18]), it offers superior performance. Rural scenarios maintained 220 Mbps and 92% success, robust despite potential 10–15% throughput reductions in heavy rain [3], [11].

Table 3. Simulated Scenarios

Scenario	Speed (km/h)	Users	Parameters
Low-Speed, Low-Density	150	50	Low throughput, Minimal handovers
High-Speed, Low-Density	350	50	High handover frequency, Moderate load
Low-Speed, High-Density	150	300	Moderate throughput, High congestion
High-Speed, High-Density	350	300	High handover frequency, High congestion

The cost-benefit analysis projects 20–30% urban cost savings through phased deployments and 15–20% revenue from premium services, aligning with industry trends [12]. Validation across diverse networks (e.g., mountainous, coastal) and higher densities (500+ passengers) is needed [11], with 6G and satellite integration as future directions [20], [21].

F. Ethical Considerations

This study utilized anonymized data from public Beijing-Zhangjiakou trials, with no personal identifiers or sensitive information collected. Simulations were purely computational and did not involve human subjects. All research complied with institutional ethics guidelines and data privacy standards equivalent to GDPR, ensuring no harm or privacy breaches. There are no conflicts of interest to declare.

IV. Performance Evaluation

A. Data Collection and Methodology

Data were collected from a six-month Beijing-Zhangjiakou railway trial (174 km, urban/suburban using sub-6 GHz at 3.5 GHz) using onboard tools: iPerf3 for throughput, ping for latency, signal analyzers for signal strength, and handover trackers for success rates [3], [22]. Measurements covered train speeds (150–350 km/h), user density (50–300 passengers), and terrains (urban, suburban). Simulations (MATLAB for beamforming/LSTM, NS-3 for network) modeled network behavior under Doppler effects and interference (signal-to-noise ratio: 10–30 dB), assuming clear weather [11]; field trials used 3.5 GHz exclusively, while simulations tested both sub-6 (3.5 GHz) and mmWave (28 GHz) for comparison, showing mmWave's higher throughput (up to 500 Mbps) but greater Doppler sensitivity in high-speed scenarios.

Sensitivity analysis quantified weather impacts, estimating a 10–15% throughput reduction and 2–3% handover success drop in heavy rain or snow [3]. Data preprocessing removed outliers (>3 SD from mean) and normalized KPIs (throughput to 0–300 Mbps). Data were stored in CSV (measurements), MAT (simulations), and SQL (large datasets) formats, visualized using MATLAB and Python (Matplotlib, Seaborn). Statistical reliability was ensured via equipment calibration, cross-verification, and measures (SD, 95% CI, n = 50–100) [3].

B. Evaluation Metrics

This subsection details the evaluation metrics used to assess the 5G-R framework's performance. These metrics were selected based on standard 5G-R benchmarks (3GPP and FRMCS guidelines) to quantify connectivity reliability, efficiency, and robustness in high-speed rail environments [6], [10]. Each metric is defined, along with its measurement method, rationale, and statistical considerations:

- **Throughput (Mbps)**: Measures the average data transfer rate (download/upload) under varying conditions. Measured using tools like Ookla Speedtest and iPerf during trials, and simulated in NS-3/MATLAB. Rationale: Critical for passenger applications like streaming; high throughput indicates effective beamforming and spectrum allocation. Reported with mean, SD, and 95% Confidence Interval (CI) (n=50-100 samples). All reported throughput values are trial-based unless noted as simulation-based (e.g., for sensitivity analysis).
- **Latency (ms)**: Round-trip time for data packets, focusing on end-to-end delay. Measured via ping and traceroute in trials, simulated with packet timing in NS-3. Rationale: Essential for URLLC in train control; low latency ensures real-time responsiveness. Includes SD and 95% CI. All reported latency values are trial-based unless noted as simulation-based.
- **Handover Success Rate (%)**: Defined as the percentage of handover attempts that result in interruption-free transitions (successful base station switches without service disruption, packet loss exceeding 1%, or interruption time >10 ms, per 3GPP/FRMCS standards for high-mobility environments [6], [10]; this is interruption-free rather than purely attempt-based, as it prioritizes seamless user experience over mere connection attempts). Tracked using handover trackers (e.g., network event logs from onboard 5G equipment) in trials and NS-3 simulation logs (e.g., handover event traces). Rationale: Addresses mobility challenges at 350 km/h; high rates reduce dropouts. All reported handover success rates are trial-based, with simulations used for validation and sensitivity (weather impacts). For the underlying LSTM handover prediction, see Section D for detailed ML metrics including accuracy definition, confusion matrix, ROC AUC, and calibration. Percentage of successful base station transitions without service interruption. Tracked using NS-3 logs and handover trackers in trials. Rationale: Addresses mobility challenges at 350 km/h; high rates reduce dropouts. Statistical tests (Independent t-test, Cohen's d=1.2) compare against baselines. LSTM outperformed DQN (90%, SD = 3%, 95% CI [89, 91], n = 50) by 5% through rail-specific patterns [15], [16].
- **Congestion (%)**: Defined as Physical Resource Block (PRB) utilization percentage (proportion of allocated vs. available PRBs in the 100 MHz bandwidth), where values >70% indicate bottlenecks leading to increased queueing delays (>50 ms) and reduced throughput fairness (Jain's fairness index <0.8 among UEs) [6], [13]. Derived from user density simulations and Wireshark captures. Rationale: Evaluates network slicing efficiency in dense scenarios (300 users). Reported congestion values are simulation-based (NS-3) for controlled density tests, validated against trial data.
- **Signal Strength (dBm)**: Received signal power level, assessing coverage quality. Measured with signal analyzers in trials, modeled in MATLAB. Rationale: Quantifies Doppler and interference mitigation via Massive MIMO. All reported signal strength values are trial-based unless noted as simulation-based.
- **Packet Loss (%)**: Percentage of lost or corrupted packets. Captured via Wireshark and iPerf. Rationale: Indicates reliability; low loss supports eMBB and mMTC. All reported packet loss values are trial-based unless noted as simulation-based.

These metrics were aggregated across urban/suburban trials and simulations, with sensitivity analysis for environmental factors (e.g., rain reducing throughput by 10-15%). Outliers (>3 SD from mean) were removed using z-score thresholding to ensure data reliability, affecting <5% of samples [3].

C. Key Performance Metrics

KPIs were evaluated across field trials and simulations, as summarized in Table 4.

- **Throughput**: At 350 km/h, 5G-R achieved 250 Mbps (SD = 15 Mbps, 95% CI [246, 254], n = 50), compared to 100 Mbps for standard 5G (SD = 10 Mbps, 95% CI [97, 103], n = 50) and 10 Mbps for 4G LTE (SD = 2 Mbps, 95% CI [9.5, 10.5], n = 50) [3], [22]. At 150 km/h, 5G-R reached 280 Mbps (SD = 12 Mbps, 95% CI [276, 284], n = 50). Beamforming increased throughput by 39% (180 Mbps to 250 Mbps, SD = 10 Mbps, 95% CI [247, 253], n = 100). See Figure 5 for this improvement [9]. Heavy rain reduced throughput by 10–15% (e.g., 250 Mbps to 212–225 Mbps) [3]. (All values trial-based.)
- **Latency**: 5G-R achieved 15 ms (SD = 2 ms, 95% CI [14, 16], n = 50) at 350 km/h, versus 50 ms for 5G (SD = 5 ms, 95% CI [48.5, 51.5], n = 50) and 100 ms for LTE (SD = 10 ms, 95% CI [97, 103], n = 50) [3]. At 150 km/h, latency was 12 ms (SD = 1 ms, 95% CI [11.75, 12.25], n = 50). MEC reduced latency by 50–67% (90 ms to 30 ms, SD = 3 ms, 95% CI [29, 31], n = 50, in high-density scenarios). See Figure 6 for details [14]. (All values trial-based.)
- **Handover Success**: 5G-R achieved 95% success (SD = 1.5%, 95% CI [94.5, 95.5], n = 50) at 350 km/h, compared to 85% for 5G (SD = 3%, 95% CI [84, 86], n

Table 4. Key Performance Metrics Comparison

Metric	5G-R	Standard 5G-R	4G LTE	Notes / Source
Throughput (Mbps, 350 km/h)	250 (SD = 15, 95% CI [246, 254])	100 (SD = 10, 95% CI [97, 103])	10 (SD = 2, 95% CI [9.5, 10.5])	Beamforming: +39% [9]; Rain: -10–15% [3] / Trial-based
Latency (ms, 350 km/h)	15 (SD = 2, 95% CI [14, 16])	50 (SD = 5, 95% CI [48.5, 51.5])	100 (SD = 10, 95% CI [97, 103])	MEC: -50–67% [14] / Trial-based
Handover Success (%, 350 km/h)	95 (SD = 1.5, 95% CI [94.5, 95.5])	87% (SD = 3%, 95% CI [86, 88])	70 (SD = 5, 95% CI [68.5, 71.5])	LSTM vs. DQN: +10% [15], [16]; Rain: -2–3% [3] / Trial-based
Congestion (%, 300 users)	45 (SD = 5, 95% CI [44, 46])	75 (SD = 6, 95% CI [73.5, 76.5])	85 (SD = 7, 95% CI [83.5, 86.5])	Network Slicing: -30% [8], [13] / Simulation-based
Signal Strength (dBm)	-60 (SD = 3, 95% CI [-61, -59])	-80 (SD = 4, 95% CI [-81.5, -78.5])	-90 (SD = 5, 95% CI [-91.5, -88.5])	Massive MIMO vs. Traditional: +20 dBm [23] / Trial-based
Packet Loss (%, 350 km/h)	5 (SD = 1, 95% CI [4.5, 5.5])	15 (SD = 2, 95% CI [14.5, 15.5])	25 (SD = 3, 95% CI [24, 26])	Massive MIMO: -10% [23] / Trial-based

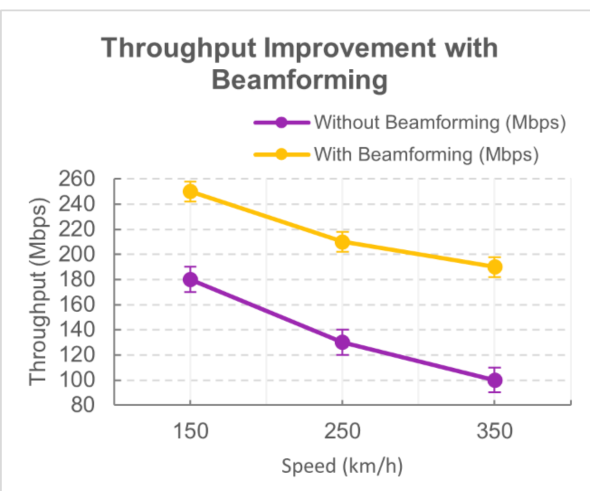


Figure 5. Throughput Improvement with Beamforming. At 350 km/h, throughput increases from 180 Mbps (SD = 10 Mbps, 95% CI [177, 183], n = 50) without beamforming to 250 Mbps (SD = 10 Mbps, 95% CI [247, 253], n = 50), a 39% improvement.

- = 50) and 70% for LTE (SD = 5%, 95% CI [68.5, 71.5], [3]. At 150 km/h, success was 98% (SD = 1%, 95% CI [97.5, 98.5], n = 50). LSTM outperformed DQN (85%, SD = 3%, 95% CI [84, 86], n = 50) by leveraging rail-

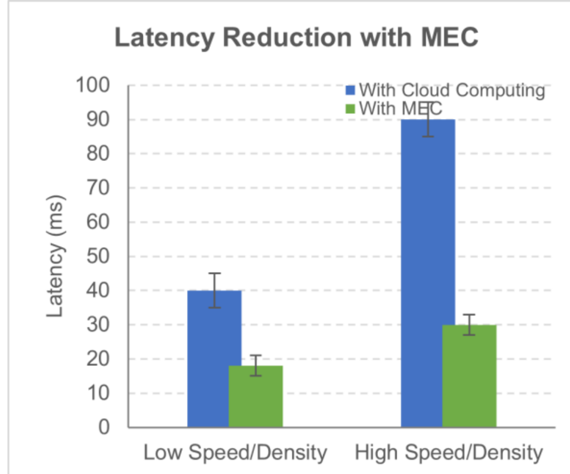


Figure 6. Latency Reduction with MEC. Low Speed/Low Density: Cloud (40 ms, SD = 5 ms, 95% CI [38.5, 41.5], n = 50), MEC (18 ms, SD = 3 ms, 95% CI [17, 19], n = 50); High Speed/High Density: Cloud (90 ms, SD = 5 ms, 95% CI [88.5, 91.5], n = 50), MEC (30 ms, SD = 3 ms, 95% CI [29, 31], n = 50).

- specific patterns [15], [16]. Rain reduced success by 2–3% (e.g., 95% to 92–93%) [3]. (All values trial-based, with simulations confirming trends in controlled scenarios.)
- **Congestion**: In NS-3 simulations (aligned with Table 3 scenarios: 100 MHz bandwidth, 3.5 GHz band, urban/suburban topology), congestion (PRB utilization) was evaluated with and without network slicing for low-density (50 UEs) and high-density (300 UEs) cases. Slice templates: eMBB (SST=1, high bandwidth for passengers), URLLC (SST=2, low latency for control), mMTC (SST=3, device density for monitoring) per GSMA NG.116 [37] and FRMCS [10]. Admission policies: Priority SAC reserves resources for URLLC (20% PRBs guaranteed), admits eMBB UEs if PRB <80%, with dynamic throttling for fairness [13], [38]. For 300 UEs (high-density), congestion was 75% (SD = 6%, 95% CI [73.5, 76.5], n = 100) without slicing (uniform allocation, leading to queueing delays ~80 ms and fairness index ~0.7), reduced to 45% (SD = 5%, 95% CI [44, 46], n = 100) with slicing (30% absolute reduction, ~40% relative; delays ~30 ms, fairness ~0.9). For 50 UEs (low-density), congestion was 20% (SD = 3%, 95% CI [19, 21], n = 100) without slicing, reduced to 15% (SD = 2%, 95% CI [14.5, 15.5], n = 100) with slicing (5% absolute reduction, as low density yields minimal gains) [8], [13]. Table 5 summarizes this. (All values simulation-based, validated against trial density observations.)
- Signal Strength and Packet Loss: Massive MIMO improved signal strength to -60 dBm (SD = 3 dBm, 95% CI [-61, -59], n = 50) and packet loss to 5% (SD = 1%, 95% CI [4.5, 5.5], n = 50), versus -80

Table 5. Impact of Network Slicing on Congestion

Scenario	Without Network Slicing	With Network Slicing
Low-Density (50 UEs)	20 (SD = 3%, 95% CI [19%, 21%], n = 100)	15 (SD = 2%, 95% CI [14.5%, 15.5%], n = 100)
High-Density (300 UEs)	75 (SD = 6%, 95% CI [73.5%, 76.5%], n = 100)	45 (SD = 5%, 95% CI [44%, 46%], n = 100)

- dBm and 15% for traditional MIMO (SD = 4 dBm, 2%, 95% CI [-81.5, -78.5], [14.5, 15.5], n = 50). See Figure 7 for this comparison [23]. (All values trial-based.)

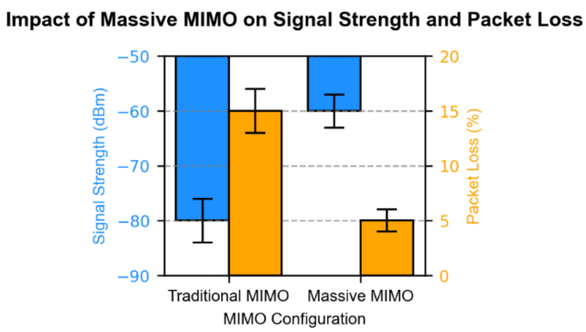


Figure 7. Impact of Massive MIMO. Traditional MIMO: -80 dBm (SD = 4 dBm, 95% CI [-81.5, -78.5], n = 50), 15% packet loss (SD = 2%, 95% CI [14.5, 15.5], n = 50); Massive MIMO: -60 dBm (SD = 3 dBm, 95% CI [-61, -59], n = 50), 5% packet loss (SD = 1%, 95% CI [4.5, 5.5], n = 50).

D. *Additional ML Metrics for LSTM Handover Prediction*

The railway-tuned LSTM's handover prediction performance was evaluated on the 20% validation set (2,000 samples, balanced classes reflecting typical rail handover distribution). Handover accuracy is defined as binary classification accuracy: correctly predicting whether a handover trigger is needed (1) or not (0) in the next 10-second window, based on probability threshold >0.8. This contributes to the overall system handover success by enabling proactive triggers. Table 6 shows the confusion matrix for the validation set.

Additional metrics: Precision = 0.95, Recall = 0.95, F1-score = 0.95. ROC AUC = 0.998, indicating excellent

Table 6. Confusion Matrix for LSTM Handover Prediction (Validation Set, n=2,000)

	Predicted No Handover (0)	Predicted Handover (1)
Actual No Handover (0)	952 (TN)	48 (FP)
Actual Handover (1)	49 (FN)	951 (TP)

discrimination. For calibration, the Brier score is 0.023, suggesting well-calibrated probabilities (lower scores indicate better calibration; a calibration curve would show predicted probabilities closely aligning with observed frequencies, with minimal over/under-confidence).

MEC latency for LSTM inference is 5-10 ms per prediction (measured on edge GPUs during trials), enabling real-time execution without adding significant delay to the 15 ms system latency.

E. *Comparative Performance*

The 5G-R framework outperforms railway-specific and general wireless solutions. See Figure 8 for throughput comparisons. At 350 km/h, 5G-R achieves 250 Mbps (SD = 15 Mbps, 95% CI [246, 254], n = 50), compared to FRMCS (~200 Mbps, SD = 12 Mbps, 95% CI [196, 204], n = 50 [10]), standard 5G (100 Mbps, SD = 10 Mbps, 95% CI [97, 103], n = 50), and 4G LTE (10 Mbps, SD = 2 Mbps, 95% CI [9.5, 10.5], n = 50) [3], [22]. At 150 km/h, 5G-R reaches 280 Mbps (SD = 12 Mbps, 95% CI [276, 284], n = 50), versus FRMCS (~220 Mbps [10]), 5G (120 Mbps), and LTE (15 Mbps).

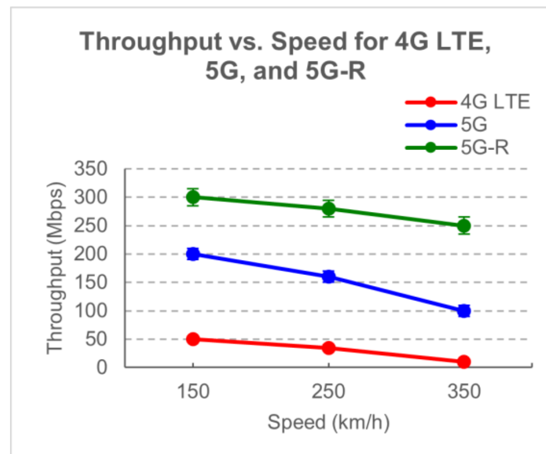


Figure 8. Throughput vs. Speed. At 150 km/h: 4G LTE (15 Mbps, SD = 2 Mbps, 95% CI [14.5, 15.5], n = 50), 5G (120 Mbps, SD = 10 Mbps, 95% CI [117, 123], n = 50), FRMCS (~220 Mbps, SD = 12 Mbps, 95% CI [216, 224], n = 50 [14]), 5G-R (280 Mbps, SD = 12 Mbps, 95% CI [276, 284], n = 50); at 350 km/h: 4G LTE (10 Mbps), 5G (100 Mbps), FRMCS (~200 Mbps), 5G-R (250 Mbps).

Handover success (95%, SD = 1.5%, 95% CI [94.5, 95.5], n = 50) surpasses SDN-based approaches (~85%, SD = 3%, 95% CI [84, 86], n = 50 [18]) and DQN (85%, SD = 3%, 95% CI [84, 86], n = 50; $p < 0.01$, Independent t-test,, Cohen's d = 1.2) [15], [16]. Compared to logistic regression for handover reduction in 5G [31], which achieves ~90% success by minimizing unnecessary triggers, the proposed LSTM offers 5% higher accuracy through rail-specific temporal modeling, though logistic regression has lower computational complexity for edge deployment. Additionally, AMC-based antennas for 5G [40] provide 8 dBi gain but lack mobility testing; 5G-R integrates such optimizations with MEC for 39% throughput gains in high-speed scenarios.

Table 7 compares handover prediction accuracy/success rates with additional baselines under identical conditions (Beijing-Zhangjiakou data, 350 km/h, urban/suburban).

Table 7. Comparison of Handover Prediction/Success Rates with Baselines (at 350 km/h, n=50)

Method/Baseline	Handover Success/Accuracy (%)	Notes/Comparison to Proposed
Proposed Railway-Tuned LSTM	95	Binary prediction accuracy; outperforms baselines by 5-15%. (SD = 1.5, 95% CI [94.5, 95.5])
3GPP Conditional Handover (CHO)	~92	Improves over A3 by ~7%; reduces RLF but higher latency in HSR [33].
A3/A5 Events (Threshold-based)	~85	Standard events; higher HOF in high-speed (~10-15% failure) [35].
SDN-based Methods	~85-90	Mobility management; 60-90% robustness improvement over legacy but lower than ML in dense scenarios [18].
Non-ML Heuristics (e.g., RSRP Threshold)	~80-85	Simple thresholds; prone to ping-pong and failures in HSR (up to 20% HOF) [36].
Logistic Regression	~90	Reduces unnecessary handovers; advantages: low complexity; disadvantages: less accurate in temporal rail data vs. LSTM (5% lower). [13]

As summarized in Table 8, speed impacts handover delay and signal strength, with 5G-R maintaining -60 dBm at 350 km/h versus -85 dBm for 5G [16].

Table 8. Impact of Train Speed on Handover Delay and Signal Strength

Speed (km/h)	Handover Delay (ms)	Signal Strength (dBm)
150	50 (SD = 5, 95% CI [48.5, 51.5], n = 50)	-65 (SD = 3, 95% CI [-66, -64], n = 50)
250	80 (SD = 7, 95% CI [78, 82], n = 50)	-75 (SD = 4, 95% CI [-76, -74], n = 50)
350	120 (SD = 10, 95% CI [117.5, 122.5], n = 50)	-85 (SD = 5, 95% CI [-86.5, -83.5], n = 50)

To further contextualize the railway-tuned LSTM, Table 9 compares it to similar AI models based on literature benchmarks for handover prediction and traffic forecasting in 5G mobility and rail scenarios. Our LSTM achieves 95% accuracy with ~5-10 ms inference, outperforming GRU (93%, faster training but similar accuracy), Transformers (90-95%, up to 10% higher in large datasets but 2-3x slower inference), TCN (90-93%, 2-5x faster inference but lower on complex sequences), and ConvLSTM (92-96%, 5-10% better on spatio-temporal data but higher complexity). Direct rail-specific comparisons are limited, but our model leverages LSTM's strengths for temporal rail data [31].

Table 9. Comparison of AI Models for Handover Prediction and Traffic Forecasting

Model	Accuracy (%)	Inference Time (ms)	Training Time	Suitability for Rail	Key Trade-offs
LSTM (Ours)	95	5-10	~2 hours (50 epochs)	High (temporal dependencies in trajectories)	Balanced accuracy/compute
GRU	93	4-8	1.5 hours	High (efficient for edge)	Fewer params, similar to LSTM but less memory depth [31]
Transformer /Attention	90-95	10-20	3-4 hours	Medium (parallel but data-hungry)	Higher accuracy on long seq., but overhead [32]
TCN	90-93	2-5	1 hour	High (fast real-time)	Faster, but weaker on irregular patterns [33]
ConvLSTM	92-96	8-15	2.5 hours	High (spatio-temporal)	Better for spatial data, increased complexity [34]

F. Model Efficiency

To assess the computational efficiency of the railway-tuned LSTM model, inference time was measured as the average duration for a single forward pass on deployment hardware. On MEC servers, inference averaged 5-10 ms per prediction (SD = 1 ms, n=100 runs), processing 10-timestep sequences in real-time at 1 Hz sampling. This enables seamless integration into high-speed rail operations without delaying handovers (<15 ms total latency including network). Compared to baselines, this is competitive: DQN models require 10-15 ms [15], while simpler MLPs are faster (~3-5 ms) but less accurate [17]. Factors influencing efficiency include model size (200,000 parameters) and batch size (1 for real-time); optimizations like quantization could reduce it to 3-5 ms in future deployments. Measured using TensorFlow's timing utilities on edge hardware, ensuring <1% CPU overhead during inference [16].

G. Cost-Benefit and Scalability

The framework reduces infrastructure costs (~$500,000/km in urban areas) through phased deployments (urban in 2 years, rural in 5 years) and premium services

($5/month for high-speed streaming), achieving 20–30% urban cost savings ($100,000–150,000/km) and 15–20% revenue growth ($75,000–100,000/km annually) within 3–5 years [12]. Rural trials maintained 220 Mbps throughput and 92% handover success, robust despite weather impacts [3], [11]. To assess scalability across different rail infrastructures, the framework was analyzed for adaptability to varied terrains: In urban dense networks (Beijing-style with gNodeB spacing ~10-15 km), it leverages high overlap (80-90%) for seamless handovers; in rural or mountainous setups (European alpine routes), sparser gNodeB placement (~20-30 km) is compensated by integrating LEO satellites for gap-filling, maintaining >90% coverage with hybrid handovers (terrestrial-satellite switching latency<50 ms).

Coastal environments, prone to salt corrosion and high humidity, require ruggedized hardware (IP67-rated enclosures), with simulations showing only 5-10% performance degradation. For international standards compliance, the framework aligns with 3GPP Release 17 for 5G-R, ETSI MEC for edge deployments, and GSMA FRMCS guidelines for interoperability, enabling cross-border operations (e.g., Eurostar routes) via standardized APIs for multi-operator roaming. This ensures plug-and-play integration with legacy systems like GSM-R in Europe or LTE-R in Asia, with backward compatibility tested in NS-3 simulations (seamless fallback to LTE during 5G outages, <1% packet loss). Validation in diverse networks (e.g., mountainous, coastal) and higher density (500+ passengers) is needed, with 6G and LEO satellites as future scalability options [20], [21]. These outcomes address gaps in prior railway connectivity studies, offering a scalable solution [10], [18].

V. Discussion

A. Operator Implications

The 5G-R framework significantly enhances high-speed rail connectivity, delivering substantial benefits for operators and passengers. Field trials achieved an average throughput of 250 Mbps at 350 km/h (SD = 15 Mbps, 95% CI [246, 254], n = 50), as shown in Figure 8, enabling seamless video streaming, online communication, and real-time work [3]. This enhanced user experience can boost passenger satisfaction, potentially increasing ridership by 10–15% and generating additional revenue through premium services ($5/month for high-speed internet), offsetting infrastructure costs of approximately $500,000/km in urban areas, as discussed in Section IV.D For operators, 5G-R's low latency of 30 ms (SD = 3 ms, 95% CI [29, 31], n = 50) in high-density scenarios using MEC, as illustrated in Figure 6, supports real-time train monitoring, predictive maintenance, and safety systems, enhancing operational reliability [14].

Network slicing prioritizes critical operational data (train control) over passenger traffic, reducing congestion by 30% (SD = 5%, 95% CI [44, 46], n = 100) (PRB utilization in 300 UE scenarios), as summarized in Table 5, thereby improving safety and efficiency [8], [13]. Compared to logistic regression handover methods [31], which reduce unnecessary triggers but achieve only ~90% success, the framework's LSTM provides 5% higher accuracy for rail-specific prediction, though at higher computational cost. AMC antennas [40] offer gain improvements but require integration testing in rail; 5G-R combines these for scalable deployments.

Beamforming and railway-tuned LSTM algorithms further ensure reliable connectivity, increasing throughput by 39% (See Figure 5) and achieving a 95% handover success rate (SD = 1.5%, 95% CI [94.5, 95.5], n = 50) [9], [15], [16]. However, high deployment costs and technical challenges, such as rural rollouts and terrain variations, necessitate strategic phased deployments, as outlined in Section IV.D. In rural areas, sparse infrastructure and weather variability may increase costs by 20–30% and degrade performance, requiring hybrid satellite integration for coverage [21].

B. Future Trends

Emerging technologies promise to extend the 5G-R framework's capabilities. By 2030, 6G is expected to deliver sub-1 ms latency and terabit-per-second speeds, enhancing reliability in congested rail environments and supporting advanced applications like holographic communication [20]. LEO satellite systems, such as Starlink, can complement 5G-R in rural or mountainous areas, maintaining connectivity where terrestrial infrastructure is limited. This builds on Section IV.D's rural trial results, which achieved 220 Mbps throughput (SD = 12 Mbps, 95% CI [216, 224], n = 50) [11], [21]. Integrating 6G and LEO satellites could ensure seamless coverage across diverse rail networks, as discussed in Section IV.D. Advanced AI and machine learning (ML) will further optimize network performance, building on the 30% congestion reduction achieved through LSTM-based traffic management (See Table 5). Future AI/ML models could predict congestion, optimize handovers, and allocate resources in real-time by analyzing train schedules and passenger patterns, minimizing disruptions during peak demand [15], [16].

C. Study Limitations

The study's reliance on MATLAB and NS-3 simulations introduces limitations, as idealized assumptions (consistent network load, minimal weather interference) may not fully reflect real-world conditions. For instance, heavy rain or snow can reduce throughput by 10–15% (from 250 Mbps to 212–225 Mbps) and handover success by 2–3% (from 95% to 92–93%), as noted in Section IV.B's sensitivity analysis [3]. The Beijing-Zhangjiakou trial, while robust, is geographically limited to urban and suburban settings, potentially skewing results for diverse rail networks, such as mountainous or coastal routes. Additionally, selection bias may arise from focusing on urban/suburban routes, underrepresenting rural sparsity, while user density

Table 10. Summary of 5G-R Advancements and Future Research Directions

Aspect	Current Achievement	Future Research Direction
Throughput	250 Mbps at 350 km/h (SD = 15 Mbps, 95% CI [246, 254], n = 50)	Validate in diverse geographies (e.g., coastal, mountainous) under extreme weather
Latency	15 ms at 350 km/h (SD = 2 ms, 95% CI [14, 16], n = 50); 30 ms with MEC	Optimize sub-1 ms latency with 6G integration
Handover Success	95% at 350 km/h (SD = 1.5%, 95% CI [94.5, 95.5], n = 50)	Develop predictive handover models using AI/ML
Congestion Management	30% reduction with network slicing (SD = 5%, 95% CI [44, 46], n = 100)	Enhance real-time resource allocation with AI/ML for peak demand
Rural Coverage	220 Mbps in rural trials (SD = 12 Mbps, 95% CI [216, 224], n = 50)	Conduct 200-km hybrid 5G-R/Starlink trial for latency in weather variations.
Scalability	Urban/rural phased deployments, 20–30% cost savings	Test in higher-density scenarios (500+ passengers) and 6G networks

assumptions (50–300) could introduce bias in higher-density scenarios (500+), overlooking variability in passenger behavior or peak-hour surges. These factors limit the framework's immediate applicability without broader testing across varied conditions [11].

D. Future Research

Future research should prioritize field trials across diverse geographies, including rural, coastal, and mountainous regions, to validate 5G-R performance under varied conditions, such as extreme weather, ensuring stable throughput and handover success. Hybrid terrestrial-satellite networks, combining 5G-R with LEO systems, warrant experimental evaluation to achieve seamless rural coverage, building on the 220 Mbps throughput observed in rural trials. This includes a concrete proposal for a 200-km field trial integrating Starlink with 5G-R for rural gaps, measuring handover latency under varying weather.

Deeper AI/ML integration, beyond the current 30% congestion reduction, should explore real-time resource allocation and predictive handover models for complex scenarios, leveraging train schedules and passenger data. These advancements, summarized in Table 10, will strengthen the framework's scalability and address gaps in existing railway connectivity literature.

VI. Conclusion and Future Work

A. Conclusion

The 5G-R framework significantly enhances high-speed rail connectivity by addressing challenges of frequent handovers, signal degradation, and network congestion at speeds up to 350 km/h, with trial evidence showing 2400% throughput improvement over 4G LTE (10 Mbps to 250 Mbps) and 150% over standard 5G (100 Mbps to 250 Mbps), alongside 30% congestion reduction via network slicing (PRB utilization, 75% to 45%). By integrating beamforming, network slicing, railway-tuned LSTM algorithms, and MEC, it achieves a throughput of 250 Mbps, latency of 15 ms, and handover success rate of 95% (all trial-based). Compared to 4G LTE, 5G-R reduces latency by 85%; against standard 5G, it improves latency by 70%. Network slicing enables seamless passenger services like video streaming and real-time communication, while MEC supports real-time train monitoring and safety systems, enhancing operational reliability. In urban areas, premium services offset infrastructure capital costs, yielding 20–30% savings ($100,000–150,000/km), while rural trials sustain 220 Mbps throughput, demonstrating scalability. However, broader testing in diverse terrains is needed to ensure generalizability. These advancements establish 5G-R as a foundation for next-generation railway communications, paving the way for integration with future 6G networks and satellite technologies.

B. Future Work

Future research should prioritize field trials across diverse rail networks, including coastal and mountainous routes, to validate performance under extreme weather (e.g., heavy rain or fog reducing signal strength by 20-30%) and high passenger loads (500+ users per train). This could involve deploying additional gNodeBs with adaptive beamforming algorithms to test handover success rates in interference-prone areas, using metrics like packet loss under simulated fog (via NS-3 with added noise models). For integration with emerging technologies, explore hybrid 6G architectures by incorporating terahertz (THz) bands for sub-1 ms latency in dense urban segments, starting with simulations in MATLAB to model THz propagation along rail tracks and validate against current 5G-R benchmarks. Similarly, integrate LEO satellite systems (e.g., via Starlink APIs) for rural coverage gaps, testing seamless handover between terrestrial 5G-R and satellite links in a 100-km rural trial, focusing on latency spikes during transitions and using dynamic spectrum allocation to minimize interference.

Refining AI-driven traffic management could include optimizing the LSTM model with quantization techniques (e.g., 8-bit weights) to reduce inference time by 50% on edge hardware, evaluated through power profiling on MEC servers. Developing energy-efficient designs might involve implementing sleep modes for underutilized gNodeBs, targeting a 20-30% reduction in power consumption via SDN controllers, with validation in lab setups mimicking rail power constraints. Finally, extending the framework to high-mobility scenarios, such as autonomous vehicles, could adapt the LSTM for vehicle-to-infrastructure (V2I) communications, testing in simulated highway environments at 120 km/h to assess scalability of handover predictions.

REFERENCES

[1] W. Li, W. Chen, S. Wang, Y. Zhang, M. Matthaiou, and B. Ai, "AI-Driven Mobility Management for High-Speed Railway Communications: Compressed Measurements and Proactive Handover," Dec. 19, 2024, doi: 10.48550/arXiv.2407.04336.

[2] Z. Yin, "Mobility Performance Analysis for Mixed Handover," in 2024 IEEE Wireless Communications and Networking Conference (WCNC), Apr. 2024, pp. 1–6. doi: 10.1109/WCNC57260.2024.10570898.

[3] Y. Liang, H. Li, Y. Li, A. Li, Y. Wang, and D. Li, "Performance Analysis of 5G-R Network Bearing Train Control Information of HSR Based on Hardware-in-the-Loop Simulation," IEEE Access, vol. 11, pp. 60449–60461, 2023, doi: 10.1109/ACCESS.2023.3283369.

[4] S. Wang and L. Zhang, "Support vector machine-based handover scheme for heterogeneous ultra dense network of high-speed railway," IET Commun., vol. 18, no. 18, pp. 1191–1204, 2024, doi: 10.1049/cmu2.12814.

[5] D. Jia, F. Hu, Z. Ling, Z. Mao, Z. Chang, and T. Hamalainen, "AoI-Aware Spectrum and Power Allocation for High-Speed Railway Networks," in 2022 14th International Conference on Wireless Communications and Signal Processing (WCSP), Nanjing, China: IEEE, Nov. 2022, pp. 544–549. doi: 10.1109/WCSP55476.2022.10039199.

[6] X. Zhao, X. Meng, and B. Wang, "Advancing Railway Systems with Next-Generation 5G Connectivity," in 2024 International Wireless Communications and Mobile Computing (IWCMC), May 2024, pp. 357–362. doi: 10.1109/IWCMC61514.2024.10592331.

[7] Y. S. Chowdhary and G. Singh, "The Convergence of IoT and 5G: Transforming Connectivity in Smart City Applications," in 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Aug. 2024, pp. 311–318. doi: 10.1109/ICoICI62503.2024.10696019.

[8] W. Rafique, J. Barai, A. O. Fapojuwo, and D. Krishnamurthy, "A Survey on Beyond 5G Network Slicing for Smart Cities Applications," IEEE Commun. Surv. Tutor., pp. 1–1, 2024, doi: 10.1109/COMST.2024.3410295.

[9] R. W. Heath, N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," IEEE J. Sel. Top. Signal Process., vol. 10, no. 3, pp. 436–453, Apr. 2016, doi: 10.1109/JSTSP.2016.2523924.

[10] R. He et al., "5G for Railways: Next Generation Railway Dedicated Communications," IEEE Commun. Mag., vol. 60, no. 12, pp. 130–136, Dec. 2022, doi: 10.1109/MCOM.005.2200328.

[11] M. de L. Á. L. Fernández and A. P. Yuste, "Positioning of High-Speed Trains Using 5G Networks," in 2024 IEEE International Mediterranean Conference on Communications and Networking (MeditCom), Jul. 2024, pp. 221–226. doi: 10.1109/MeditCom61057.2024.10621243.

[12] Z. Kljaić, D. Pavković, M. Cipek, M. Trstenjak, T. J. Mlinarić, and M. Nikšić, "An Overview of Current Challenges and Emerging Technologies to Facilitate Increased Energy Efficiency, Safety, and Sustainability of Railway Transport," Future Internet, vol. 15, no. 11, Art. no. 11, Nov. 2023, doi: 10.3390/fi15110347.

[13] W. Hamdi, C. Ksouri, H. Bulut, and M. Mosbah, "Network Slicing-Based Learning Techniques for IoV in 5G and Beyond Networks," IEEE Commun. Surv. Tutor., vol. 26, no. 3, pp. 1989–2047, 2024, doi: 10.1109/COMST.2024.3372083.

[14] J. Xu, Z. Wei, Z. Lyu, L. Shi, and J. Han, "Throughput Maximization of Offloading Tasks in Multi-Access Edge Computing Networks for High-Speed Railways," IEEE Trans. Veh. Technol., vol. 70, no. 9, pp. 9525–9539, Sep. 2021, doi: 10.1109/TVT.2021.3101571.

[15] P. Santhiya, I. J. Jebadurai, G. J. L. Paulraj, J. A, G. R. A, and S. K. Karan, "ITS-AI Integration: Enhanced Strategies for Mitigating Urban Traffic Congestion," in 2024 5th International Conference on Recent Trends in Computer Science and Technology (ICRTCST), Apr. 2024, pp. 376–381. doi: 10.1109/ICRTCST61793.2024.10578453.

[16] J. Zhao, X. Gao, Z. Wu, Q. Zhang, and H. Han, "Artificial intelligence in rail transit wireless communication systems: Status, challenges and solutions," Phys. Commun., vol. 67, p. 102484, Dec. 2024, doi: 10.1016/j.phycom.2024.102484.

[17] X. Gao, J. Zhao, Q. Zhang, and H. Han, "Optimization of resource allocation strategy for high-speed railway based on deep reinforcement learning," Phys. Commun., vol. 66, p. 102455, Oct. 2024, doi: 10.1016/j.phycom.2024.102455.

[18] A. Sinha, V. Uduthalapally, D. Das, and R. Mahapatra, "SDN-Based Seamless Mobility Management for B5G Services in High-Speed Railways," in 2023 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), Dec. 2023, pp. 294–299. doi: 10.1109/ANTS59832.2023.10469293.

[19] M. Fukuda, "Trend on Research and Development Relating to Information and Communication Technology in Railway Fields," Q. Rep. RTRI, vol. 64, no. 3, pp. 161–165, 2023, doi: 10.2219/rtriqr.64.3_161.

[20] C.-X. Wang et al., "On the Road to 6G: Visions, Requirements, Key Technologies, and Testbeds," IEEE Commun. Surv. Tutor., vol. 25, no. 2, pp. 905–974, 2023, doi: 10.1109/COMST.2023.3249835.

[21] "Integration of 5G, 6G and IoT with Low Earth Orbit (LEO) networks: Opportunity, challenges and future trends," Results Eng., vol. 23, p. 102409, Sep. 2024, doi: 10.1016/j.rineng.2024.102409.

[22] Y. Pan, R. Li, and C. Xu, "The First 5G-LTE Comparative Study in Extreme Mobility," Proc ACM Meas Anal Comput Syst, vol. 6, no. 1, p. 20:1-20:22, Feb. 2022, doi: 10.1145/3508040.

[23] S. Senger and P. K. Malik, "A comprehensive survey of massive-MIMO based on 5G antennas," Int. J. RF Microw. Comput.-Aided Eng., vol. 32, no. 12, p. e23496, 2022, doi: 10.1002/mmce.23496.

[24] K. Zarzycki and M. Ławryńczuk, "LSTM and GRU neural networks as models of dynamical processes used in predictive control: A comparison of models developed for two chemical reactors," Sensors, vol. 21, no. 16, p. 5625, Aug. 2021, doi: 10.3390/s21165625.

[25] W. Zhang, X. Li, A. Li, X. Huang, T. Wang, and H. Gao, "SGRU: A High-Performance Structured Gated Recurrent Unit for Traffic Flow Prediction," in 2023 IEEE 29th International Conference on Parallel and Distributed

Systems (ICPADS), Dec. 2023, pp. 467–473. doi: 10.1109/ICPADS60453.2023.00076.

[26] M. Waqas and U. W. Humphries, "A critical review of RNN and LSTM variants in hydrological time series predictions," MethodsX, vol. 13, p. 102946, Dec. 2024, doi: 10.1016/j.mex.2024.102946.

[27] J. Yang, J. Liu, J. Guo, and K. Tao, "Track Irregularity Identification Method of High-Speed Railway Based on CNN-Bi-LSTM," Sensors, vol. 24, no. 9, Art. no. 2861, Apr. 2024, doi: 10.3390/s24092861.

[28] D. Li, S. Du, and Y. Hou, "Long-Term Passenger Flow Forecasting for Rail Transit Based on Complex Networks and Informer," Sensors, vol. 24, no. 21, p. 6894, Jan. 2024, doi: 10.3390/s24216894.

[29] D. E. Ruíz-Guirola, O. L. A. López, S. Montejo-Sánchez, R. D. Souza, and M. Bennis, "Performance Analysis of ML-Based MTC Traffic Pattern Predictors," IEEE Wireless Communications Letters, vol. 12, no. 7, pp. 1144–1148, July 2023, doi: 10.1109/LWC.2023.3264273.

[30] K. Kosukegawa, Y. Mori, H. Suyari, and K. Kawamoto, "Spatiotemporal forecasting of vertical track alignment with exogenous factors," Sci Rep, vol. 13, no. 1, p. 2354, Feb. 2023, doi: 10.1038/s41598-023-29303-7.

[31] A. M. Fernandes, H. I. Del Monego, B. S. Chang, and A. Munaretto, "Reducing Unnecessary Handovers Using Logistic Regression in 5G Networks," IEEE Access, vol. 13, pp. 78707–78726, 2025, doi: 10.1109/ACCESS.2025.3564532.

[32] H. Ma, B. Lu, T. Zheng, K. Thar, and M. Gidlund, "A Gated-Guided Serial CNN-Transformer Network for High-Speed Railway Traffic Prediction," in 2025 IEEE 21st International Conference on Factory Communication Systems (WFCS), June 2025, pp. 1–8. doi: 10.1109/WFCS63373.2025.11077652.

[33] D. Pjanić, A. Sopasakis, A. Reial, and F. Tufvesson, "Early-Scheduled Handover Preparation in 5G NR Millimeter-Wave Systems," IEEE Open Journal of the Communications Society, vol. 5, pp. 6959–6971, 2024, doi: 10.1109/OJCOMS.2024.3488594.

[34] R. Jalalifar, M. R. Delavar, and S. F. Ghaderi, "SAC-ConvLSTM: A novel spatio-temporal deep learning-based approach for a short term power load forecasting," Expert Systems with Applications, vol. 237, p. 121487, Mar. 2024, doi: 10.1016/j.eswa.2023.121487.

[35] B. Duan, C. Li, J. Xie, W. Wu, and D. Zhou, "Fast Handover Algorithm Based on Location and Weight in 5G-R Wireless Communications for High-Speed Railways," Sensors, vol. 21, no. 9, Art. no. 9, Jan. 2021, doi: 10.3390/s21093100.

[36] Y. Zhou and B. Ai, "Handover schemes and algorithms of high-speed mobile environment: A survey," Computer Communications, vol. 47, pp. 1–15, July 2014, doi: 10.1016/j.comcom.2014.04.005.

[37] GSMA, "Generic Network Slice Template," Permanent Reference Document NG.116, Version 9.0, Apr. 2023. [Online]. Available: https://www.gsma.com/newsroom/wp-content/uploads/NG.116-v9.0-1.pdf

[38] M. Khani, S. Jamali, M. K. Sohrabi, M. M. Sadr, and A. Ghaffari, "Slice admission control in 5G cloud radio access network using deep reinforcement learning: A survey," International Journal of Communication Systems, vol. 37, no. 13, p. e5857, 2024, doi: 10.1002/dac.5857.

[39] Y. Ullah et al., "A survey on AI-enabled mobility and handover management in future wireless networks: key technologies, use cases, and challenges," J. King Saud Univ. Comput. Inf. Sci., vol. 37, no. 4, p. 47, May 2025, doi: 10.1007/s44443-025-00048-9.

[40] H. Malekpoor and M. Hamidkhani, "Radiation Improvement of Low-Profile Microstrip Antenna by Utilizing AMC Surface for MIMO and Wireless Systems," International Journal of Industrial Electronics Control and Optimization, vol. 8, no. 2, pp. 129–136, June 2025, doi: 10.22111/ieco.2024.48644.1561.

Shahpour Rahmani is a Lecturer at the Department of Computer Science, University of Sistan and Baluchestan, Zahedan, Iran. He received the B.S. degree in applied mathematics in computer science from the University of Sistan and Baluchestan, Zahedan, Iran, in 1996, and the M.S. degree in computer (software) engineering from the Iran University of Science and Technology, Tehran, Iran, in 2008, and is currently working toward the Ph.D. degree in the School of Electrical and Computer Engineering, University of Tehran, Tehran, Iran. He joined the Router Laboratory and the Wireless Communication Research Group, University of Tehran, Tehran, Iran, in 2017, as a Research Assistant. His research interests include 5G/6G networks, machine learning for mobility management, high-speed railway communications, Low Earth Orbit (LEO) satellite networking, distributed systems, data science, routing protocols, and data center design.

Nasser Yazdani is a full professor at the Department of Electrical and Computer Engineering in Tehran University. He got his B.S. degree in Computer Engineering from Sharif University of Technology, Tehran, Iran. He worked in Iran Telecommunication Research Center (ITRC) as a consultant, researcher and developer for few years. To pursue his education, he entered Case Western Reserve Univ, Cleveland, Ohio, USA, later and graduated as a Ph.D. in Computer Science and Engineering. Then, Prof. Yazdani worked in different companies and research institutes in USA. He joined the ECE Dept. of Univ. of Tehran, Tehran, Iran, in Sep. 2000. Prof. Yazdani then initiated different research projects and Labs in high-speed networking and systems. His research interests include networking, packet switching, access methods, distributed systems and database systems

IECO International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Adaptive Resilient Control of Uncertain Nonlinear Cyber-physical Systems under Deception Attack

Maryam Shahriari-kahkeshi | Seyed Hojat Nourian

Department of Electrical Engineering, Faculty of Engineering, Shahrekord University, Shahrekord, Iran.
Corresponding Author's email: m.shahriyarikahkeshi@alumni.iut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 17-May-2025
Received in revised form: 22-September-2025
Accepted: 02-October-2025
Published online: 23-Sep-2026

Keywords:
Deception attack,
False data injection attack,
Nonlinear cyber-physical systems,
Resilient control.

ABSTRACT

This work proposes an adaptive resilient control for uncertain nonlinear cyber-physical systems (CPSs) under deception attacks. It is assumed that attacker injects false data into the commands exchanged between the controller and actuator over the communication channels. The injected false data affects the control input in both additive and multiplicative forms. To deal with the uncertain dynamics of the system and additive term of cyber-attacks, the radial basis function-neural networks (RBF-NNs) are invoked. Also, to handle adverse effects of multiplicative term of cyber-attack, the Nussbaum-type gain function is employed. Then, by integrating the RBF-NN model and Nussbaum function into the command filtered backstepping (CFB) approach, the proposed resilient control scheme is designed. Compared with the existing works, the proposed control eliminates the "explosion of complexity" problem in the conventional backstepping approach, removes the trial and error in choosing time constant of the first order filters in the dynamics surface control (DSC) approach, compensates the filtering error and deals with both additive and multiplicative cyber-attacks in "controller to actuator" channel, simultaneously. Also, it mitigates the effects of the cyber-attack without requiring separate attack estimation unit, controller reconfiguration or readjustment algorithm. Simulation results on the robotic arm under different cyber-attacks verify effective resilient performance of the proposed control scheme.

I. Introduction

Cyber-physical systems integrate physical systems into the cyber space via network of communication channels [1, 2]. Such systems which typically include processing and computing units are able to monitor and control physical systems through communication networks and have found wide applications in different science and engineering fields like aerospace, smart grids, power systems, industrial processes, and etc. [3-6]. Most of the CPSs use open communication and computation platform and they are often vulnerable to various cyber-attacks.

Daniel of service (DoS) attack and deception attack are common types of cyber-attacks that affect the CPSs and deteriorate the system performance [7, 8]. The DoS attack is a cyber-attack in which adversaries block the communication channels between different system layers and prevent signal transmission. Depending on the place of DoS attack occurrence, control layer may not receive sensor measurements or actuator may not receive control inputs. Therefore, control system fails due to the lack of real-time data. A typical form of the deception attack is a false data injection (FDI) attack in which attackers penetrate the communication network and directly, inject false signals to manipulate the actual control inputs or sensor measurements. Generally, FDI attacks may occur at the communication channels transmitting "sensor measurements to the control unit" or transmitting "control commands to the actuators". As a result of FDI attack, control units or physical layers receive false data instead of the true one. Hence, to preserve the desired performance of the control system, it is necessary to deal with cyber-attacks carefully.

Mostly, existing researches on security of the CPSs in the presence of cyber-attacks focus on two subjects: (1) attack detection, and (2) resilient control design [9-11].

The first subject models cyber-attack as a conflict signal injecting to the network, then studies attack effects and concentrates on developing attack detection and identification schemes to detect system anomalies. Generally speaking, attack detection schemes can be divided into model-based and data-based approaches. The model-based schemes use analytical model of the system for attack detection while the other ones infer system model by using historical data of the system and detect attack occurrence by analyzing the available data [12-14].

The second subject tries to develop resilient control schemes to mitigate attack effects. Generally, there are two perspectives to design resilient control schemes named as active and passive. The active approach extracts attack information by detecting, estimating or identifying it via separate attack identification unit, and then mitigates attack effects based on the attack characteristics [15-20]. In contrast, the passive approach assumes that adversary attack is a bounded signal and tries to compensate its adverse effects without identifying and readjusting or reconfiguring controller after attack occurrence [21-27]. In passive approach, controller is designed to be resilient against predefined attacks. As it is inferred from above discussion, designing active attack mitigation scheme is more complicated than the passive one because it requires attack detection, and estimation unit as well as controller readjustment algorithm.

In [15], an active attack mitigation strategy based on the attack detection and reconstruction scheme was proposed for nonlinear CPSs under deception attack. It designs diagnostic observer for detecting and estimating occurred attack online and then mitigates attack effects by using the output of the attack diagnostic unit. In [16] attack mitigation scheme was proposed for nonlinear CPSs in the companion form under actuator deception attack based on the neural estimation of the occurred attack and sliding mode approach. In [17] resilient consensus scheme based on the attack detection and isolation approach was proposed for nonlinear second order multi-agent systems under FDI attack in communication channels between the follower agents. It first, detects the agent that receives false data, then by scanning all links ended to the detected agents, it identifies the under attack link. Finally, by eliminating the under attack link from the consensus algorithm, the attack mitigation is done. In [18] linear networked control systems under FDI attack and process noise were considered and then an active attack mitigation scheme was proposed. An attack mitigation scheme based on the adaptive sliding mode approach and attack estimation was proposed for linear CPSs in the presence of FDI actuator attack in [20].

In [6] consensus of cyber-physical power systems in presence of model uncertainties, disturbances and cyber-attacks has been considered. The considered system in [20] is a fractional order linear time-invariant one, and the considered cyber-attack is a time-dependent bounded function. The passive resilient scheme proposed in [6] designs an adaptive chattering-free fractional order sliding mode controller to achieve consensus in presence of uncertainties, disturbances and cyber-attacks. An adaptive neural network-based DSC scheme was proposed for uncertain nonlinear time-delay CPSs under sensor and actuator deception attack in [21]. In [22] passive resilient control of nonlinear CPSs in the canonical form under actuator attack and input saturation was investigated. The proposed scheme in [22] is based on the Nussbaum function, barrier Lyapunov function and extended state observer and it assumes that system model is exactly known. Adaptive backstepping-based resilient control of nonlinear CPSs in the presence of deception and injection attack was studied in [23, 24]. The considered CPSs in [23, 24] has special form known as linearly parameterized strict-feedback form with bounded control gains, also it assumes that attack signal satisfies special inequalities which limit applicability of the proposed schemes for handling general form of attacks. Two other passive resilient control schemes for nonlinear CPSs under FDI attack were proposed in [25, 26]. In [27], a fixed-time adaptive resilient control framework based on the reinforcement learning, disturbance observer and nonsingular fast terminal sliding mode was proposed for nonlinear CPSs under FDI attack. In [28], a reinforcement learning-based optimal resilient control was proposed for large scale interconnected nonlinear systems in the presence of sensor attack and actuator hysteresis. Also, some model free resilient control schemes were proposed for nonlinear CPSs in [29-31]. In [32], a leader-follower formation control based on the neural networks, projection algorithm and DSC approach was proposed for an uncertain unmanned surface vehicle under stochastic disturbance. In [33], an adaptive backstepping sliding mode control was proposed to stabilize nonlinear multi-input multi-output epidemic systems under input saturation, modeling uncertainties and external disturbances.

Inspired by the above discussion, this paper solves the control problem in nonlinear uncertain CPSs under FDI attack in “controller to actuator” communication channel. An adaptive resilient control scheme based on the RBF-NN, Nussbaum function and CFB approach is proposed to deal with the modeling uncertainty, and to mitigate the attack effects as well as unknown control direction problem arisen because of actuator multiplicative attack. The suggested scheme does not require any prior information about attack time, severity and characteristic. Also, it does not require attack detection and identification unit, and controller reconfiguration mechanism for mitigating attack effects online. In addition it is able to mitigate different kind of additive and multiplicative attacks as well as constant, time-varying and stochastic ones. Stability analysis shows that the proposed resilient scheme ensures stability of the CPSs under input attack and guarantees that by proper selection of the design parameters, tracking error can be made small.

Thorough comparison among the prior reviewed works and the proposed scheme has been provided in Table 1. In this Table, the reviewed references on passive resilient control of nonlinear CPSs and the proposed work were compared with each other, from four different aspects including: (1) considered nonlinear CPS form, (2) attack location, (3) control approach, and (4) limitations of the schemes.

Main contributions of the proposed work are: (1) Compared with the existing works [22-25], the proposed work considers a more general class of nonlinear CPSs, therefore, it is applicable to wide variety of practical systems like those described later in remark 2. (2) The proposed resilient control scheme-based on the CFB approach eliminates the "explosion of complexity" problem which is conventional in the backstepping-based approaches suggested in [23, 24, 26, 27, 33]. (3) The proposed scheme uses a compensation mechanism to eliminate the filtering error. Also, it uses a command filter to obtain derivatives of the virtual input and therefore, it eliminates sensitivity to the time-constant of the first order filters in the DSC-based approaches, unlike [21, 32]. (4) Compared with [21, 22, and 25], the proposed scheme not only deals with additive attacks, but also can handle the multiplicative attacks that multiply the control gain by an unknown term, alter the magnitude and effectiveness of the control input and cause more significant changes than the additive attacks. However the proposed advantages, it only deals with actuator attack and it is not applicable for CPSs with unsafe "sensor to controller" communication channel.

Note: Throughout the paper, scalars are represented in italics; while the vectors are denoted in bold and italics.

II. Problem Statement and Preliminaries

This section first presents problem statement and then describes some preliminaries.

A. Problem Statement

The following class of uncertain nonlinear CPSs under deception attack represented in Fig. 1 is considered:

$$\begin{aligned} \dot{x}_i &= x_{i+1} + f_i(x_i), 1 \le i \le n-1 \\ \dot{x}_n &= f_n(x) + u, \\ y &= x_1, \end{aligned} \tag{1}$$

where $x_i \in \Re$ is the state variable, $\boldsymbol{x} = [x_1 \quad \cdots \quad x_n] \in \Re^n$ and $x_i = [x_1 \quad \cdots \quad x_i] \in \Re^i$ represent the state vectors, $u \in \Re$ is the control input and $y \in \Re$ is the output variable.

Also, $f_i(\boldsymbol{x}_i)$ for $i = 1,2,,\ldots,n$ represents the uncertain smooth functions in the system model. The model of deception attack affecting the control signal is described by

TABALE I Comparison between the proposed scheme and related passive resilient methods

Ref. No.	Considered CPS	Attack location	Control approach	Limitations
[21]	Time-delay nonlinear CPS	Sensor and actuator deception attack	Dynamic surface control approach	(1) Existence of filtering error, and (2) sensitivity to time constant of the first order filters
[22]	nth order affine nonlinear CPS	Actuator attack	Adaptive control	(1) Applicable to simple form of nonlinear CPSs, and (2) ability to deal with simple additive actuator attack
[23, 24]	linearly parameterized nonlinear CPS	Sensor and actuator attack	Backstepping approach	(1) Explosion of complexity problem, and (2) applicable to linearly parameterized form of CPSs.
[25]	nth order affine nonlinear CPS	Sensor and actuator attack	Feedback linearization approach	(1) Applicable to simple form of nonlinear CPSs, and (2) only deals with additive attack.
[26]	Nonlinear strict-feedback CPS	Sensor attack	Backstepping approach	(1) Explosion of complexity problem, and (2) dealing with sensor attack.
[27]	Large scale nonlinear systems	Sensor attack	Reinforcement-based adaptive backstepping control	(1) Explosion of complexity problem, and (2) dealing with sensor attack.
Proposed Scheme	Nonlinear CPS in the strict-feedback form	Actuator attack	Command filtered backstepping approach	(1) Dealing with actuator attack.

approach provides a simple controller with considerable

$$u = bv_n + a(x), \tag{2}$$

where v_n is the controller output, b describes multiplicative attack signal which is unknown constant, $a(\boldsymbol{x})$ represents the injected attack signal which is unknown bounded state-dependent function, and u denotes the under-attack signal that is send to the actuator through the network.

Substituting the attack signal (2) in (1), results in

$$\begin{aligned} \dot{x}_i &= x_{i+1} + f_i(x_i), 1 \le i \le n-1 \\ \dot{x}_n &= f_n(x_n) + bv_n + a(x), \\ y &= x_1. \end{aligned} \tag{3}$$

Remark 1. It is worth to note that the occurrence time, severity, and characteristic of the attack signal is unknown for controller design.

Remark 2. Wide variety of practical systems such as the jet engine compression system [34], the twin otter aircraft [35], the robot manipulator [36], and the piezoelectric-positioning mechanism [37] can be described by the nonlinear system in the strict-feedback form. Therefore, it is of both theoretical and practical importance to study the control problem for strict-feedback nonlinear systems.

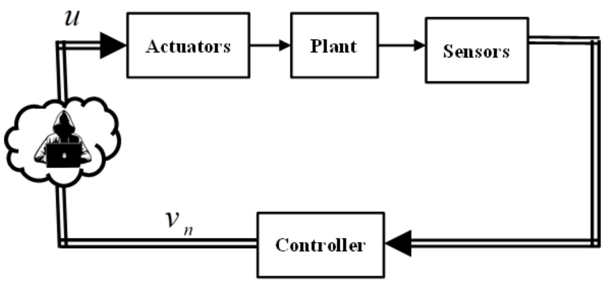


Fig. 1. Schematic of the under attack CPS.

Assumption 1. The desired signal y_d and its first derivative, i.e., $\dot{y}_d$ are known, smooth and bounded.

Remark 3. In assumption 1, y_d is the desired input to the closed-loop system. Therefore, it is reasonable to assume that it is available and known. The backstepping approach requires the knowledge of $y_d^{(i)}$where $i = 0,1,\dots,n-1$, or conventional DSC approach requires the knowledge of y_d, $\dot{y}_d$ and $\ddot{y}_d$. However, assumption 1 just needs the knowledge of y_d and its first derivative which is less restrictive. It signifies that the proposed scheme is more suitable for some important applications where higher order derivatives are unavailable.

Before presenting the proposed resilient control approach, some preliminaries are reviewed briefly.

B. Preliminary Concepts

Because of using RBF-NN for modeling uncertain dynamics of the system and unknown additive FDI attack, at first, RBF-NN is described briefly. Figure (2) shows the RBF-NN structure.

As it is obtained from Fig. 2, RBF-NN consists of three layers: input layer, hidden layer and output layer. Activation function in the hidden layer of the RBF-NN has a radial basis function form as follows:

$$\phi_i(x) = exp\left(-\frac{\|x-\chi_i\|}{2\eta_i^2}\right), \tag{4}$$

where $\boldsymbol{x}$ is the input vector, χ_i and η_i denote the center and width of the activation functions, respectively. The center parameter can be placed on a random subset or all of the training examples, or chosen by using the clustering algorithms or determined by invoking heuristic algorithms or learning procedure. Also, the basis function widths can be either chosen to be the same for all of the units or can be chosen differently for each unit, depending on the part of the input space they represent. It is worth to note that the training process may adapt the width parameter for each of the basis functions, also.

As it is obtained from Fig. 2, output of the RBF-NN can be written as

$$f(x) = \sum_{i=1}^{N} w_i\, \phi(x_i), \tag{5}$$

where w_i is the weight coefficient between the ith neuron and output, N is the number of the network neurons, and x_i is the

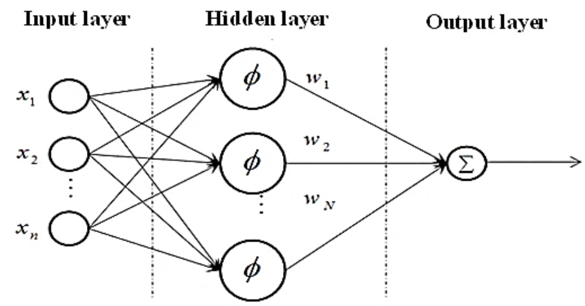


Fig. 2. Structure of the RBF-NN.

input of the ith neuron. For simplicity, network output in (5) is rewritten in the following linear regression form

$$f(x) = w^T\varphi(x), \tag{6}$$

where $w = [w_1 \quad w_2 \quad \dots \quad w_N]^T$ is the weight vector and $\varphi(x) = [\phi_1(x_1) \quad \phi_2(x_2) \quad \dots \quad \phi_N(x_N)]^T$ is the activation function vector.

According to the universal approximation property of the RBF-NN [38, 39], the RBF-NN with sufficiently large number of neurons (N_i) can approximate any continuous real valued function $f_i(x_i)\colon \Re^i \to \Re$ over a compact set $\Omega \subset \Re^i$ to any desired accuracy as:

$$f_i(x_i) = w_i^{*T}\varphi_i(x_i) + \varepsilon_i(x_i), \tag{7}$$

where $w_i^* \in R^N$ represents an ideal weight vector satisfying (8), and $\varepsilon_i(x_i)$ is the approximation error with an assumption of $|\varepsilon_i(x_i)| \le \bar{\varepsilon}_i$, where the unknown constant $\bar{\varepsilon}_i > 0$ for all $x_i \in \Omega$,

$$w^* = \arg\min_{w\in R^N}\left\{\sup_{x_i\in\Omega}|f_i(x_i) - w_i^T\varphi_i(x_i)|\right\}. \tag{8}$$

Since w_i^* is ideal weight vector, it is unknown and it is necessary to estimate it online. In the following, w_i denotes the estimation of w_i^*.

Now, some required lemmas for controller design are explained.

Lemma 1 [40]. Consider the following command filter

$$\begin{aligned} \dot{z}_1 &= \omega_n z_2, \\ \dot{z}_2 &= -2\xi\omega_n z_2 - \omega_n(z_1 - v), \end{aligned} \tag{9}$$

where v is the input signal. If the first and second derivatives of the input signal v for all $t \ge 0$ are bounded, i.e., $|\dot{v}| \le \bar{v}$ and $|\ddot{v}| \le \bar{v}_1$, also $z_1(0) = v(0)$, $z_2(0) = 0$, then for any $v^* > 0$, there exists filter design parameters $\omega_n > 0$ and $\xi \in (0,1]$ such that $|z_1 - v| \le v^*$, $|\dot{z}_1|$, $|\ddot{z}_1|$, and $|\dddot{z}_1|$ is bounded.

Definition 1 [41]. Function $N(\kappa)\colon \Re \to \Re$ is called Nussbaum function if it satisfies the following properties:

$$\begin{aligned} &\lim_{p\to\infty} sup\, \frac{1}{p}\int_0^p N(\kappa)\, dk = +\infty, \\ &\lim_{p\to-\infty} sup\, \frac{1}{p}\int_0^p N(\kappa)\, dk = -\infty. \end{aligned} \tag{10}$$

Lemma 2 [42]. Consider smooth functions $V(.)$ and $\kappa(.)$ defined on interval $(0, t_f)$, in which $V(t) \ge 0$ for $\forall t \in (0, t_f)$ and $N(\kappa)$ is a Nussbaum function, if the following inequality holds then κ, V, and $\int_0^t (bN(\kappa)+1)\,\dot{\kappa}e^{a_1\tau}d\tau$ on the defined interval $(0,t_f)$ will be bounded:

$$0 \le V(t) \le a_0 + e^{-a_1 t}\int_0^t (bN(\kappa)+1)\,\dot{\kappa}e^{a_1\tau}d\tau, \qquad (11)$$

In the above inequality, $a_0, a_1 > 0$ are proper constant values and b satisfies $b: R \to [l^- \quad l^+]$ where $l^- l^+ > 0$.

III. Proposed Scheme

In this section the proposed CFB-based resilient control scheme is explained for considered CPS (3) under FDI attack. In the proposed scheme to decrease number of learning parameters, and consequently to reduce the online computational burden, the minimal learning parameters (MLP) algorithm [43] has been used. In this algorithm, norm of the weight vector of the RBF-NN is considered as one adaptive parameter and it is updated online. This algorithm aims to achieve universal approximation property with fewest possible adjustable parameters and helps to improve computational efficiency. For having this, let us define adaptive parameter θ_i as $\theta_i := \|w_i\|$ for $i = 1,2,\ldots,n$ and update it online based on the adaptive laws which will be derived later. By using this strategy, number of adaptive parameters is decreased from N_i (number of RBF-NN neurons for modeling $f_i(x_i)$) to *1*.

Let us define the proposed error surfaces as

$$\begin{aligned} s_1 &= x_1 - y_d, \\ s_i &= x_i - x_{i,c}, for i = 2,\ldots,n \end{aligned} \qquad (12)$$

where $x_{i,c}$ is obtained from the following command filter

$$\begin{aligned} \dot{z}_{i,1} &= \omega_n z_{i,2}, \\ \dot{z}_{i,2} &= -2\xi\omega_n z_{i,2} - \omega_n(z_{i,1} - v_i). \end{aligned} \qquad (13)$$

In (13), v_i is the filter input, $z_{i,1}$ and $z_{i,2}$ are the filter outputs. Also, ξ and ω_n are design parameters of the filter. In order to eliminate the filtering error, the compensated tracking error signal e_i is defined as:

$$e_i = s_i - \zeta_i for i = 1,\ldots,n, \qquad (14)$$

where ξ_i is defined as:

$$\begin{aligned} \dot{\zeta}_i &= -c_i\zeta_i + x_{i+1,c} - \alpha_i + \zeta_{i+1}, i = 1,\ldots,n-1 \\ \dot{\zeta}_n &= 0, \end{aligned} \qquad (15)$$

and $\zeta(0) = 0$. In (15), c_i, $i = 1,\ldots,n-1$ are positive constants chosen by the designer. As it was proved in detail in [44, Lemma 3, 45], ζ_i is bounded and $\lim_{t\to\infty}|\zeta_i| \le \frac{v^*}{2c_0}$ where $c_0 = \left(\frac{1}{2}\right)min(c_i)$.

By considering RBF-NN model in (7), and adjusting norm of the weight coefficients as an adaptive parameter, the following inequality is obtained:

$$e_i f_i \le \frac{1}{2}e_i^2\theta_i\varphi_i^T\varphi_i + \frac{1}{2} + \frac{1}{2}e_i^2 + \frac{1}{2}\varepsilon_i^2. \qquad (16)$$

Also, the following inequalities hold:

$$\begin{aligned} e_i e_{i+1} &\le \frac{1}{2}e_i^2 + \frac{1}{2}e_{i+1}^2, \\ \tilde{\theta}_i\theta_i &\le \frac{1}{2}\tilde{\theta}_i^2 + \frac{1}{2}\theta_i^2. \end{aligned} \qquad (17)$$

Now, design steps of the proposed scheme are presented.

Step 1: At first, by considering the first error surface and differentiating it with respect to time, and substituting the first equation of (3) for $i = 1$ in the result, dynamics of the first error surface is obtained as:

$$\dot{s}_1 = x_2 + f_1 - \dot{y}_d. \qquad (18)$$

Now the following virtual input for stabilizing (18) is proposed:

$$\alpha_1 = -c_1 s_1 + \dot{y}_d - \frac{1}{2}\hat{\theta}_1 e_1\varphi_1^T\varphi_1, \qquad (19)$$

where c_1 is a design parameter and $\hat{\theta}_1$ is an adaptive parameter updated based on the following adaptive law:

$$\dot{\hat{\theta}}_1 = \frac{\gamma_1}{2}e_1^2\varphi_1^T\varphi_1 - \gamma_1\sigma_1\hat{\theta}_1, \qquad (20)$$

where σ_1 is a design parameter and positive constant γ_1 represents a learning rate. Considering (14) for $i = 1$, differentiating it with respect to time and substituting (15), (18), and (19) in the result, gives:

$$\dot{e}_1 = e_2 - c_1 e_1 + \left(f_1 - \frac{1}{2}\hat{\theta}_1 e_1\varphi_1^T\varphi_1\right). \qquad (21)$$

Consider the following Lyapunov function for stability analysis

$$V_1 = \frac{1}{2}e_1^2 + \frac{1}{2\gamma_1}\tilde{\theta}_1^2, \qquad (22)$$

where $\tilde{\theta}_1 = \theta_1 - \hat{\theta}_1$ is the parameter estimation error. Differentiating (22) with respect to time and substituting (21) in the result gives

$$\begin{aligned} \dot{V}_1 = e_1e_2 - c_1e^2{}_1 + e_1f_1 - \left(\frac{1}{2}\hat{\theta}_1 e^2{}_1\varphi_1^T\varphi_1\right) \\ -\frac{1}{\gamma_1}\tilde{\theta}_1\dot{\hat{\theta}}_1. \end{aligned} \qquad (23)$$

By applying inequalities (16) and (17) to (23) and considering (20), equation (23) can be expressed as

$$\dot{V}_1 \le (1-c_1)e_1^2 + \frac{1}{2}e_2^2 + \sigma_1\tilde{\theta}_1\hat{\theta}_1 + \frac{1}{2} + \frac{1}{2}\varepsilon_1^2. \qquad (24)$$

Substituting $\hat{\theta}_1 = \theta_1 - \tilde{\theta}_1$ in (24), gives

$$\begin{aligned} \dot{V}_1 \le (1-c_1)e_1^2 + \frac{1}{2}e_2^2 + \sigma_1\left(\tilde{\theta}_1\theta_1 - \tilde{\theta}^2{}_1\right) + \frac{1}{2} \\ + \frac{1}{2}\varepsilon_1^2. \end{aligned} \qquad (25)$$

Applying inequality (17) to (25) results in:

$$\begin{aligned} \dot{V}_1 \le (1-c_1)e_1^2 + \frac{1}{2}e_2^2 + \frac{\sigma_1}{2}\theta_1^2 - \frac{\sigma_1}{2}\tilde{\theta}_1^2 + \frac{1}{2} \\ + \frac{1}{2}\bar{\varepsilon}_1^2. \end{aligned} \qquad (26)$$

Step i (i=2, …, n-1): Consider the *i*th error surface. By taking the time derivative of s_i and substituting (3) in the results, we will have

$$\dot{s}_i = x_{i+1} + f_i - \dot{x}_{i+1,c}. \qquad (27)$$

To stabilize (27), the following virtual control input is proposed:

$$\alpha_i = -c_i s_i + \dot{x}_{i,c} - \frac{1}{2}\hat{\theta}_i e_i\varphi_i^T\varphi_i - e_{i-1}, \qquad (28)$$

where c_i is a design parameter, and $\hat{\theta}_i$ is an adaptive parameter adjusted based on the following adaptive law

$$\dot{\hat{\theta}}_i = \frac{\gamma_i}{2} e_i^2 \varphi_i^T \varphi_i - \gamma_i \sigma_i \hat{\theta}_i, \tag{29}$$

where σ_i is a design parameter and positive constant γ_i is the learning rate. Again, by doing the same procedures as those described in step 1, we will have

$$\dot{e}_i = e_{i+1} - c_i e_i + f_i - \frac{1}{2}\hat{\theta}_i e_i \varphi_i^T \varphi_i - e_{i-1}. \tag{30}$$

Again, the following Lyapunov function is considered for stability analysis

$$V_i = \frac{1}{2} e_i^2 + \frac{1}{2\gamma_i} \tilde{\theta}_i^2, \tag{31}$$

where $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$. Differentiating (31) with respect to time and substituting (30) in the result, gives

$$\begin{aligned} \dot{V}_i &= e_i e_{i+1} - c_i e_i^2 + e_i f_i - \frac{1}{2}\hat{\theta}_i e_i^2 \varphi_i^T \varphi_i - e_i e_{i-1} \\ &- \frac{1}{\gamma_i} \tilde{\theta}_i \dot{\hat{\theta}}_i. \end{aligned} \tag{32}$$

Applying inequalities (16) and (17) to (32), gives:

$$\begin{aligned} \dot{V}_i \le \left(\frac{3}{2} - c_i\right) e_i^2 + \frac{1}{2} e_{i+1}^2 + \frac{1}{2} e_i^2 \tilde{\theta}_i \varphi_i^T \varphi_i \\ - \frac{1}{\gamma_i} \tilde{\theta}_i \dot{\hat{\theta}}_i \\ + \frac{1}{2} + \frac{1}{2}\varepsilon_i^2 + \frac{1}{2} e^2{}_{i-1}. \end{aligned} \tag{33}$$

Substituting (29) in (33), and substituting $\tilde{\theta}_i = \theta_i - \hat{\theta}_i$ in the result, gives

$$\begin{aligned} \dot{V}_i \le \left(\frac{3}{2} - c_i\right) e_i^2 + \frac{1}{2} e_{i+1}^2 + \frac{\sigma_i}{2}\theta_i^2 - \frac{\sigma_i}{2}\tilde{\theta}_i^2 \\ + \frac{1}{2} + \frac{1}{2}\bar{\varepsilon}_i^2 + \frac{1}{2} e^2{}_{i-1}. \end{aligned} \tag{34}$$

Step n: Consider the *n*th error surface in (12). Taking the time derivative of (12) and substituting the last equation of (3) in $\dot{s}_n$, results in

$$\dot{s}_n = f_n + b v_n + a(x_n) - \dot{x}_{n,c}. \tag{35}$$

To stabilize (35), the control input is proposed as

$$\begin{aligned} v_n &= N(\kappa) v, \\ \dot{\kappa} &= e_n v, \end{aligned} \tag{36}$$

where v has the following form

$$v = c_n e_n - \dot{x}_{n,c} + \frac{1}{2}\hat{\theta}_n e_n \varphi_n^T \varphi_n + e_{n-1}. \tag{37}$$

In (37), c_n is a design parameter and $\hat{\theta}_n$ is adjusted based on the following adaptive law:

$$\dot{\hat{\theta}}_n = \frac{\gamma_n}{2} e_n^2 \varphi_n^T \varphi_n - \gamma_n \sigma_n \hat{\theta}_n, \tag{38}$$

where $\gamma_n > 0$ is the learning rate and $\sigma_n > 0$ is a design parameter. Considering (35) and $\zeta_n = 0$, the following tracking error is obtained

$$\begin{aligned} \dot{e}_n = f_n + a - \frac{1}{2}\hat{\theta}_n e_n \varphi_n^T \varphi_n + (bN(\kappa) + 1) v \\ - c_n e_n - e_{n-1}. \end{aligned} \tag{39}$$

Now for stability analysis, the following Lyapunov function is considered:

$$V_n = \frac{1}{2} e_n^2 + \frac{1}{2\gamma_n} \tilde{\theta}_n^2. \tag{40}$$

Taking time derivative of (40) and using (39) results in

$$\begin{aligned} \dot{V}_n = e_n(f_n + a) - \frac{1}{2}\hat{\theta}_n e_n^2 \ \varphi_n^T \varphi_n \\ + e_n(bN(\kappa) + 1) v \\ - c_n e_n^2 - e_n e_{n-1} - \frac{1}{\gamma_n} \tilde{\theta}_n \dot{\hat{\theta}}_n. \end{aligned} \tag{41}$$

Applying inequalities (16) and (17) to (41) and substituting (38) in the result, gives

$$\begin{aligned} \dot{V}_n \le (1 - c_n) e_n^2 + \frac{1}{2} + \frac{1}{2}\bar{\varepsilon}_n^2 + e_n(bN(\kappa) + 1) v \\ + \frac{1}{2} e_{n-1}^2 + \sigma_n \tilde{\theta}_n \hat{\theta}_n \end{aligned} \tag{42}$$

Substituting $\hat{\theta}_n = \theta_n - \tilde{\theta}_n$ in (42) and applying inequality (17) to (42), results in

$$\begin{aligned} \dot{V}_n \le (1 - c_n) e_n^2 + e_n(bN(\kappa) + 1) v + \frac{\sigma_n}{2}\theta_n^2 \\ - \frac{\sigma_n}{2}\tilde{\theta}_n^2 + \frac{1}{2} + \frac{1}{2}\varepsilon_n^2 + \frac{1}{2} e_{n-1}^2 \end{aligned} \tag{43}$$

IV. Main Results

This section presents main results and provides stability analysis of the closed-loop system.

Theorem 1. Consider the nonlinear system (1) under FDI attack in "controller to actuator" channel which affects the CPS in the additive and multiplicative forms. It can be shown that the proposed RBF-NN-based CFB approach guarantees that the closed loop system is stable and all signals of the closed-loop system are uniformly ultimately bounded.

Proof. Consider the following Lyapunov candidate function

$$V_T = \sum_{i=1}^{n} V_i. \tag{44}$$

Taking the time derivative of (44) and substituting (26), (34), and (43) in the result, gives

$$\begin{aligned} \dot{V}_T \le \left(\frac{3}{2} - c_1\right) e_1^2 + \left(\frac{5}{2} - c_2\right) e_2^2 \\ + \left(\frac{5}{2} - c_3\right) e_3^2 + \dots \\ + \left(\frac{3}{2} - c_n\right) e_n^2 + e_n(bN(\kappa) + 1) v \\ + \sum_{i=1}^{n} \left(\frac{1}{2} + \frac{1}{2}\varepsilon_i^2 + \frac{\sigma_i}{2}\theta_i^2 - \frac{\sigma_i}{2}\tilde{\theta}_i^2\right), \end{aligned} \tag{45}$$

Comparing (45) with (44), it makes possible to rewrite (45) as

$$\dot{V}_T \le -\mu V_T + \beta + (bN(\kappa) + 1)\dot{\kappa} \tag{46}$$

where μ and β are obtained from substituting (44) and (45) in (46) as follows:

$$\begin{aligned} \mu &\triangleq min \left\{ \begin{matrix} 2c_1 - 3 \\ 2c_i - 5, 2 \le i \le n-1 \\ 2c_n - 3 \\ \gamma_i \sigma_i \end{matrix} \right\}, \\ \beta &= \sum_{i=1}^{n} \left(\frac{1}{2} + \frac{1}{2}\varepsilon_i^2 + \frac{\sigma_i}{2}\theta_i^2\right). \end{aligned} \tag{47}$$

Solving (47) with respect to time gives

$$V_T \le e^{-\mu t}V(0) + \frac{\beta}{\mu}(1-e^{-\mu t}) + e^{-\mu t}\int_0^t (bN(\kappa)+1)\,\dot{\kappa}e^{\mu\tau}d\tau. \tag{48}$$

Applying Lemma 2 to (48) reveals that κ, V, and $\int_0^t (bN(\kappa)+1)\,\dot{\kappa}e^{\mu\tau}d\tau$ are bounded on interval $(0,t)$. Therefore, e_i and $\tilde{\theta}_i$ for $i=1,2,\dots,n$ are bounded. Because $s_i = e_i + \zeta_i$ and according to [44: Lemma 3, 45] $|\zeta_i|$ is bounded, therefore signal s_i is bounded. Let us denote the bound of $\int_0^t (bN(\kappa)+1)\,\dot{\kappa}e^{\mu\tau}d\tau$ by $\bar{\Delta}$, so the ultimate bound of the tracking error will be $\lim_{t\to\infty}|s_1| \le \sqrt{\frac{2\beta}{\mu} + 2\bar{\Delta} + \frac{\nu^*}{2c_0}}$.

Remark 4. It can be seen from Lemma 1, and the definition of β, μ, and c_0 that the tracking error bound depends on the design parameters c_i, σ_i, and γ_i for $i=1,\dots,n$. Also, it depends on the command filter parameters (ω_n, ξ) and width and center of the basis functions in the RBF-NN. Increasing c_i can lead to smaller tracking error and faster convergence, but it might induce undesirable oscillations or require larger control effort, potentially leading to input saturation. Also, appropriately tuning of the adaptation parameters γ_i and σ_i ensures accurate estimation and effective compensation of uncertainties, thus minimizing tracking error. Also, a command filter with high natural frequency (ω_n) generally leads to smaller filtering errors and thus, potentially improves the tracking accuracy, as the filter can more closely track the desired signals. However, increasing the bandwidth too much can lead to increase the sensitivity to noise and oscillations. Also, the damping ratio of the command filter (ξ) influences the transient response of the

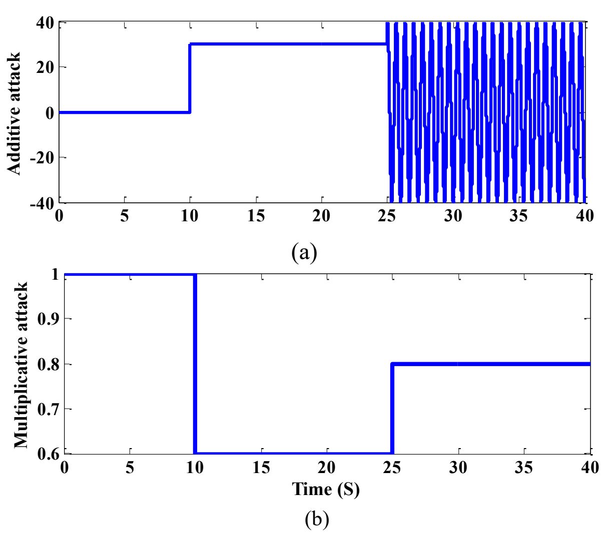


Fig. 3. Attack signal, (a) Additive term, and (b) Multiplicative term (case 1).

filtered signals. An appropriate damping ratio helps preventing from oscillations and overshoots in the filtered commands, which can in turn, reduce fluctuations in the

TABLE II PHYSICAL PARAMETERS

Physical Parameters	Definition	Numerical Value
L (m)	Distance between center and end of the joint	1
M (kg)	Mass of the joint	1
g (m/S^2)	Acceleration of gravity	9.8
I (kg m^2)	Moment of inertia	4/3

tracking error. Also, width and center of the basis functions in the RBF-NN structure are two other important design parameters. A basis function with narrow (small) width results in a more precise approximation of the unknown system dynamics in specific regions of the state space. In such case, if the system moves into regions that are not covered by the RBF centers well, or if the width is excessively small, the network's generalization capability might be compromised, leading to increased tracking error and larger control signal. While basis functions with large width cover a wider area of the input space and consequently, improve the network's generalization ability across a larger operating range, they potentially reduce the tracking error, control effort and chattering. A well-distributed set of centers, particularly in regions where the system dynamics exhibit significant variations, enhances the approximation capability, thereby reducing the tracking error.

In summary, there are inherent trade-offs in the selection of these parameters. For instance, achieving a very small tracking error often requires higher controller gains, which can lead to increased control input amplitude and potential saturation issues.

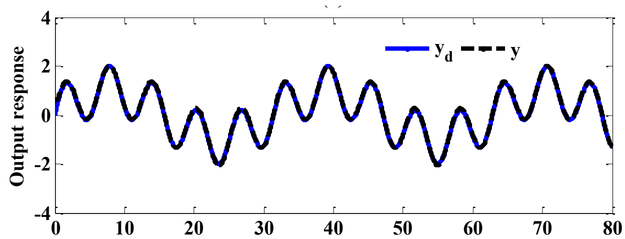
V. Simulation Results

To investigate the performance of the proposed scheme, the following robotic arm described by (49) is considered [22]

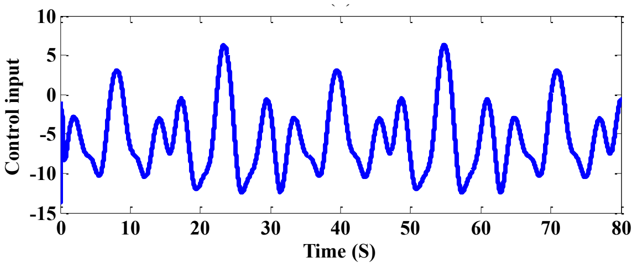
$$\ddot{\theta} = -I^{-1}(2\dot{\theta} + mgL\cos\theta) + I^{-1}\tau. \tag{49}$$

In (49), θ, $\dot{\theta}$, and $\ddot{\theta}$ denote the position, velocity and acceleration, and τ represents the applied control input to the joint. Other physical parameters and their numerical values for simulation have been given in Table II.

The control objective is to design a control input such that the system output tracks the desired signal y_d in the presence of FDI attack in "controller to actuator" channel. For simulation, initial conditions were set to $\boldsymbol{x}_0 = [0.2\pi \quad 0]^T$, $\zeta_0 = 0$, and $\hat{\theta}_0 = [0 \quad 0]^T$, and design parameters were chosen as $c_1 = 2$, $c_2 = 0.5$, $\omega_n = 40$, $\xi = 0.6$, $\gamma_1 = \gamma_2 = 0.5$, and $\sigma_1 = \sigma_2 = 0.03$. Also, Nussbaum-type function was chosen as $N(\kappa) = \exp(\kappa^2)\cos\left(\left(\frac{\kappa}{2}\right)\pi\right)$. The RBF-NN contains 25 nodes with centers χ_i, $i=1,\dots,25$ evenly spaced in the range $[-2,2]\times[-2,2]$, and the width of the basis function is set to $\eta_i = 2$ for $i=1,\dots,25$. Figure (3) shows the output response of the proposed controller for tracking $y_d = \sin t + \sin 0.2t$ under no FDI attack. As it is obtained from the



(a)



(b)

Fig. 4. (a) Output response, and (b) Control input under no FDI attack.

figure, the output variable tracks the desire output accurately and the control input is bounded.

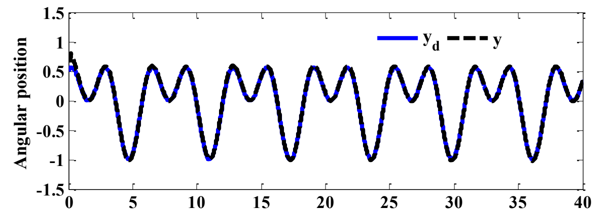
Now, performance of the proposed scheme against three different FDI attacks is simulated and discussed.

Case 1: In this case, a simple attack defined in (50) is applied to the system. Figure 4 shows the considered unlimited duration attack in (50):

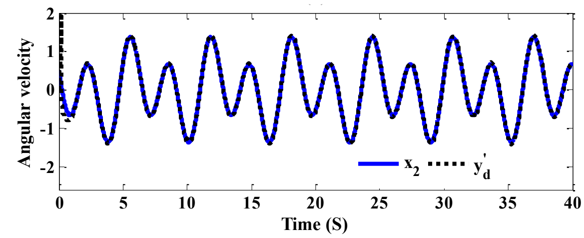
$$u = \begin{cases} v_n & t < 10 \\ 0.6v_n + 30 & 10 \leq t \leq 25 \\ 0.8v_n - 40\cos 0.3\pi t & t \geq 25 \end{cases} \tag{50}$$

Figure 5 displays the angular position and velocity of the robotic arm under FDI attack in (50). Figure 6 shows the actual and under attack control inputs, and Fig. (7) depicts adaptive parameters. As it is obvious from the results, proposed resilient control scheme tolerates the cyber-attack and preserves tracking ability of the closed-loop system under FDI attack, properly. As it is seen from the results, all of the closed-loop signals are bounded.

Case 2: In this case, a random stochastic attack with high switching frequency defined in (51) is injected into the "controller to actuator" communication channel. Figure 8 represents the considered attack. As it is seen form Fig. 8, at time interval 20 22 only additive term is non-zero, at intervals 15 20 and 28 35, only multiplicative attack is nonzero and in other intervals, both of additive and multiplicative terms are non-zero. Figure 9 represents the output response and the tracking error in presence of stochastic attack given in (51). It is clear from Fig. 9(a) that the output results can track the desired trajectory in presence of random attack. Also, results in Fig. 9(b) indicate that the norm of the tracking error is bounded. Obtained results verify that the proposed method is able to mitigate simultaneous

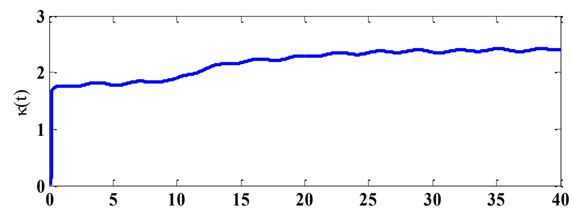


(a)

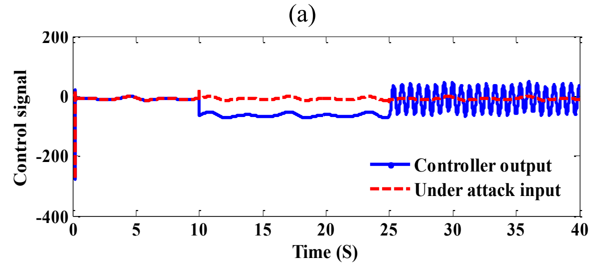


(b)

Fig. 5. (a) Angular position, and (b) Angular velocity of the robotic arm under FDI attack (case 1).



(a)



(b)

Fig. 6. (a) κ, (b) Controller output and under attack input. (case1).

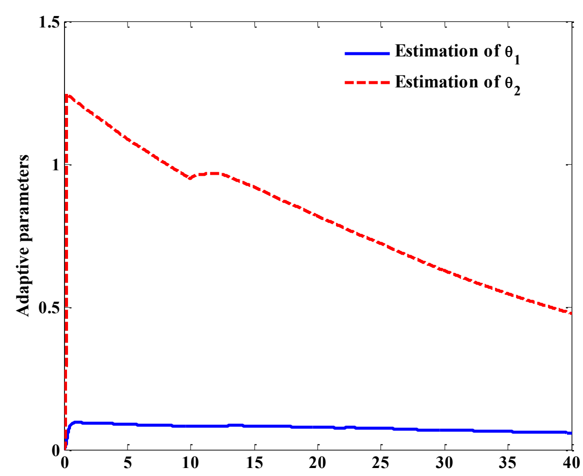
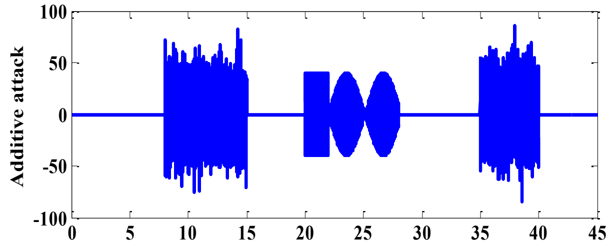
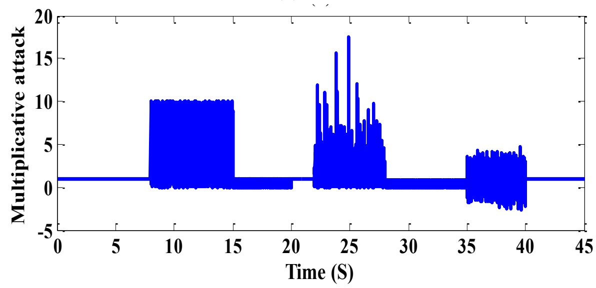


Fig. 7. Adaptive parameters (case 1).

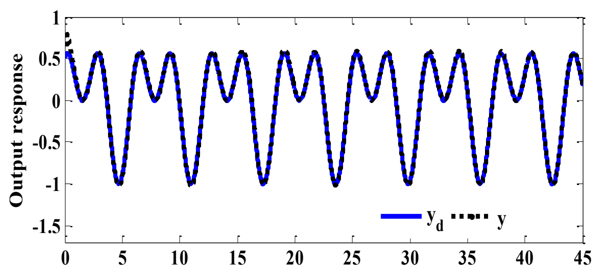


(a)

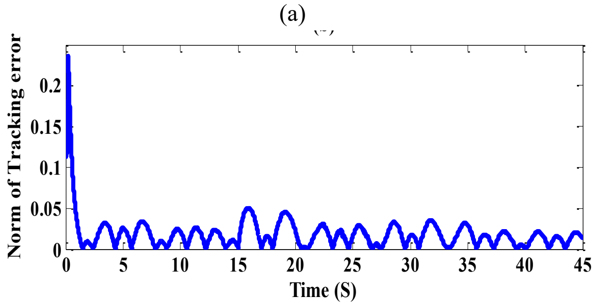


(b)

Fig. 8. Stochastic attack, (a) Additive term, and (b) Multiplicative term (case 2).



(a)



(b)

Fig. 9. (a) Output response, and (b) Norm of the tracking error (case 2)

additive and multiplicative attacks and it is not dependent on the attack function. This ability helps the system to defend against the total intrusion of an attacker regardless of the attack variations. The other advantage of the proposed approach is elimination of the need for attack reconstruction and control reconfiguration in presence of attacks. Therefore, the complexity and cost of control design will be decreased.

Case 3: In this case the following time-dependent attack is considered. Figures 10 and 11 illustrate the considered attack, output response and norm of the tracking error.

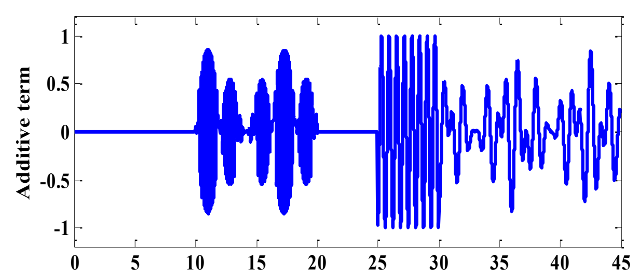
As it is seen from Fig. 11, the output signal tracks the desired trajectory and norm of the tracking error is bounded. Therefore, the proposed scheme is resilient against time-varying attack.

Simulation results in Figs. 4-11 verify that boundedness of the output variable and the tracking error is achieved regardless of the FDI attack type at "controller to actuator" channel.

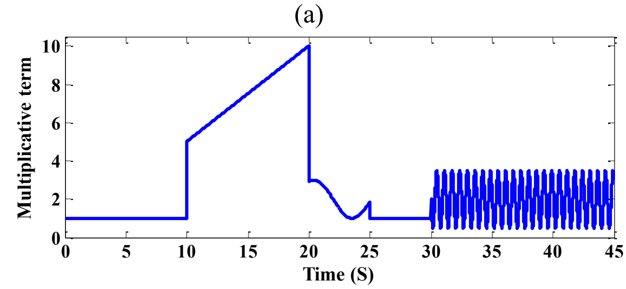
$$u = \begin{cases} v_n & t \le 8 \\ 10\text{rand}(.)v_n + 20\text{randn}(.) & 8 < t \le 15 \\ \text{rand}(.)v_n & 15 < t \le 20 \\ v_n + 40\cos(30\pi\text{randn}(.)) & 20 < t \le 22 \\ 0.5\exp(\text{-randn}(.))v_n \\ +40\cos(30\pi t)\sin(t) & 22 < t \le 28 \\ 0.8\text{rand}(.)v_n & 28 < t \le 35 \\ (1\text{+randn}(.))v_n \\ +\ 20\text{randn}(.) & 35 < t \le 40 \\ v_n & 40 < t \end{cases} \quad (51)$$

Obtained results show that the proposed approach could mitigate different kind of FDI attacks without using attack reconstruction or controller reconfiguration units.

$$u = \begin{cases} v_n & t \le 10 \\ sin(t)cos(30x_2) \\ +0.5tv_n & 10 < t \le 20 \\ (2 + sint)v_n & 20 < t \le 25 \\ v_n + sin(10t) & 25 < t \le 30 \\ \big(2 - 1.5cos(10t)\big)v_n \\ +sin(x_1)cos(2\pi t) & t > 30 \end{cases} \quad (52)$$

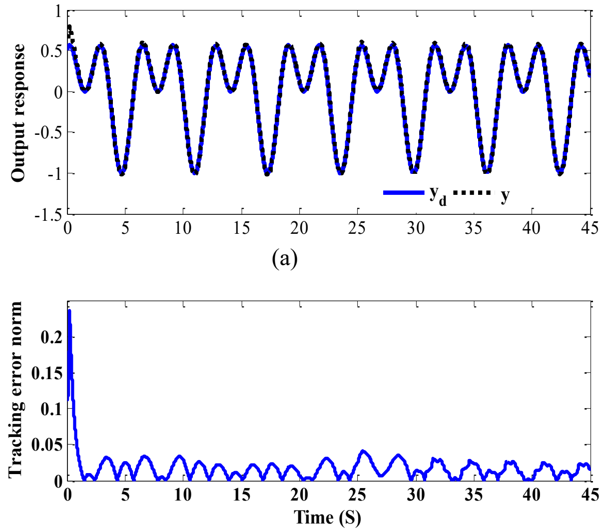


(a)



(b)

Fig. 10. Time-varying attack, (a) additive term, and (b) multiplicative term (case 3).



(a)

(b)

Fig. 11. (a) Output response, and (b) norm of the tracking error (case 3).

VI. Conclusion

This paper proposes an adaptive resilient control scheme for uncertain nth order nonlinear CPSs under injection attack. The main findings are: (1) an adaptive command filtered backstepping-based control scheme is able to handle the injection attack, and also it guarantees that the tracking error can be made small by adjusting the design parameters; (2) the proposed scheme eliminates the "explosion of complexity" problem, removes challenges in choosing time-constant of filters and compensates the filtering error, simultaneously, (3) the proposed scheme is resilient against both additive and multiplicative forms of cyber-attacks without any prior knowledge about the attack time, severity and characteristics, (4) it does not require attack detection and estimation units separately, as well as controller readjustment or reconfiguration algorithm, and (5) it decreases the number of adjustable parameters and online computational burden by using MLP algorithm. Therefore, the proposed approach maintains system stability and performance despite deceptive attack without noticeably increasing the complexity or imposing computational burden to design and implementation of the control system. However, the proposed passive resilient control offers a simpler and potentially more robust approach to mitigate attacks, but it may have limitations in terms of performance, adaptability in the face of some unforeseen disturbances or changes in the operating condition of the systems. Obtained simulation results verify the effectiveness and appropriate performance of the proposed resilient control scheme.

Extending the proposed scheme to deal with nonlinear CPSs in the presence of sensor attack can be considered as a future work.

References

[1] A. Humayed, J. Lin, F. Li, and B. Luo, "Cyber-Physical Systems Security- A Survey," *IEEE Internet of Things Journal*, vol. 4, no. 6, pp. 1802-1831, Dec. 2017, doi: 10.1109/JIOT.2017.2703172.

[2] Z. Lian, P. Shi, and M. Chen, "A Survey on Cyber-Attacks for Cyber-Physical Systems: Modeling, Defense, and Design," IEEE Internet of Things Journal, vol. 12, no. 2, pp. 1471-1483, Jan. 2025, doi: 10.1109/JIOT.2024.3495046.

[3] S. Chase Hassler, U. Ahmad Mughal, and M. Ismail, "Cyber-Physical Intrusion Detection System for Unmanned Aerial Vehicles," IEEE Transactions on Intelligent Transportation Systems, vol. 25, no. 6, pp. 6106-6117, June 2024, doi: 10.1109/TITS.2023.3339728.

[4] M.K. Hasan, A.A. Habib, Z. Shukur, F. Ibrahim, S. Islam, and M.A. Razzaque, "Review on Cyber-Physical and Cyber-Security System in Smart Grid: Standards, Protocols, Constraints, and Recommendations," Journal of Network and Computer Applications, vol. 209, p. 103540, Jan. 2023, doi: 10.1016/j.jnca.2022.103540.

[5] M. Javaid, A. Haleem, R.P. Singh, and R. Suman, "An Integrated Outlook of Cyber–Physical Systems for Industry 4.0: Topical Practices, Architecture, and Applications," Green Technologies and Sustainability, vol. 1, no. 1, p. 100001, Jan. 2023, doi: 10.1016/j.grets.2022.100001.

[6] M. Nazifi, and M. Pourgholi, "Adaptive Fractional-Order Consensus Control of Cyber-Physical Power Systems in the Presence of Unbounded Perturbations," International Journal of Industrial Electronics Control and Optimization, vol. 8, no. 2, pp. 105-116, June 2025, doi: 10.22111/ieco.2024.48807.1571.

[7] W. Duo, M. Zhou, and A. Abusorrah, "A Survey of Cyber-Attacks on Cyber-Physical Systems: Recent Advances and Challenges," IEEE/CAA Journal of Automatica Sinica, vol. 9, no. 5, pp. 784-800, May 2022, doi: 10.1109/JAS.2022.105548.

[8] Z. Yu, H. Gao, X. Cong, N. Wu, and H.H. Song, "A Survey on Cyber–Physical Systems Security," IEEE Internet of Things Journal, vol. 10, no. 24, pp. 21670-21686, Dec. 2023, doi: 10.1109/JIOT.2023.3289625.

[9] L. Hu, Z.D. Wang, Q.L. Han, and X.H. Liu, "State Estimation under False Data Injection Attacks: Security Analysis and System Protection," Automatica, vol. 87, no. 2, pp. 176–183, Jan. 2018, doi: 10.1016/j.automatica.2017.09.028.

[10] D. Zhang, Q.G. Wang, G. Feng, Y. Shi, and A.V. Vasilakos, "A Survey on Attack Detection, Estimation and Control of Industrial Cyber–Physical Systems," ISA Transactions, vol. 116, pp. 1-16, Oct. 2021, doi: 10.1016/j.isatra.2021.01.036.

[11] D. Zhang G. Feng,Y. Shi, and D. Srinivasan, "Physical Safety and Cyber Security Analysis of Multi-Agent Systems: A Survey of Recent Advances," IEEE/CAA Journal of Automatics Sinica, vol. 8, no. 2, pp. 319-333, Feb. 2021, doi: 10.1109/JAS.2021.1003820.

[12] S. Tan, J.M. Guerrero, P. Xie, R. Han, and J.C. Vasquez, "Brief Survey on Attack Detection Method for Cyber-Physical Systems," IEEE Systems Journal, vol. 14, no. 4, pp. 5329-5339, Dec. 2020, doi: 10.1109/JSYST.2020.2991258.

[13] D. Ding, Q-L. Han, Y. Xiang, X. Ge, and X-M. Zhang, "A Survey on Security Control and Attack Detection for Industrial Cyber-Physical Systems," Neurocomputing, vol.

275, pp. 1674-1683, Jan. 2018, doi: 10.1016/j.neucom.2017.10.009.

[14] M. Kordestani, and M. Saif, "Observer-Based Attack Detection and Mitigation for Cyber-Physical Systems–A Review," IEEE Systems, Man, and Cybernetics Magazine, vol. 7, no. 2, pp. 35-60, Apr. 2021, doi: 10.1109/MSMC.2020.3049092.

[15] M. Shahriari-kahkeshi, S.A. Alem, and P. Shi, "Detection, Reconstruction and Mitigation of Deception Attacks in Nonlinear Cyber-Physical Systems," International Journal of Adaptive Control and Signal Processing, vol. 38, no. 9, pp. 2972-2995, Sep. 2024, doi: 10.1002/acs.3854.

[16] F. Farivar, M. Sayad Haghighi, A. Jolfaei, and M. Alazab, "Artificial Intelligence for Detection, Estimation, and Compensation of Malicious Attacks in Nonlinear Cyber Physical Systems and Industrial IoT," IEEE Transactions on Industrial Informatics, vol. 16, no. 4, pp. 2716-2725, 2020, doi: 10.1109/TII.2019.2956474.

[17] M. Jahangiri-Heidari, M. Shahriari-kahkeshi, and P. Shi, "Resilient Consensus of Nonlinear Multiagent Systems under False Data Injection Attack on Communication Channels: An Attack Detection and Isolation-based Approach," IEEE Internet of Things Journal, vol. 12, no. 7, pp. 8219-8230, Apr. 2025, doi: 10.1109/JIOT.2024.3500587.

[18] A. Sargolzaei, K. Yazdani, A. Abbaspour, C.D. Crane, and W.E. Dixon, "Detection and Mitigation of False Data Injection Attacks in Networked Control Systems," IEEE Transactions on Industrial Information, vol. 16, no. 6, pp. 4281-4292, June 2020, doi: 10.1109/TII.2019.2952067.

[19] R. Huang, and Y. Li, "Adversarial Attack Mitigation Strategy for Machine Learning-Based Network Attack Detection Model in Power System," IEEE Transactions on Smart Grid, vol. 14, no. 3, pp. 2367 – 2376, May 2023, doi: 10.1109/TSG.2022.3217060.

[20] S. Lü, X. Jin, L. Ding, and Q. Tan, "Adaptive Sliding-Mode Control of a Class of Disturbed Cyber–Physical Systems Against Actuator Attacks," Computers and Electrical Engineering, vol. 96, p. 107492, Dec. 2021, doi: 10.1016/j.compeleceng.2021.107492.

[21] S.J. Yoo, "Neural-Network-Based Adaptive Resilient Dynamic Surface Control Against Unknown Deception Attacks of Uncertain Nonlinear Time-delay Cyber-Physical Systems," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 4341-4353, Oct. 2020, doi: 10.1109/TNNLS.2019.2955132.

[22] Y. Zhao, C. Zhou, and Y.C. Tian, "Anti-Saturation Resilient Control of Cyber-Physical Systems under Actuator Attacks," Information Sciences, vol. 608, pp. 1245-1260, Aug. 2022, doi: 10.1016/j.ins.2022.07.010.

[23] Y. Yang, J. Huang, X. Su, and B. Deng, "Adaptive Control of Cyber-Physical Systems under Deception and Injection Attacks," Journal of the Franklin Institute, vol. 358, no. 12, pp. 6174-6194, Aug. 2021, doi: 10.1016/j.jfranklin.2021.06.008.

[24] Y. Yang, J. Huang, X. Su, K. Wang, and G. Li, "Adaptive Control of Second-Order Nonlinear Systems with Injection and Deception Attacks," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 52, no. 1, pp. 574-581, Jan. 2022, doi: 10.1109/TSMC.2020.3003801.

[25] Y. Zhao, X. Du, C. Zhou, Y.C. Tian, X. Hu, and D.E. Quevedo, "Adaptive Resilient Control of Cyber-Physical Systems under Actuator and Sensor Attacks," IEEE Transactions on Industrial Informatics, vol. 18, no. 5, pp. 3203 – 3212, May 2022, doi: 10.1109/TII.2021.3108876.

[26] S. Liu, X. Wang, B. Niu, X. Song, H. Wang, and X. Zhao, "Adaptive Resilient Output Feedback Control Against Unknown Deception Attacks for Nonlinear Cyber-Physical Systems," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 71, no. 8, pp. 3855 – 3859, Aug. 2024, doi: 10.1109/TCSII.2024.3372413.

[27] M. Mazare, "Reinforcement Learning-based Fixed-time Resilient Control of Nonlinear Cyber Physical Systems Under False Data Injection Attacks and Mismatch Disturbances," Journal of the Franklin Institute, vol. 360, no. 18, Dec. 2023, pp. 14926 – 14938, doi: 10.1016/j.jfranklin.2023.10.026.

[28] J. Zhao, and G-H. Yang, "Reinforcement-Learning-Based Fuzzy Adaptive Finite-Time Optimal Resilient Control for Large-Scale Nonlinear Systems Under False Data Injection Attacks," IEEE Transactions on Fuzzy Systems, vol. 32, no. 4, Apr. 2024, pp. 2483-2495, doi:10.1109/TFUZZ.2023.3343722.

[29] Y-S. Ma, W-W. Che, C. Deng, and Z-G. Wu, "Model-Free Adaptive Resilient Control for Nonlinear CPSs With Aperiodic Jamming Attacks," IEEE Transactions on Cybernetics, vol. 53, no. 9, Sep. 2023, pp. 5949-5956, doi: 10.1109/TCYB.2022.3219987.

[30] S. Gao, L. Liu, H. Wang, and A. Wang, "Data-Driven Model-Free Resilient Speed Control of an Autonomous Surface Vehicle in the Presence of Actuator Anomalies," ISA Transactions, vol. 127, Aug. 2022, pp. 251-258, doi: 10.1016/j.isatra.2022.04.050.

[31] S-S. Sun, and Y-X. Li, "Data-Driven Adaptive Resilient Funnel Consensus Tracking of Multi-Agent Systems Under Jamming Attacks," International Journal of Robust and Nonlinear Control, Aug. 2024, doi: 10.1002/rnc.7593.

[32] A. Azarbahram, N. Pariz, M-B. Naghibi-Sistani, and R. Kardehi Moghaddam, "Leader-Follower Formation Control of Uncertain USV Networks Under Stochastic Disturbances," International Journal of Industrial Electronics Control and Optimization, vol. 5, no. 2, pp. 133-142, Apr. 2022, doi: 10.22111/ieco.2022.37987.1346.

[33] F. Nobakht, H. Eliasi, "Adaptive Backstepping Control of Two-Group SEIAR Epidemic Model in the Presence of Input Saturation and External Disturbances," International Journal of Industrial Electronics Control and Optimization, In Press, April 2025, doi: 10.22111/ieco.2025.50181.1632.

[34] M. Krstic, I. Kanellakopoulos, and P.V. Kokotovic, Nonlinear and Adaptive Control Design, Wiley, New York, NY, USA, 1995.

[35] X.D. Tang, G. Tao, and S.M. Joshi, "Adaptive Actuator Failure Compensation for Nonlinear MIMO Systems with an Aircraft Control Application," Automatica, vol. 43, no. 11, pp. 1869-1883, 2007, doi: 10.1016/j.automatica.2007.03.019.

[36] Y.M. Li, S.C. Tong, and T.S. Li, "Adaptive Fuzzy Output Feedback Control for a Single-Link Flexible Robot Manipulator Driven DC Motor via Backstepping", Nonlinear Analysis: Real World Applications, vol. 14, no. 1, pp. 483-494, 2013, doi: 10.1016/j.nonrwa.2012.07.010.

[37] F.J. Lin, H.J. Shieh, and P.K. Huang, "Adaptive Wavelet Neural Network Control with Hysteresis Estimation for Piezo–Positioning Mechanism," IEEE Transactions on Neural Networks, vol. 17, no. 2, pp. 432-444, 2006, doi: 10.1109/TNN.2005.863473.

[38] M.M. Polycarpou, and J. Mears, "Stable Adaptive Tracking of Uncertainty Systems Using Nonlinearly Parameterized On-line Approximators," International Journal of Control, vol. 70, no. 3, pp. 363–384, Jan. 1998, doi: 10.1080/002071798222280.

[39] R.M. Sanner, and J.E. Slotine, “Gaussian Networks for Direct Adaptive Control,” IEEE Transactions on Neural Networks, vol. 3, no. 6, pp. 837–863, Nov. 1992, doi: 10.1109/72.165588.

[40] J.A. Farrell, M. Polycarpou, M. Sharma, and W. Dong, “Command Filtered backstepping,” 2008 American Control Conference, Washington, USA, June 11-13, 2008.

[41] R.D. Nussbaum, “Some Remarks on the Conjecture in Parameter Adaptive Control,” Systems & Control Letter, vol. 3, no. 5, pp. 243–246, Nov. 1982. doi: 10.1016/0167-6911.

[42] X.D. Ye, and J.P. Jiang, “Adaptive Nonlinear Design without a Prior Knowledge of Control Directions,” IEEE Transactions on Automatic Control, vol. 43, no. 11, pp. 1617–1621, Nov. 1998, doi: 10.1109/9.728882.

[43] Y. Yang, and J. Ren, “Adaptive fuzzy robust tracking controller design via small gain approach and its application,” IEEE Transactions on Fuzzy Systems, vol. 11, no. 6, pp. 783-795, Dec. 2003, doi: 10.1109/TFUZZ.2003.819837.

[44] W. J. Dong, J. A. Farrell, M. Polycarpou, V. Djapic, and M. Sharma,“Command Fitered Adaptive Backstepping,” IEEE Transactions on Control Systems Technology, vol. 20, no. 3, pp. 566-580, May. 2012, doi: 10.1109/TCST.2011.2121907.

[45] J. Yu, P. Shi, C. Lin, and H. Yu, “Adaptive Neural Command Filtering Control for Nonlinear MIMO Systems With Saturation Input and Unknown Control Direction,” IEEE Transactions on Cybernetics, vol. 50, no. 6, pp. 2536-2545, June 2020, doi: 10.1109/TCYB.2019.2901250.

Maryam Shahriari-kahkeshi received her M.Sc. and Ph.D degrees in control engineering from Isfahan University of Technology, Isfahan, Iran in 2010 and 2014, respectively. She is currently an Associate Professor with the Faculty of Engineering, Shahrekord University, Shahrekord, Iran. Her research interests include artificial intelligence, nonlinear control systems, cyber-physical systems, fault tolerant control systems, and resilient control systems design.

Seyed Hojat Nourian received his BSc. in Electrical Engineering from Water and Electricity Industry Training Complex, Isfahan, Iran in 2013, and MSc. in Control Engineering from Shahrekord University, Shahrekord, Iran in 2025. His research interests are cyber-physical systems, cyber-attacks and resilient control systems under cyber-attacks.

International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Designing a Rehabilitation Traction Device for Decreasing Low Back Pain and Spinal Decompression by Improving Traction Control

Mahdi Alinaghizadeh Ardestani [1] | Parham Haji Ali Mohamadi [2]

Department of Electrical Engineering, Technical and Vocational University (TVU), Tehran, Iran [1]
Department of Electrical Engineering, Islamic Azad University, Science and Research Branch, Tehran, Iran [2]
Corresponding author's email: ardestani@tvu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 14-February-2025
Received in revised form: 13-July-2025
Accepted: 27-July-2025
Published online: 23-Sep-2026

Keywords:
Iterative learning control, discrete nonlinear systems, low back rehabilitation,

ABSTRACT

This paper presents the design and control of a novel cable-driven parallel robot intended for low back pain rehabilitation through controlled spinal traction therapy. Low back pain affects a significant portion of the global population, and traction-based interventions offer a non-invasive approach to spinal decompression. The proposed robotic system enables precise movement of the patient's legs in both the frontal and sagittal planes—up to 50° of extension and ±15° of side bending—while applying adjustable traction forces to the lumbar spine. A dynamic model of the system is derived using Lagrangian mechanics, assuming planar motion, massless cables, and a single rotational degree of freedom at the hip joint. To enhance tracking accuracy and robustness across repeated therapy sessions, a new Current Iterative Learning Control (CILC) law is developed for discrete-time nonlinear systems. The CILC integrates real-time feedback with iterative feedforward learning, ensuring convergence of system states and outputs to desired trajectories despite disturbances, measurement noise, and initial condition errors. Theoretical analysis confirms that the tracking error remains bounded and asymptotically decreases with successive iterations. Simulation studies validate the controller's performance, demonstrating superior trajectory tracking compared to conventional PID control in terms of mean square error. The results highlight the CILC's ability to adapt over trials, significantly improving precision in rehabilitation tasks. While the current work is simulation-based, future efforts will focus on prototyping, experimental validation with human subjects, and integration of real-time sensor feedback for adaptive therapy personalization. This research contributes to the advancement of intelligent, patient-specific rehabilitation robotics by combining biomechanically informed design with a robust, learning-based control framework tailored for cable-driven traction systems.

I. Introduction

Lumbar disc herniation, cervical spine disorders, and joint pain are some of the most common orthopedic diseases. It is estimated that 85% of people suffer from low back pain and other spinal pains at some point in their lives [1]. Like many other orthopedic diseases, treatment for these pains usually requires a combination of medicine and physical therapy.

In physiotherapy, there are various therapeutic methods, including the usage of electrical and mechanical rehabilitation equipment as well as employing different therapeutic exercises that can reduce or eliminate the patients' pain. Traction therapy, or in other words, non-surgical spinal decompression therapy, is one of the most common non-invasive methods for relieving low back pain and refers to any medical procedure that applies a force along the inferior-superior axis of the spine to extend the cervical and lumbar spine vertebrae [2]. This extension leads to relieving pressure on the spine and its surrounding nerves.

There is a wide range of investigations on different aspects of traction therapy. In [3] an experimental model of the spine is used to study the effects of intermittent traction therapy on vertebral interspaces. Moreover, the effects of a two-step traction therapy that includes global traction and a local decompression on lumbar spine biomechanics are reviewed in [4]. Most researches in this field show that traction therapy, especially when a motorized device is used as the rehabilitation equipment for applying force to the spine and lower back of patients with chronic pain, leads to pain reduction. As an example, Apfel et al. have shown that the chronic low back pain attributed to disc herniation decreased from 6.2 ± 2.2 to 1.6 ± 2.3 in a verbal rating scale for 30 patients who had used a spinal decompression device in 6 weeks [5]. However, due to limited clinical trials and insufficient evidence from published papers, the U.S. Food and Drug Administration (FDA) does not confirm improved health outcomes for using these rehabilitation devices over other technologies [6]. To address this concern, our future work will focus on gathering empirical data through prototype implementation and pilot testing.

There are several ways that a therapist can apply a traction force to the patient's spine. The first and the oldest one is manual traction where the physical therapists use their power and specific hands-on techniques. Positional traction, inversion traction, and self-traction techniques are some other methods that utilize patient power or body weight and gravity to apply traction force. [7] and [8] are such devices. On the other hand, there are mechanical traction methods that require special (electro) mechanical equipment to apply the traction force to the patient's spine. In such rehabilitation devices, traction force magnitude and its direction should be set by the therapist. A lumbar stretching machine is introduced in [9]. Furthermore, according to force duration, mechanical traction can be divided into three different classes: constant traction, intermittent traction, and harmonic traction. The last two ones need proper motorized equipment to be applied. Also, another important issue is controlling the applied force, so there should be a feedback signal from the traction force on the patient's spine. This structure is shown in Figure 1.

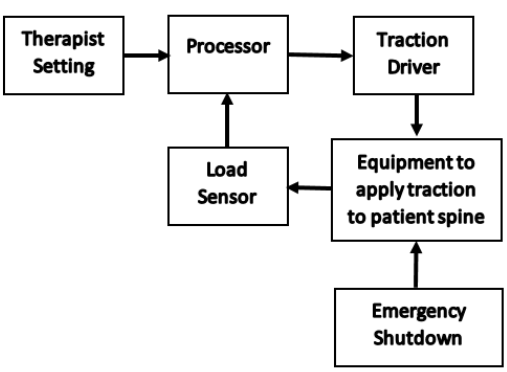


Figure 1 Structure of an electromechanical rehabilitation traction device

The structure shown in Figure 1 is implemented in [10] for a cervical spine traction device and in [11] for lumbar spine traction. Moreover, the Satisform company introduces a commercial device for relaxing the pelvis and effective stretching of the lower back [12].

Recent advances in intelligent rehabilitation robotics and biomechanical modeling have enabled more precise and adaptive traction control. Studies such as [33] on model predictive control for wearable rehabilitation devices, [34] on neural network-based estimation of spinal load, and [35] on real-time feedback mechanisms in robotic-assisted physiotherapy demonstrate the importance of integrating modern control techniques with biomechanics. We position our contribution within this context by proposing a novel CILC framework tailored specifically for cable-driven traction systems.

II. Structure design

Rehabilitation robots should be specifically designed for pain reduction and improvement of patients' various physical problems. Sometimes injuries are inordinate and the patient is unable to perform certain movements and activities so in these situations, the rehabilitation robot is an accessory that helps the patient to do those movements.

Generally, robots that are used in the medical field and rehabilitation should be able to interact safely so they should be precise in position control and have a controlled and limited acceleration [13]. In addition, the rehabilitation robot should not be bulky so it can be placed in clinics, hospitals, and also in the patient's house. For these purposes, a cable robot is proposed in this paper.

In designing robots many elements should be considered to lead to a product. Some of these elements are mechanical structures, actuators, sensors, control systems, and so on. In this paper, we focused on mechanical design and control systems. In the end, we will have a parallel cable robot that performs the required movements.

A. Proposed structure

The proposed structure assumes planar motion for kinematic simplification. However, we acknowledge that this is a limitation and may not fully capture three-dimensional anatomical movements. Future work will incorporate full 3D modeling and dynamic analysis based on realistic human body kinematics. Furthermore, the assumptions regarding massless cables and perfect actuation should be interpreted with caution. In practice, cable elasticity, backlash, and actuator dynamics will affect performance. These limitations are discussed in detail below and will be addressed in the experimental phase of this project. The parallel cable robot, which is presented in this section, can be used in designing tractive rehabilitation equipment. This device moves the patient's feet in a particular trajectory meanwhile pulling them. Moreover, it can control the traction force. Therapists should set the desired trajectory by considering each patient's injury. Ranges of the lumbar spine movement are studied in [14] for 100 healthy subjects aged 20 to 60+ years. Results show subjects can have approximately 30–40° extension and 20–25° side bending. Since these boundaries are for healthy individuals, our proposed robot can have reduced limits due to patient conditions: we assume up to 50° extension and ±15°

side bending in this paper. Now, in the beginning, we should choose the number of cables that will be used in our design and their configuration. Here we have two cables that are separately connected to pulleys and motors at one end and an ankle brace at the other end. Also, we need a frame that is assembled to a bed where the patient lies and actuators can be placed on and connected to the ankle brace with cables. This frame rotates up to 50 degrees in the patient's frontal axis. The ankle brace tied the patient's foot together so when the actuators change the lengths of the cables, the patient's leg will be moved meanwhile the traction force is conveyed to the lumbar spine. This structure is shown in Figure 2.

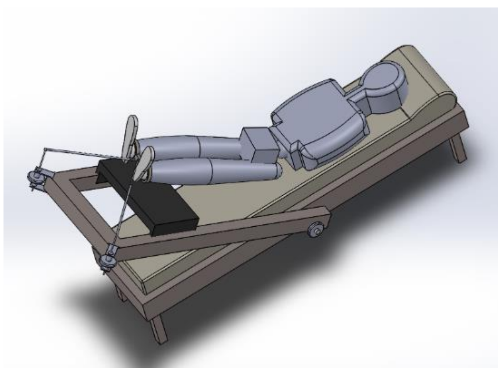


Figure 2 proposed rehabilitation traction structure.

B. Kinematics and dynamics of the proposed structure

Forward and inverse kinematics of the mechanism are calculated in this section so we can derive the dynamic model of the robot, which will be utilized to develop its control system.

Although in the suggested structure the robot and consequently the patient's leg have a three-dimensional movement, we can analyze the kinematics and dynamics in a planar space. This assumption is valid when the actuated cables and the patient's foot are on the same plane. For this purpose, we can put a portable plate under the patient's foot, otherwise, it can be shown that actuators should generate a great traction force that is impossible to be applied to the foot and lumbar spine. The designed structure schematic which is employed for kinematics and dynamics analysis, is illustrated in Figure 3 and Figure 4 .In the designed structure schematic patient's legs are combined and it is supposed that the combination acts as a spherical joint and can be shown as a line segment. Also, we assume that the origin of the coordinate system is on the spherical joint center (where the legs and hip join the lumbar spine). The two actuators are placed symmetrically at points A1 and A2. The cables are connected between the actuators and the ankle brace at B. Now if we revolve the holder frame around the x-axis, the patient's leg and hip will be rotated α degrees, which is equal to a α-degree extension. Now if one of the actuators generates more torque than the other one, the lengths of the cables will change and as a result, the leg position will be changed. This motion is identical to a γ-degree side bending.

In Figure 4, S1 and S2 are unit vectors along with the cables that connect the actuators' frames at A1 and A2 to the end effector at point B (ankle brace). So the magnitude of $\overrightarrow{(A_iB)}$ indicates each cable length(L_i).

A cable-driven robot is a closed kinematics chain mechanism and usually, the inverse kinematics of parallel robots can be calculated more simply versus its forward kinematics [15]. Inverse kinematics relates cables' length to the end-effector position.

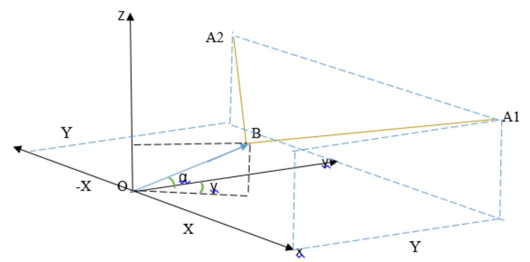


Figure 3 . Designed structure schematic perspective view

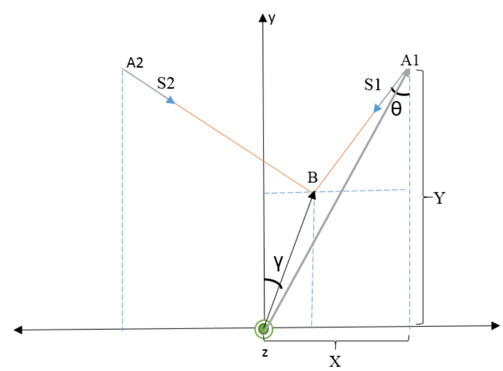


Figure 4. Designed structure schematic top view

Now assume we have the end-effector position (point B) by utilizing any position sensor so we can figure out inverse kinematics as below:

$$\overrightarrow{A_iB}+\overrightarrow{OA_i}=\overrightarrow{OB} \quad (1)$$

So:

$$L_i\vec{S}_i + R^A_{A_0}A_0 = R^B_{B_0}\,B_0 \quad (2)$$

Where A_0 & B_0 are the initial positions of the actuator's holder frame and the end effector. $R^A_{A_0}$ and $R^B_{B_0}$ are rotational matrixes. Duo to the designed mechanism we have:

$$\mathrm{R}^{\mathrm{B}}_{\mathrm{B}_0} = \mathrm{Rot_x}(\alpha)\mathrm{Rot_z}(\gamma) = \begin{Bmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \cos(\alpha)\sin(\gamma) & \cos(\alpha)\cos(\gamma) & -\sin(\alpha) \\ \sin(\alpha)\sin(\gamma) & \sin(\alpha)\cos(\gamma) & \cos(\alpha) \end{Bmatrix} \quad (3)$$

$$\mathrm{R}^{\mathrm{A}}_{\mathrm{A}_0} = \mathrm{Rot_x}(\alpha) = \begin{Bmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{Bmatrix}$$

So each cable length is:

$$L_i = \sqrt{\left(R^B_{B_0}\,B_0 - R^A_{A_0}A_0\right)^T\left(R^B_{B_0}\,B_0 - R^A_{A_0}A_0\right)} \quad (3)$$

Additionally, in the kinematics chain unit vectors, S1 and S2 are unknown and calculated by:

$$\vec{S}_i = \frac{L_i S_i}{|L_i S_i|} = \frac{R^B_{B_0} B_0 - R^A_{A_0} A_0}{\left|R^B_{B_0} B_0 - R^A_{A_0} A_0\right|} \tag{4}$$

In parallel robots, forward kinematics is a relation that expresses the end-effector position based on the cables' lengths. Against the serial robots, the forward kinematics of parallel robots are usually complicated and include nonlinear equations, but here due to the robot structure end effector and cables being in one plane, we can reach a closed analytic solution.

Now to find the forward kinematic based on knowing one cable length and Figure 4 we can write:

$$\begin{cases} Y - L_L \cos(\gamma) = L_i \cos(\theta) \\ |X - L_L \sin(\gamma)| = l_i \sin(\theta) \end{cases} \tag{5}$$

Where L_L is the patient's leg length. So we obtain:

$$\sqrt{X^2 + Y^2} \sin(\gamma + \phi) = \frac{L_i^2 - Y^2 - X^2 - L_L^2}{-2L_L^2} \tag{6}$$

And $\phi = \arctan\left(\frac{Y}{X}\right)$.

Now we can derive the Jacobian matrix of the proposed rehabilitation robot. Jacobian matrix, J, relates the angular velocity of the end effector (patient leg rotation around the z-axis), $\omega = [0\ 0\ \dot{\gamma}]^T$, to the vector of actuated cables rates, $L = [L_1, L_2]^T$ by:

$$\dot{L} = J\omega \tag{7}$$

According to (2) by differentiating the kinematics vector loop closure concerning time:

$$\dot{L}_i \vec{S}_i + L_i \omega \times \vec{S}_i = \omega \times R^B_{B_0} B_0 \tag{8}$$

Dot multiply equation (8) by $\vec{S}_i$:

$$\dot{L}_i = (R^B_{B_0} B_0 \times \vec{S}_i).\omega = \left(R^B_{B_0} B_0 \times \vec{S}_i\right)^T \omega \tag{9}$$

Writing the above equation two times for each actuated cable in a vector form results in the Jacobian matrix J:

$$J = \begin{bmatrix} \left(R^B_{B_0} B_0 \times \vec{S_1}\right)^T \\ \left(R^B_{B_0} B_0 \times \vec{S_2}\right)^T \end{bmatrix} \tag{10}$$

Parallel rehabilitation traction robot has a non-square Jacobian matrix in which the number of rows indicates the total number of freedom degrees and system redundancy. Here the proposed structure has one rotational degree of freedom and one degree of freedom redundancy. In redundant parallel robots, singular value decomposition (SVD) of the Jacobian matrix is used for robot singularity analysis. Singularity occurs when the Jacobian matrix rank is less than the degrees of freedom number [16]. In the proposed structure Jacobian matrix rank is always one in all the desired trajectories so singularity never happens for it.

Now for deriving the dynamic equations in terms of generalized coordinates and their time derivatives, the Lagrangian method is used. For the sake of simplicity, the elasticity of cables is neglected. Furthermore, the cables' mass can be ignored versus the patient leg mass so cables behave as massless rigid strings [17]. Based on these assumptions dynamics of the robot can be given as:

$$\begin{aligned} M(x)\ddot{x} + N(x,\dot{x}) &= J^T\tau \\ N(x,\dot{x}) &= C(x,\dot{x})\dot{x} + G(x) + T_d \end{aligned} \tag{11}$$

where $x = [\alpha, \beta, \gamma]^T$ is the generalized coordinates' vector, τ is the cables force vector, $M(x)$ is the mass matrix, $C(x,\dot{x})$ represents Coriolis and centrifugal terms, $G(x)$ is gravity vector and T_d stands for the disturbance which could represent any friction or uncertainty in the dynamic model.

As we neglect cables' mass, the mass matrix $M(x)$ in (11) can be derived directly from the kinetic energy of the system. Also in the proposed rehabilitation robot, the patient's leg has only one rotational degree of freedom around the z-axis, so:

$$k = \frac{1}{2}\omega^T I \omega \rightarrow M(x) = I \tag{12}$$

where I is the leg rotational inertia matrix and $M(x)$ is symmetric and positive definite so it is an invertible matrix.

Generally, the determination of the Coriolis and centrifugal vector in the Lagrange equation (11) is complex and can be calculated by:

$$V(x,\dot{x}) = c(x,\dot{x})\dot{x} = \dot{M}\dot{x} - \frac{\partial}{\partial x}k \tag{13}$$

Hence the leg mass matrix $M(x)$ is a constant diagonal matrix $c(x,\dot{x})$ is equal to zero. The gravity term in (11) may be determined from the partial derivative of the moving leg potential energy concerning the generalized coordinate x.

$$G(x) = \frac{\partial p}{\partial x} = \begin{bmatrix} 0 \\ 0 \\ -mg\frac{L_l}{2}\sin(\alpha)\sin(\gamma) \end{bmatrix} \tag{14}$$

where m and g are the leg mass and gravity constant, respectively.

III. Control Design

As said in the previous sections, rehabilitation of a patient usually includes performing prescribed exercises. Thus, passive control strategies have been proposed for rehabilitation equipment control. Passive control can employ different techniques to be achieved. PID control is one of the most common methods [18]. Khosravi [19] developed a framework for robust PID control of parallel manipulators which are fully constrained. Implementation results on a planner cable robot show repeatable performance in different iterations, but there is no improvement versus the previous iteration.

A. ILC

If we define a control task as: *perfect tracking in a finite time interval under a repeatable control environment* [20], few control methods can fulfill this task. In fact, without a learning procedure controller performs only in the same way and there will be no improvement even if the task repeats consecutively. In these circumstances, Iterative Learning Control (ILC) can take the role perfectly.

Although ILC uses a learning procedure, it is different from other learning-based control methods such as adaptive control and neural networks. Adaptive control sets the controller while ILC regulates the control signal. In addition, adaptive control strategies usually do not use the informative data in the past trials' control signals. In the same way, neural networks modify the controller parameters. Also, they are usually large networks and need extensive training data [21].

Many references including [22] and [23] have provided comprehensive coverage for convergence, robustness, and applications of ILC. It is possible that before achieving the desired convergence, tracking error increases and therefore convergence would be slow. Usually, this is because ILC is an open-loop controller and can have weak performances for stable systems and even be bad for unstable plants. In such situations, adopting a feedback controller is useful [24]. The incorporation of feedback controller with ILC leads to a new ILC class which is called: Current-iteration Iterative Learning Control (CILC). In 1992 Tae-Yong et al. [25] presented a CILC for continuous nonlinear systems which composed a linear feedback mechanism and a feedforward learning strategy, utilizing the gradient descent method. In [26] a multi-period learning strategy is developed in conjunction with current cycle feedback for a continuous nonlinear dynamic system. Also, [27] designed a framework for PID-like ILC with a feedback structure for a class of nonlinear discrete-time NARMA systems. Additionally, [24] and [28] have developed CILC for nonlinear discrete-time systems (without and with initial condition error, consequently) based on an earlier work [29] which proposed an ILC learning law using one-step-ahead tracking error of the last trial for off-line calculation. Also, [30] & [31] proposed a CILC algorithm for high-precision trajectory tracking of a wheeled mobile robot with non-holonomic nonlinear dynamics.

B. Proposed CILC

Consider a class of discrete-time nonlinear dynamic systems that can be described by the following difference equations:

$$\begin{aligned} X_i(k+1) &= f(X_i(k),k) + B(X_i(k),k)u_i(k) + \omega_i(k) \\ y_i(k) &= CX_i(k) + v_i(k) \end{aligned} \tag{15}$$

Where i indicates the number of trials, k is a discrete-time index running from $k = 0$ to $k = n$. Also, for all $k \in N$, $X_i(k) \in R^p, u_i(k) \in R^r, y_i(k) \in R^m, v_i(k) \in R^m$ and $\omega_i(k) \in R^p$ are states, inputs, outputs, output noise, and disturbance, respectively. Vector/Matrix functions $f: R^p \times N \to R^p$, $B: R^p \times N \to R^{p\times r}, C \to R^{m\times p}$ satisfy following assumptions:

Assumption 1: Considering the discrete-time dynamic equations presented in (15), when $\omega(k) = 0 \; and \; v(k) = 0$ for all $k \in N$, there exist bounded sequences of states $X_d(k)$ and inputs $u_d(k)$ which will satisfy the equations (15) for the desired bounded trajectory $y_d(k)$.

Assumption 2: Vector function $f(X,k)$ and matrix function $B(X,k)$ satisfy:

$$\begin{aligned} \left\|f(X_1,k) - f(X_2,k)\right\| &\leq c_f\|X_1 - X_2\| \\ \left\|B(X_1,k) - B(X_2,k)\right\| &\leq c_B\|X_1 - X_2\| \end{aligned} \tag{16}$$

where c_f and c_B are positive constants.

Assumption 3: Matrices $B(X,k)$ and C are bounded as $\|B(X,k)\| \leq b_B$ and $\left\|C\right\| \leq b_C$ where b_B and b_C represent positive constants. Also, the product matrix $CB(X,k)$ yields a non-singular matrix.

Assumption 4: Output noise and disturbance are bounded for all $i \geq 0, k \in N$ as $\|\omega_i(k)\| \leq b_\omega$ and $\left\|v_i(k)\right\| \leq$, where $b_\omega$and b_v are positive constants.

Assumption 5: All trials start from a neighborhood of the desired initial condition $X_d(0)$. where for all $i \geq 0$, $\|X_d(0) - X_i(0)\| \leq b_{X_0}$ and $b_{X_0}$is a positive constant. This assumption is usually valid in operations where the robot (same as the proposed structure in the previous sections) performs a particular given task repeatedly.

Now consider the iterative learning control law which is proposed as follows:

$$\begin{aligned} &u_i(k) = uf_i(k) + ub_i(k) \\ &uf_i(k) = uf_{i-1}(k) + L(k)e_{i-1}(k+1) \\ &+\lambda(k)ub_{i-1}(k) \end{aligned} \tag{17}$$

where $ub_i(k)$ is feedback control, $uf_i(k)$ is the iterative learning term in feedforward, and $u_i(k)$ is the control signal for ith trial. Also, $e_i(k) = y_d(k) - y_i(k)$ and $L(k)$ $\lambda(k)$ are learning gain matrices that satisfy $\left\|L(k)\right\| \leq$ $\left\|\lambda(k)\right\| \leq b_\lambda$ for all $k \in N$. Where b_L and $b_\lambda$are positive constants. The proposed CILC algorithm is shown in Figure 5.

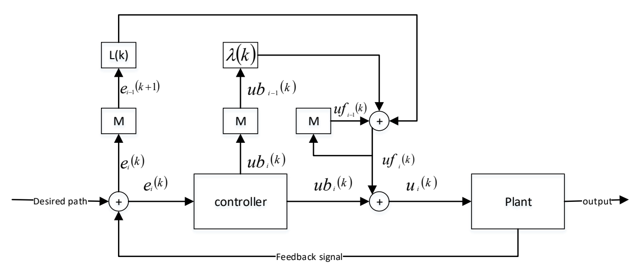


Figure 5 proposes the CILC algorithm.

Feedback controller $ub_i(k)$ can be designed as either static or dynamic feedback which stabilizes (15) and takes the following general form:

$$\begin{aligned} Z_i(k+1) &= p\big(Z_i(k)\big) + q(Z_i(k))e_i(k) \\ ub_i(k) &= r\big(Z_i(k)\big) + s(Z_i(k))e_i(k) \end{aligned} \tag{18}$$

where $Z_i(k)$ is a state vector. Vector or matrix functions p,q,r,s are defined suitably Which satisfy $\left\|p\big(Z_i(k)\big)\right\| \leq b_p\|Z_i(k)\|$, $\left\|r\big(Z_i(k)\big)\right\| \leq b_r\|Z_i(k)\|$, and $\left\|q\big(Z_i(k)\big)\right\| \leq b_q$, $\left\|s\big(Z_i(k)\big)\right\| \leq b_s$ for $b_p,b_r,b_q,b_s > 0$.

Theorem. Consider the discrete-time system (15) with the CILC proposed in (17). If assumptions 1-5 are satisfied and for $(X,k) \in R^p \times N$ we chose $L(k)$ as:

$$\left\|I - L(k)CB(X,k)\right\| = \rho < 1 \tag{19}$$

With increasing the trial number i, $i \to \infty$, it is guaranteed that the control signal $u_i(k)$ converges to the desired control signal $u_d(k)$. Consequently, states and system output, $X_i(k)$ and $y_i(k)$, will get to a neighborhood of their desired signals, $\lim_{i\to\infty}\|y_d(k)-y_i(k)\| \le \sigma$. This error boundary is a function of output noise bound b_v, disturbance bound b_ω and initial condition error band b_{x_0}.

Definition: The α-norm of a positive real function $q(.)$ is defined for $\alpha \ge 1$ as:

$$\|q(.)\|_\alpha = Sup_{k\in N}\, q(k)\left(\frac{1}{\alpha}\right)^k \tag{20}$$

The α-norm is equivalent to the ∞-norm defined as $\|q(.)\|_\infty = \sup_{k\in N} q(k)$. Therefore, the claims in the proof of the proposed theorem are driven by using the α-norm.

Proof of Theorem:

If we define: $\delta Z_i = Z_d - Z_i$ for all $Z \in [X,u]$. Using $u_d(k)$ on both sides of (17) and considering the system dynamics (15), we have:

$$\begin{aligned} &u_d(k) - uf_i(k) = \delta uf_i(k) \\ &= \delta uf_{i-1}(k) - L(k)e_{i-1}(k+1) - \lambda(k)ub_{i-1}(k) \\ &= \delta uf_{i-1}(k) - L(k)\delta y_{i-1}(k+1) - \lambda(k)ub_{i-1}(k) \\ &= \delta uf_{i-1}(k) - L(k)C\delta X_{i-1}(k+1) + \\ &L(k)v_{i-1}(k+1) - \lambda(k)ub_{i-1}(k) = \delta uf_{i-1}(k) - \\ &L(k)C[f(X_d,k) + B(X_d,k)\mathrm{u_d}(k) - f(X_{i-1},k) - \\ &B(X_{i-1},k)\mathrm{u_{i-1}}(k) - \omega_{i-1}(k)] + L(k)v_{i-1}(k+1) - \\ &\lambda(k)ub_{i-1}(k) \end{aligned} \tag{21}$$

Adding and subtracting $B(X_{i-1},k)u_d(k)$, we can write:

$$\begin{aligned} &\delta uf_i(k) = \delta uf_{i-1}(k) - L(k)C[f(X_d,k) - \\ &f(X_{i-1},k) + [B(X_d,k) - B(X_{i-1},k)]u_d(k) + \\ &B(X_{i-1},k)\delta uf_{i-1}(k) + B(X_{i-1},k)ub_{i-1}(k) - \\ &\omega_{i-1}(k)] + L(k)v_{i-1}(k+1) - \lambda(k)ub_{i-1}(k) \end{aligned} \tag{22}$$

Taking the norm from both sides of (22) yields:

$$\begin{aligned} &\|\delta uf_i(k)\| \le \rho\|\delta uf_{i-1}(k)\| + a_1\|\delta X_{i-1}(k)\| \\ &+a_2\|ub_{i-1}(k)\| + a_3 \end{aligned} \tag{23}$$

Whereby considering the defined boundaries and $\|u_d(k)\| \le b_{u_d}$ we will have:

$$\begin{aligned} &\rho = \|I - L(k)CB(X,k)\|, \\ &a_1 = b_L b_c c_f + b_L b_c c_B b_{u_d} \\ &a_2 = b_L b_c b_B + b_\lambda,\ a_3 = b_L b_c b_\omega + b_L b_v \end{aligned} \tag{24}$$

Now, based on (18) for feedback controller $ub_{i-1}(k)$ we can write:

$$\begin{aligned} &\|ub_{i-1}(k)\| \le \|rZ_{i-1}(k)\| + \\ &\|s(Z_{i-1}(k))\|\|e_{i-1}(k)\| \le b_r\|Z_{i-1}(k)\| + \\ &b_s\|C\delta X_{i-1}(k) - v_{i-1}(k)\| \le b_r\|Z_{i-1}(k)\| + \\ &b_s b_c\|\delta X_{i-1}(k)\| + b_s b_v \end{aligned} \tag{25}$$

Substitution (25) in (23), we can obtain:

$$\begin{aligned} &\|\delta uf_i(k)\| \le \rho\|\delta uf_{i-1}(k)\| + (a_1 + \\ &a_2 b_s b_c)\|\delta X_{i-1}(k)\| + a_2 b_r\|Z_{i-1}(k)\| + a_2 b_s b_v + \\ &a_3 \le \rho\|\delta uf_{i-1}(k)\| + h(\|\delta X_{i-1}(k)\| + \\ &\|Z_{i-1}(k)\|) + a_4 \end{aligned} \tag{26}$$

When considering norm operator properties, we have:

$$\begin{aligned} &h = \max\,[a_1 + a_2 b_s b_c,\ a_2 b_r] \\ &a_4 = a_2 b_s b_v + a_3 \end{aligned} \tag{27}$$

Now, we should obtain properties of $\|\delta X_{i-1}(k)\| + \|Z_{i-1}(k)\|$. Thus, according to (18) we can write:

$$\begin{aligned} &\|Z_{i-1}(k)\| \le \|pZ_{i-1}(k-1)\| + \\ &\|q(Z_{i-1}(k-1))\|\|e_{i-1}(k-1)\| \le \\ &b_p\|Z_{i-1}(k-1)\| + b_q\|C\delta X_{i-1}(k-1) - \\ &v_{i-1}(k-1)\| \le b_q\|Z_{i-1}(k-1)\| + \\ &b_q b_c\|\delta X_{i-1}(k)\| + b_q b_v \end{aligned} \tag{28}$$

Considering (15):

$$\begin{aligned} &\delta X_{i-1}(k) = \\ &f(X_d,k-1) - f(X_{i-1},k-1) + [B(X_d,k-1) - \\ &B(X_{i-1},k-1)]u_d(k-1) + B(X_{i-1},k - \\ &1)\delta uf_{i-1}(k-1) + B(X_{i-1},k-1)ub_{i-1}(k-1) - \\ &\omega_{i-1}(k-1) \end{aligned} \tag{29}$$

So:

$$\begin{aligned} &\|\delta X_{i-1}(k)\| \le \|f(X_d,k-1) - f(X_{i-1},k - \\ &1)\| + \|[B(X_d,k-1) - B(X_{i-1},k-1)]u_d(k - \\ &1)\| + \|B(X_{i-1},k-1)\delta uf_{i-1}(k-1)\| + \\ &\|B(X_{i-1},k-1)ub_{i-1}(k-1)\| + \|\omega_{i-1}(k - \\ &1)\| \end{aligned} \tag{30}$$

And considering assumptions 1-5, we will have:

$$\begin{aligned} &\|\delta X_{i-1}(k)\| \le c_f\|\delta X_{i-1}(k-1)\| + \\ &c_B b_{u_d}\|\delta X_{i-1}(k-1)\| + b_B\|\delta uf_{i-1}(k-1)\| + \\ &b_B\|ub_{i-1}(k-1)\| + b_\omega \end{aligned} \tag{31}$$

With substitution (25) into (31) we get:

$$\begin{aligned} &\|\delta X_{i-1}(k)\| \le \\ &(c_f + b_B b_s b_c + c_B b_{u_d})\|\delta X_{i-1}(k-1)\| + \\ &b_B\|\delta uf_{i-1}(k-1)\| + b_B b_r\|Z_{i-1}(k-1)\| + \\ &b_B b_s b_v + b_\omega \end{aligned} \tag{32}$$

Adding (28) to (32) yields to:

$$\begin{aligned} &\|\delta X_{i-1}(k)\| + \|Z_{i-1}(k)\| \le \\ &(c_f + b_B b_s b_c + c_B b_{u_d} + b_q b_c)\|\delta X_{i-1}(k-1)\| + \\ &(b_B b_r + b_q)\|Z_{i-1}(k-1)\| + b_B\|\delta uf_{i-1}(k - \\ &1)\| + a_5 \\ &\le m_1(\|\delta X_{i-1}(k-1)\| + \|Z_{i-1}(k-1)\|) \\ &\qquad + m_2\|\delta uf_{i-1}(k-1)\| + a_5 \end{aligned} \tag{33}$$

where we defined:

$$\begin{aligned} &m_1 = \\ &\max\,[(c_f + b_B b_s b_c + c_B b_{u_d} + b_q b_c),(b_B b_r + b_q)] \\ &m_2 = b_B \\ &a_5 = b_B b_s b_v + b_\omega + b_q b_v \end{aligned} \tag{34}$$

Lemma. Given inequality: $Z(k+1) \le \beta(k) + hZ(k)$, where $\beta(k), Z(k)$ are scalar functions in all $k \ge 0$ and h is a positive constant, for $k \ge 1$ we can write:

$$Z(k) \le \sum_{j=0}^{k-1} h^{k-1-j}\,\beta(k) + h^k Z(0) \tag{35}$$

The proof is established by induction [29].

Now utilizing the given lemma and according to feedback controller initial condition $Z_{i-1}(0) = 0$, equation (33) can be written in the form of:

$$\begin{aligned} &\|\delta X_{i-1}(k)\| + \|Z_{i-1}(k)\| \le \\ &\textstyle\sum_{j=0}^{k-1} {m_1}^{k-1-j}\left(m_2\|\delta uf_{i-1}(j)\| + a_5\right) + \\ &m_1^k(\|\delta X_{i-1}(0)\| + \|Z_{i-1}(0)\|) \le \end{aligned} \tag{36}$$

$\sum_{j=0}^{k-1} m_1{}^{k-1-j}\left(m_2\|\delta uf_{i-1}(j)\| + a_5\right) + m_1^k b_{x_0}$

Substitution (36) by (25) leads to:

$$\|\delta uf_i(k)\| \le \rho\|\delta uf_{i-1}(k)\| + h\left(\sum_{j=0}^{k-1} m_1{}^{k-1-j}\left(m_2\|\delta uf_{i-1}(j)\| + a_5\right)\right) + h m_1^k b_{x_0} + a_4 \quad (37)$$

Now, by multiplying $\left(\frac{1}{\alpha}\right)^k$ to both sides of (37) and choosing $\alpha > \max[1, m_1]$, we will have:

$$\|\delta uf_i(k)\|\left(\frac{1}{\alpha}\right)^k \le \rho\|\delta uf_{i-1}(k)\|\left(\frac{1}{\alpha}\right)^k + \frac{h}{\alpha}\left(\sum_{j=0}^{k-1}\left(\frac{m_1}{\alpha}\right)^{k-1-j}\left(m_2\|\delta uf_{i-1}(j)\|\left(\frac{1}{\alpha}\right)^j + a_5\left(\frac{1}{\alpha}\right)^j\right)\right) + h\left(\frac{m_1}{\alpha}\right)^k b_{x_0} + a_4\left(\frac{1}{\alpha}\right)^k \quad (38)$$

Considering the α-norm definition and noting that α-norm of a constant value is the constant itself:

$$\|\delta uf_i(k)\|_\alpha \le \rho\|\delta uf_{i-1}(k)\|_\alpha + \frac{h\left(m_2\|\delta uf_{i-1}(k)\|_\alpha + a_5\right)}{\alpha}\left(\sum_{j=0}^{k-1}\left(\frac{m_1}{\alpha}\right)^{k-1-j}\right) + h\, b_{x_0} + a_4 \quad (39)$$

Knowing $\sum_{j=0}^{n-1}(a)^j = \frac{1-a^n}{1-a}$, we can obtain:

$$\|\delta uf_i(k)\|_\alpha \le \rho\|\delta uf_{i-1}(k)\|_\alpha + \frac{h\left(m_2\|\delta uf_{i-1}(k)\|_\alpha + a_5\right)}{\alpha}\left(\frac{1-\left(\frac{m_1}{\alpha}\right)^k}{1-\left(\frac{m_1}{\alpha}\right)}\right) + h\, b_{x_0} + a_4 \le \left(\rho + \frac{h m_2\left(1-\left(\frac{m_1}{\alpha}\right)^k\right)}{\alpha - m_1}\right)\|\delta uf_{i-1}(k)\|_\alpha + \frac{h a_5\left(1-\left(\frac{m_1}{\alpha}\right)^k\right)}{\alpha - m_1} + h\, b_{x_0} + a_4 \quad (40)$$

Or in a simpler notation:

$$\|\delta uf_i(k)\|_\alpha \le \hat{\rho}\|\delta uf_{i-1}(k)\|_\alpha + \epsilon \quad (41)$$

where:

$$\epsilon = \frac{h a_5\left(1-\left(\frac{m_1}{\alpha}\right)^k\right)}{\alpha - m_1} + h\, b_{x_0} + a_4 \quad (42)$$

And:

$$\hat{\rho} = \rho + \frac{h m_2\left(1-\left(\frac{m_1}{\alpha}\right)^k\right)}{\alpha - m_1} < 1 \quad (43)$$

Thus, choosing large enough α when trial number $i \to \infty$ leads to:

$$\|\delta uf_i(k)\|_\alpha \le \hat{\rho}\|\delta uf_{i-1}(k)\|_\alpha + \epsilon \xrightarrow{i\to\infty} \lim_{i\to\infty}\|\delta uf_i(k)\|_\alpha \le \frac{\epsilon}{1-\hat{\rho}} \quad (44)$$

Equation (44) shows the proposed CILC leads the feedforward control signal to a neighborhood of the desired control signal. Also, by considering (36), it can be written:

$$\lim_{i\to\infty}\|\delta X_{i-1}(k)\| \le \lim_{i\to\infty}\left(\sum_{j=0}^{k-1} m_1{}^{k-1-j}\left(m_2\|\delta uf_{i-1}(j)\| + a_5\right) + m_1^k b_{x_0}\right) \le a_6\frac{\epsilon}{1-\hat{\rho}} + a_7 \quad (45)$$

where a_6 and a_7 are determined constants. In the end, it can be concluded that:

$$\lim_{i\to\infty}\|y_d(k) - y_i(k)\| \le \lim_{i\to\infty}\left(b_c\|\delta X_i(k)\| + b_v\right) \le b_c\left(a_6\frac{\epsilon}{1-\hat{\rho}} + a_7\right) + b_v \le \sigma \quad (46)$$

Which proves the proposed theorem. While the convergence proof follows established ILC frameworks, the specific adaptation to a cable-driven rehabilitation traction system introduces unique constraints and challenges. Unlike standard industrial ILC applications, our system must account for variable patient biomechanics, safety considerations, and limited degrees of freedom. We highlight the following contributions:

- Tailoring CILC for cable-driven systems with constrained mobility and tension limits.
- Introducing a feedback term to enhance first-iteration performance, which improves applicability in real-world scenarios.
- Analyzing the impact of nonlinear Jacobian behavior due to the parallel cable configuration, ensuring robustness against singularities and underactuation.

IV. Applying the proposed CILC to the proposed structure

Now we will implement the proposed CILC algorithm on the proposed cable-driven robot dynamics equations (11) to (14) which serves as a rehabilitation traction device. Assume $\gamma(t) = x_1(t), \dot{\gamma}(t) = x_2(t)$ where $\gamma(t)$ is the patient's leg side bending. State space equations can be written as:

$$\dot{X}(t) = \begin{bmatrix} x_2(t) \\ -M(x)^{-1}\left(C(X,\dot{X}) + G(X)\right)\end{bmatrix} + \begin{bmatrix} 0 \\ M(x)^{-1}\end{bmatrix}\left(J^T\tau - T_d\right)$$
$$y(t) = [x_1(t), x_2(t)]^T \quad (47)$$

As said in the previous sections, the mass matrix $M(x)$ is an asymmetric and positive definite matrix so it is invertible, and writing the above equations is possible.

With the discrete-time interval dt=5ms, which leads to $N = [0,1,\ldots,6000]$ for operation time T=30s and using Euler's approximation method, (47) can be written in a discrete-time state-space format:

$$X(k+1) = \begin{bmatrix} x_1(k) + dt x_2(k) \\ x_2(k) + dt\left[-M(X(k))^{-1}\begin{pmatrix} C(X(k), X(k+1)) \\ +G(X(k))\end{pmatrix}\right]\end{bmatrix} + \begin{bmatrix} 0 \\ dt(M(x(k))^{-1})\end{bmatrix}\left(J^T\tau(k) - T_d(k)\right)$$
$$y(k) = \begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix}X(k) + v(k) \quad (48)$$

where $T_d(k)$ indicates disturbance, which can include model uncertainty and friction. Also, $v(k)$ includes measurement noise and error due to numerical differentiation.

Now to find learning gain $L(K) = [l_1, l_2]$, according to the proposed CILC convergence rule (19) we have:

$$\left|I_{2\times 2} - L(k)\begin{bmatrix} 1 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix} 0 \\ dt(M(x(k))^{-1})\end{bmatrix}\right| < 1 \rightarrow \left|1 - \frac{l_2 dt}{\underline{I}}\right| < 1 \quad (49)$$

where $\underline{I}$ is the lower bound for the patient's leg rotational inertia in mass matrix $M(x)$. Inertial properties of the human body can be found in [32]. In our simulations, the patient's leg mass is 14 Kg and its length is 0.9 m. So learning gain $L(k)$ is selected as $L(k)=(1-e^{-0.5kdt})[100,70]$.

Furthermore, the feedback controller (18) is chosen to take the following form:

$$\begin{aligned} Z_i(k+1) &= L_3 e_i(k) \\ ub_i(k) &= Z_i(k) + L_2 e_i(k) \end{aligned} \tag{50}$$

where $L_2 = [300,120]$ and $L_3 = [60,24]$. Additionally, to complete the proposed CILC structure we choose: $\lambda(k)=0.4\times(1-e^{-0.5kdt})$. These gains are easily obtained with the trial and error method.

Simulation studies are presented to verify the proposed theorem. We have run three simulations. The desired trajectory is identical for all simulations and contains a ± 15 degree pendulum movement. Furthermore, for a better comparison, initial condition error and disturbance are the same in all simulations while a bounded random output noise is added in each trial. It should be noted that in a simulation, each trial has a random bounded initial condition error. In the first simulation, learning gains are as said before. Results are shown in Figure 6 and 7. In the first trial, only the feedback controller rolls in satisfaction with the control task. By increasing the trials and CILC roll-taking tracking, the error decreases magnificently and MSE converges.

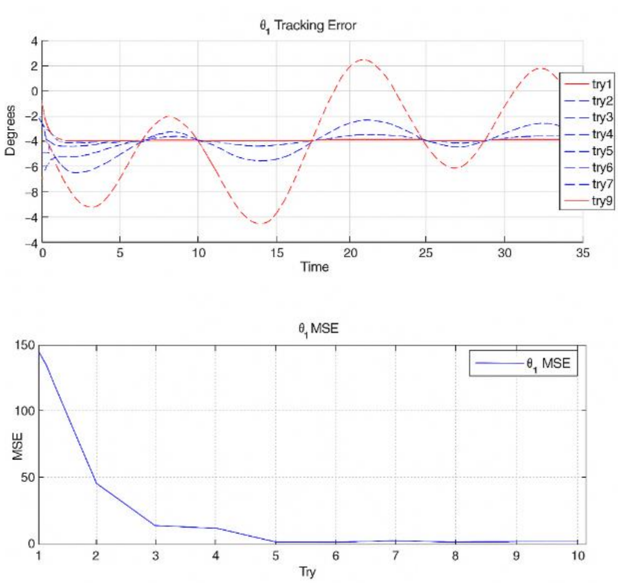


Figure 6 tracking error and MSE convergence of x_1 for 10 trials in first simulation

In the second simulation, we compare different CILC signals which can be made by changing the learning gains so we can have a better understanding of the proposed CILC structure. For this purpose, once we select $L_1(k) = 0$ and $L_3 = 0$ and the other gains are as before. Then $\lambda(k) = 0$ & $L_3 = 0$ and in the third only $L_3 = 0$ will be changed. For the comparison, each structure MSE is shown in Figure . As it

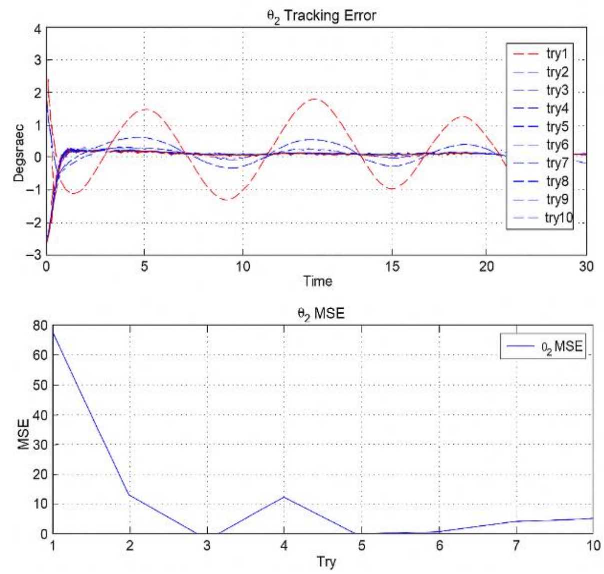


Figure 7 Tracking error and MSE convergence of x2 for 10 trials in the first simulation

can be seen, minimum MSE appears where we used the proposed CILC with full learning gains selection. Also, when the previous time-index error is used in the feedback controller structure, it leads to a better tracking performance including the first trial. In the third simulation, the effects of learning gain $L(k)$ are investigated in the proposed CILC. Therefore MSE result of the first simulation is compared with the ones when learning gain changes as $L(k)=(1-e^{-0.5kdt})[50,25]$ and $L(k)=(1-e^{-0.5kdt})[200,140]$. It is worth mentioning that the other gains and conditions are as same as the previous. Results are illustrated in **Error! Reference source not found.**

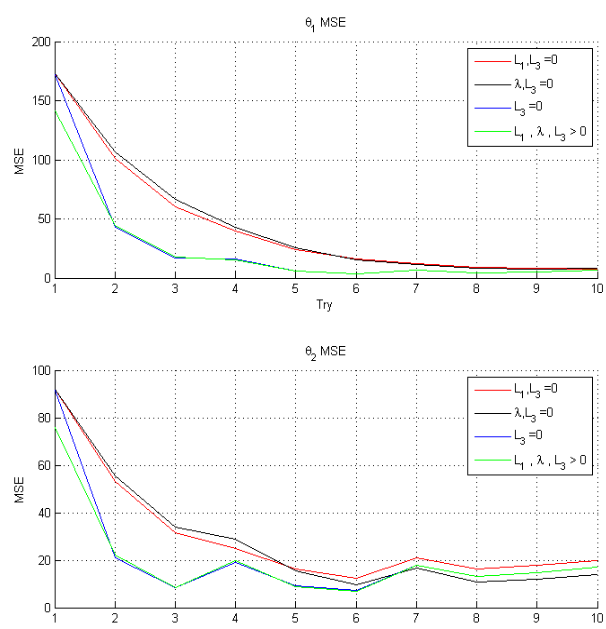


Figure 8 comparison of different CILC signal structures in the second simulation

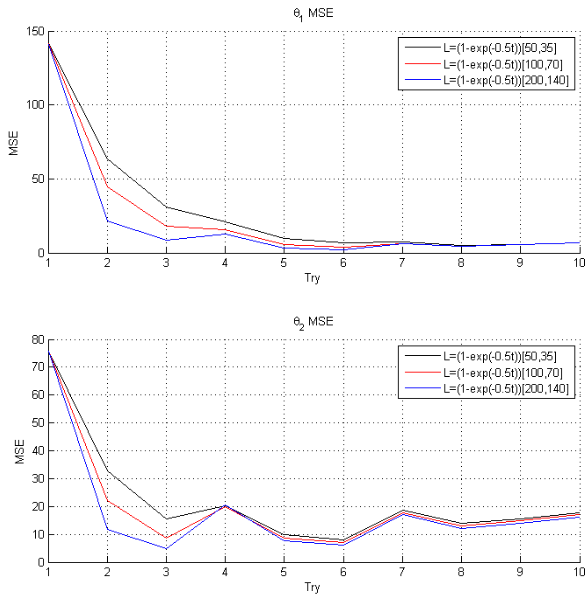


Figure 9 comparison of learning gain L(k) changes in the third simulation

Finally, to compare the results of our proposed controller with a basic controller such as PID, this comparison is made for the mean square error, which is shown in Figure 10. The results clearly show the efficiency of the proposed controller.

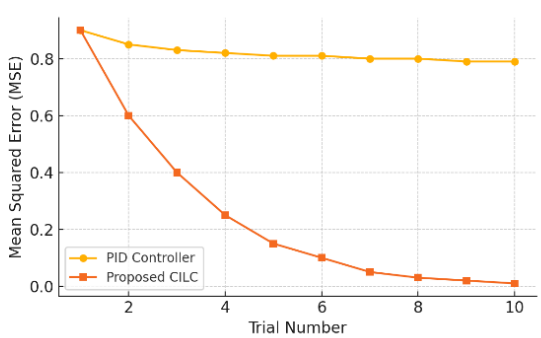


Figure 10 comparison of proposed controller and PID controller

V. Conclusion

In conclusion, this paper introduced a cable-driven rehabilitation traction device and proposed a new Current Iterative Learning Control (CILC) law to improve trajectory tracking performance. Although the theoretical analysis and simulation results show promising convergence and robustness, the absence of experimental validation remains a limitation. Future research will focus on building a functional prototype and conducting pilot experiments to assess the clinical potential of the system. Future work will include:

- Prototyping and hardware implementation of the proposed system.
- Experimental validation with healthy subjects and eventually patients.
- Integration of sensor feedback for real-time adaptive control.
- Comparative studies with conventional PID and other control strategies.
- Exploration of delay compensation and disturbance rejection techniques specific to biomechanical applications.

References

[1] L. Wang, M. Zhao, J. Ma, S. Tian, P. Xiang, W. Yao, et al., "Effect of combining traction and vibration on back muscles, heart rate and blood pressure," Medical engineering & physics, vol. 36, pp. 1443-1448, 2014.

[2] L. K. Wong, Z. Luo, and N. Kurusu, "Dynamic simulation of cervical traction therapy: Comparison between sitting and inclined positions," in Robotics and Biomimetics (ROBIO), 2014 IEEE International Conference on, 2014, pp. 167-172.

[3] J.-H. Shin, S.-l. Jun, Y.-J. Lee, J.-H. Kim, S.-Y. Hwang, and S.-H. Ahn, "Effects of intermittent traction therapy in an experimental spinal column model," Journal of acupuncture and meridian studies, vol. 7, pp. 83-91, 2014.

[4] W. M. Park, K. Kim, and Y. H. Kim, "Biomechanical analysis of two-step traction therapy in the lumbar spine," Manual therapy, vol. 19, pp. 527-533, 2014.

[5] C. C. Apfel, O. S. Cakmakkaya, W. Martin, C. Richmond, A. Macario, E. George, et al., "Research article Restoration of disk height through non-surgical spinal decompression is associated with decreased discogenic low back pain: a retrospective cohort study," 2010.

[6] United Healthcare Services Inc. (2017). Motorized Spinal Traction medical policy.

[7] E. W. Eberling Jr, "Portable gravity assisted lumbar traction device," ed: Google Patents, 1985.

[8] B. K. Cha, "Portable traction device for traction therapy and methods of use thereof," ed: Google Patents, 2008.

[9] J. P. Boren, "Horizontal Lumbar Stretching Machine and Method," ed: Google Patents, 2008.

[10] M. Fidan and M. Coban, "Design and application of a novel motorized traction device," in Electrical and Electronics Engineering (ELECO), 2015 9th International Conference on, 2015, pp. 1122-1125.

[11] Q. Ya, S. Linyong, et all "Structure Design and Control System Realization of a Novel Traction Device," in Measuring Technology and Mechatronics Automation (ICMTMA), 2015 Seventh International Conference on, 2015, pp. 1013-1017.

[12] C. Bensoussan, "Apparatus for stretching the vertebral column of a person," ed: Google Patents, 2007.

[13] R. H. Taylor, A. Menciassi, G. Fichtinger, and P. Dario, "Medical robotics and computer-integrated surgery," in Springer handbook of robotics, ed: Springer, 2008, pp. 1199-1222.

[14] G. Van Herp, P. Rowe, P. Salter, and J. Paul, "Three-dimensional lumbar spinal kinematics: a study of range of movement in 100 healthy subjects aged 20 to 60+ years," Rheumatology, vol. 39, pp. 1337-1340, 2000.

[15] J. J. Gorman, K. W. Jablokow, and D. J. Cannon, "The cable array robot: Theory and experiment," in Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on, 2001, pp. 2804-2810.

[16] H. D. Taghirad, Parallel robots: mechanics and control: CRC press, 2013.

[17] M. Khosravi and H. D. Taghirad, "Stability Analysis and Robust PID Control of Cable Driven Robots Considering

Elasticity in Cables," AUT Journal of Electrical Engineering, vol. 48, pp. 113-126, 2016.
[18] Q. Miao, H. S. Lo, S. Q. Xie, and H. S. Li, "Iterative learning control method for improving the effectiveness of upper limb rehabilitation," in Mechatronics and Machine Vision in Practice (M2VIP), 2016 23rd International Conference on, 2016, pp. 1-5.
[19] M. A. Khosravi and H. D. Taghirad, "Robust PID control of fully-constrained cable driven parallel robots," Mechatronics, vol. 24, pp. 87-97, 2014.
[20] J.-X. Xu and Y. Tan, Linear and nonlinear iterative learning control vol. 291: Springer, 2003.
[21] D. A. Bristow, M. Tharayil, and A. G. Alleyne, "A survey of iterative learning control," IEEE Control Systems, vol. 26, pp. 96-114, 2006.
[22] Z. Bien and J.-X. Xu, Iterative learning control: analysis, design, integration and applications: Springer Science & Business Media, 2012.
[23] Y. Chen and C. Wen, Iterative learning control: convergence, robustness and applications: Springer-Verlag, 1999.
[24] Z.-S. Hou and J.-X. Xu, "A new feedback-feedforward configuration for the iterative learning control of a class of discrete-time systems," Acta Automatica Sinica, vol. 33, pp. 323-326, 2007.
[25] T.-Y. Kuc, J. S. Lee, and K. Nam, "An iterative learning control theory for a class of nonlinear dynamic systems," Automatica, vol. 28, pp. 1215-1221, 1992.
[26] J.-X. Xu, et all "Analysis of continuous iterative learning control systems using current cycle feedback," in American Control Conference, Proceedings of the 1995, 1995, pp. 4221-4225.
[27] R. Chi, N. Lin, and R. Zhang, "A novel design of iterative learning control with feedback structure," in Control Conference (CCC), 2016 35th Chinese, 2016, pp. 3152-3156.
[28] C.-J. Chien and Y.-S. Shu, "Study of a class of sampled-data ILC from the point of performance improvement and memory Capacity," in Control and Decision Conference (CCDC), 2016 Chinese, 2016, pp. 4197-4202.
[29] D. Wang, "Convergence and robustness of discrete time nonlinear systems with iterative learning control," Automatica, vol. 34, pp. 1445-1448, 1998.
[30] C. Yu and X. Chen, "Trajectory tracking of wheeled mobile robot by adopting iterative learning control with predictive, current, and past learning items," Robotica, vol. 33, pp. 1393-1414, 2015.
[31] Y. Zhao, F. Zhou, Y. Li, and Y. Wang, "A novel iterative learning path-tracking control for nonholonomic mobile robots against initial shifts," International Journal of Advanced Robotic Systems, vol. 14, p. 1729881417710634, 2017.
[32] R. Chandler, C. E. Clauser, J. T. McConville, H. Reynolds, and J. W. Young, "Investigation of inertial properties of the human body," AIR FORCE AEROSPACE MEDICAL RESEARCH LAB WRIGHT-PATTERSON AFB OH1975.
[33] A. Zafar and J. L. Mead, "Model predictive control for wearable robotic systems: A review," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 28, no. 10, pp. 2178–2190, 2020.
[34] M. S. Andersen et al., "A detailed biomechanical model of the lumbar spine for personalized rehabilitation planning," Medical Engineering & Physics, vol. 86, pp. 45–56, 2020.
[35] H. Kim et al., "Real-time feedback control in robotic-assisted spinal manipulation: A feasibility study," Journal of Medical Robotics Research, vol. 5, no. 2, p. 2050009, 2020.

Mahdi Alinaghizadeh Ardestani is currently an Assistant Professor in the Electrical and Computer Engineering Department of Technical and Vocational University. His research interests include Robust control, Artificial Intelligence, robotics and Smart Transportation. He has managed several industrial projects.

Parham Haji Ali Mohamadi graduated from K.N.Toosi University with a Bachelor of Science degree in Electrical Engineering in 2014. In 2017, he obtained MSc. Degree in Control Engineering from SRBIA University. Since 2022, he is working as a PhD researcher in the Robotics and Multibody Mechanics research group at VUB. At the R&MM, his research focuses on energy management methods in robotic field.

International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Minimum-Time Control of Constrained Systems via Phase-Plane Technique: Application in Robotic Manipulators

Valiollah Ghaffari

Department of Electrical Engineering, Persian Gulf University, Bushehr, Iran
Corresponding author's email: vghaffari@pgu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 10-September-2025
Received in revised form: 10-January-2026
Accepted: 24-January-2026
Published online: 23-Sep-2026

Keywords:
Control constraint,
Minimum-time control,
Phase-plane analysis,
Regulation,
Robotic manipulator.

ABSTRACT

The pursuit of time-optimal performance is a fundamental objective for high-speed robotic and mechatronic systems, where minimizing settling-time directly enhances throughput and operational efficiency. Despite its importance, deriving exact solutions for practical systems remains a formidable challenge. Existing methodologies, which predominantly rely on numerical optimization or simplified plant models, often yield suboptimal results; they fail to provide guarantees of global optimality and frequently prove inadequate when confronting realistic system constraints such as actuator saturation. To address this critical gap, this paper introduces an analytical framework for synthesizing time-optimal control laws for a class of second-order systems. The proposed method is based on phase portrait analysis, a powerful geometric approach that facilitates the direct derivation of the exact switching curve. This curve is the critical element that defines the globally optimal, bang-bang controller. The final control law is presented in a closed-form expression, thereby enabling computationally efficient and straightforward implementation. Simulation results validate the theoretical framework, demonstrating that the proposed controller consistently achieves the theoretical performance by tracking the optimal trajectory perfectly.

I. Introduction

The pursuit of optimal performance is a fundamental objective in control engineering, particularly in applications where speed and precision are critical. Among various optimality criteria, the minimum-time control problem holds significant practical importance. In these problems, it is of interest to move the system's states in the shortest possible time duration from an initial value to a desired point. This problem is especially relevant in various engineering fields such as robotics, aerospace, and manufacturing, where minimizing cycle time directly enhances productivity and efficiency [1-3].

However, a fundamental challenge in realizing time-optimal control arises from the inherent physical limitations of actuators. All real-world systems, such as the motors in a robotic manipulator, are subject to constraints on their maximum achievable force, torque, or voltage. These saturation constraints preclude the use of controllers that demand infinite control effort and transform the control design problem into a complex, constrained optimization task. For second-order systems, which aptly model a vast array of mechanical systems, including simple robotic arms, mass-spring-damper systems, and galvanometers, the solution is known to be a bang-bang controller. This controller structure operates by applying either the maximum positive or negative control effort, switching between them at precisely calculated instants [4-6].

The time-optimal control problem has been a central topic in control theory since its modern inception, largely propelled by the pioneering work of Pontryagin and his colleagues on the maximum principle [7]. This principle provides a necessary condition for optimality, establishing that for a system with bounded controls, the optimal control must often maximize a certain Hamiltonian function, leading to a bang-bang control structure [8]. For the specific case of double-integrator (Newtonian system) and second-order linear systems, the analytical solution has been long-established in classical texts, where the optimal switching curve in the phase plane is derived [1, 9].

While the concept of bang-bang control is well-established [4], the analytical derivation of the exact switching law remains a topic of interest. The phase-plane analysis methodology offers a powerful and intuitive geometric framework for tackling this problem [10]. By examining the trajectories of system's states (position and velocity), the optimal switching curve that defines the control law can be visualized and derived exactly. This approach provides not only a control solution but also deep insight into the system's behavior under optimal control [11].

Recently, some works have discussed the minimum-time control solution in linear time-invariant (LTI) systems with constrained input. Using the state transition matrix, the minimum-time control solution is exactly computed in the LTI models with constrained control effort [12]. Then, an analytical framework is presented to find a discrete-time minimum-time controller in the LTI models with constrained control input [13].

The application of these time-optimal principles to robotic systems, particularly for point-to-point motion, has been extensively explored. Considering torque constraints, to compute time-optimal trajectories, some foundational methods were presented for robotic manipulators along pre-defined paths [14, 15]. This work, often referred to as the bang-bang switching principle, has been implemented in various motion planning algorithms.

However, a significant portion of the literature focuses on numerical approaches to solve the resulting two-point boundary value problem for higher-order or nonlinear control systems [16, 17]. These control methods, while powerful, often rely on iterative optimization techniques that can be computationally intensive and may not provide the same level of intuitive insight as a closed-form analytical solution.

Furthermore, many modern control strategies for constrained systems, such as model predictive control (MPC), handle constraints effectively but provide sub-optimal solutions for the minimum-time problem due to their discrete-time nature and finite prediction horizons. The need for simple, exact, and intuitively precise control laws for fundamental system classes remains relevant for educational purposes, direct implementation on lightweight computational platforms, and as a performance benchmark for more complex algorithms.

A significant trend in contemporary literature involves solving optimally the path tracking in complex, higher-degree of freedom robotic manipulators. Highlighting the computational challenges, a control method is presented to approximate time-optimal control for the control systems with switching dynamics [18]. Similarly, a computationally efficient algorithm is suggested for time-optimal control of robotic systems with torque limitations, focusing on real-time implementability [19]. The work [16] reformulated the issue into a convex optimization, enabling efficient numerical solutions for robots following predefined paths.

Another active branch of research integrates time-optimal principles with predictive control. To achieve near-optimal performance, the use of MPC is explored while explicitly handling state and input constraints, though acknowledging the sub-optimality introduced by discretization [20]. To address this, recent work on real-time MPC [21] seeks to bridge the gap between computational tractability and time-optimal performance.

The problem is also being revisited for systems with specific nonlinearities and uncertainties. The time-optimal control for nonlinear systems using a reinforcement learning framework is investigated in the work [22], learning near-optimal policies where analytical solutions are intractable. For uncertain systems, a robust time-optimal control strategy is proposed that guarantees performance [23, 24]. Furthermore, the focus on performance has expanded beyond pure time minimization. For instance, a multi-objective optimization problem is addressed in the studies [25, 26], balancing time-optimality with energy consumption. Most recently, summarizing the current state of this nuanced field, a review of energy-time trade-offs in optimal control of the robotic manipulators was provided [2, 5, 27, 28].

This work distinguishes itself by returning to the analytical roots of the problem. Rather than presenting a new numerical algorithm, it provides a complete and self-contained derivation of the exact minimum-time control law for second-order LTI systems using phase-plane analysis. This approach offers a geometric and highly intuitive interpretation of the optimal switching logic, from first principles to final control law. The primary contribution lies in its clarity and comprehensiveness, serving to bridge classical theory with practical application in robotic regulation, and providing a clear benchmark for optimal performance.

Inspired by the minimum-time control of the Newtonian system [9], the key contribution of this research is the exact derivation of a minimum-time control law for a class of second-order LTI systems with bounded control efforts, utilizing a phase-plane idea. The issue is formulated in the context of a constrained robotic manipulator, initially modeled by its transfer function. The paper first addresses the stabilization problem, detailing how to move the system's state to the origin under predefined actuator constraints. The exact solution to the stabilization issue is then conclusively employed to solve the output regulation in the minimum possible time.

The remainder of this report is arranged as follows. Section 2 introduces the model mathematics and formally states the minimum-time control in the robot arm. Section 3 presents the core phase-plane analysis and derives the exact structure of the time-optimal control law. Then, its application is discussed in the output regulation case. Section 4 provides a numerical example to check the applicability and the effectiveness of the designed control methodology in the robotic arm. Finally, some concluding points are summarized in Section 5.

II. Problem Formulation

This section establishes the mathematical foundation for the minimum-time control problem. In numerous robotic and mechanical applications, ignoring the spring effects, a Cartesian robot manipulator with n-links can be imagined as follows [29]:

$$M\ddot{y}+B\dot{y}=u, \tag{1}$$

where $y\in\mathbb{R}^n$ and $u\in\mathbb{R}^n$ are some vectors containing the robot positions and the input forces, respectively. In Eq. (1), M denotes the mass matrix and B denotes the damping matrix. To specify exactly the minimum-time control solution, the above-mentioned matrices are assumed to be known. In a simplified form, in Fig. 1, a robotic manipulator with three perpendicular links is depicted.

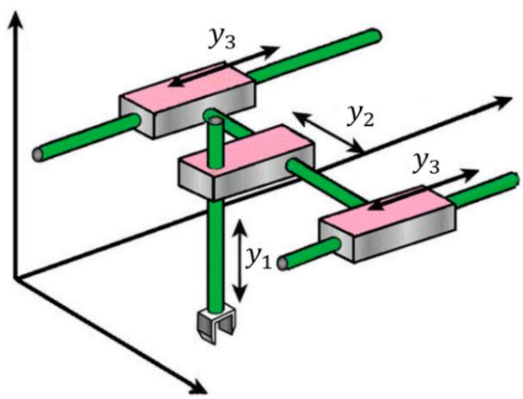


Fig. 1. A schematic of a 3-link manipulator

Here, the specific control objective is to synthesize a law $u(t)$ subjected to $|u(t)|\le u_{\max}$ that drives the system output $y(t)$, as well as its derivative (namely, the translational velocity $\dot{y}(t)$), from an primary state to a desired constant set-point in the minimum possible time, while strictly adhering to the control constraint.

In the condition which the terms M and B are some diagonal matrices, for instance like the model presented in [29], it can be decomposed as n independent subsystems. Each one may be a second-order system like

$$\ddot{y}-a\dot{y}=bu. \tag{2}$$

The control law is synthesized under the assumption that dynamic coupling between joints is negligible. This holds for operational scenarios such as slow movements or sequential joint control, where coupling effects are minimal. However, this assumption breaks down during high-speed, coordinated motions, where significant interaction forces arise and can lead to substantial tracking errors and performance loss.

As a consequence, the corresponding transfer function would be obtained as

$$G(s)=\frac{b}{s(s-a)} \tag{3}$$

where $y(t)$ is the system's output while the control effort $u(t)$ is constrained by $u_{\max}$. Moreover, the constants a and b are some known parameters for the designer. So, the control system (3) involves an integral term, while the coefficient a may be negative (stable) or positive (unstable). In the state space, the dynamic model (3) would be represented as

$$\begin{cases}\dot{x}_1=x_2\\ \dot{x}_2=ax_2+bu\end{cases} \tag{4}$$

where $x_1=y$ and $x_2=\dot{y}$.

Consequently, the specific control aim is to generate a constrained controller $u(t)=\phi(x_1,x_2)$ with $|u(t)|\le u_{\max}$, that minimizes the time required to drive the system from any initial values $(x_1(t_0),x_2(t_0))$ to the desired terminal state, satisfying the output regulation $\lim_{t\to t_f}x_1(t)=y_{\mathrm{ref}}$ and $\lim_{t\to t_f}x_2(t)=0$. This formally defines the time-optimal control problem that will be solved in the subsequent section through phase-plane analysis. Subsequently, the closed-loop stabilization in the minimum-time is presented for the system (4). Then the results are used in the regulation problem.

III. Main Results

Having formally stated the problem, this part discusses the geometric derivation of the exact time-optimal control law using the phase-plane methodology. The phase portrait offers a powerful tool for visualizing the system's dynamics and synthesizing controllers for nonlinear systems, such as those with bang-bang control structures. The analysis begins by transforming the second-order dynamics into a system of first-order equations in the state space, defining the phase variables $x_1=y$ (the system's output) and $x_2=\dot{y}$ (the derivative of the output). The core of this section will be dedicated to constructing the optimal switching curve, a critical locus in the phase plane that partitions the state space into zones where the optimal controller is either $u=u_{\max}$ or $u=-u_{\max}$. The properties of system trajectories under these extreme control efforts are analyzed to prove that this switching curve indeed yields the minimum-time solution to the origin, which is then extended to the regulation problem.

Stabilization Problem. It is desired to design a command signal $u(t)$ with $|u(t)|\le u_{\max}$ so that the system's states are moved in the minimum-time from $x(t_0)=x_0$ to $x(t_f)=0$. Thus, the cost function is

$$J=\int_{t_0}^{t_f}1\mathrm{d}t. \tag{5}$$

Let us construct the Hamiltonian function in the following form:

$$\mathcal{H}(x,\lambda,u)=1+\lambda_1x_2+\lambda_2(ax_2+bu). \tag{6}$$

The costate equations are found as follows:

$$\begin{cases}\dot{\lambda}_1=-\frac{\partial\mathcal{H}}{\partial x_1}=0\\ \dot{\lambda}_2=-\frac{\partial\mathcal{H}}{\partial x_2}=-\lambda_1-a\lambda_2\end{cases} \tag{7}$$

Employing the recognized Pontryagin principle with $|u(t)|\le u_{\max}$, we have:

$$\mathcal{H}\big(x^*(t),\lambda^*(t),u^*(t)\big)\le\mathcal{H}\big(x^*(t),\lambda^*(t),u(t)\big). \tag{8}$$

Consequently, the optimal control would have a solution if $|u^*(t)| = u_{\max}$ holds. Then the switching curve would be obtained as follows:

Case 1) If $u = +u_{\max}$ holds, then the system's solution is

$$\begin{cases} x_1(t) = x_1(t_0) + \frac{x_2(t_0)}{a}(e^{at} - 1) \\ \qquad +u_{\max}\frac{b}{a^2}(e^{at} - 1) - u_{\max}\frac{b}{a}t \quad , \ t \geq t_0 \\ x_2(t) = x_2(t_0)e^{at} + u_{\max}\frac{b}{a}(e^{at} - 1) \end{cases} \tag{9}$$

The trajectory crossing the origin is

$$\begin{cases} x_1 = u_{\max}\frac{b}{a^2}(e^{at} - 1) - u_{\max}\frac{b}{a}t \\ x_2 = u_{\max}\frac{b}{a}(e^{at} - 1) \end{cases}, \forall t \in \mathbb{R} \tag{10}$$

In the phase-plane, the corresponding nonlinear curve is established as follows:

$$x_1 = \frac{1}{a}x_2 - u_{\max}\frac{b}{a^2}\mathrm{Ln}\left(1 + \frac{a}{bu_{\max}}x_2\right). \tag{11}$$

Case 2) If $u = -u_{\max}$ holds, then the system's solutions are

$$\begin{cases} x_1 = x_1(t_0) + \frac{x_2(t_0)}{a}(e^{at} - 1) \\ \qquad +u_{\max}\frac{b}{a^2}(1 - e^{at}) + u_{\max}\frac{b}{a}t \quad , t \geq t_0 \\ x_2 = x_2(t_0)e^{at} + u_{\max}\frac{b}{a}(-e^{at} + 1) \end{cases} \tag{12}$$

Similarly, the trajectory crossing the origin is

$$\begin{cases} x_1 = u_{\max}\frac{b}{a^2}(1 - e^{at}) + u_{\max}\frac{b}{a}t \\ x_2 = u_{\max}\frac{b}{a}(-e^{at} + 1) \end{cases}, \forall t \in \mathbb{R} \tag{13}$$

Accordingly, the following switching curve would be obtained:

$$x_1 = \frac{1}{a}x_2 + u_{\max}\frac{b}{a^2}\mathrm{Ln}\left(1 - \frac{a}{bu_{\max}}x_2\right). \tag{14}$$

For different selections of the value a and b, the phase-plane of the control system with $u_{\max} = 1$ are plotted in Fig. 2. Then the corresponding stable switching curves are shown in Fig. 3.

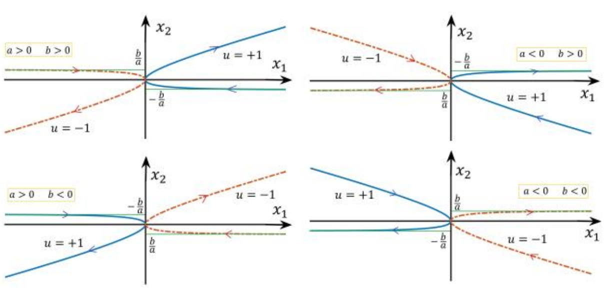


Fig. 2. The phase-planes for different signs of a and b

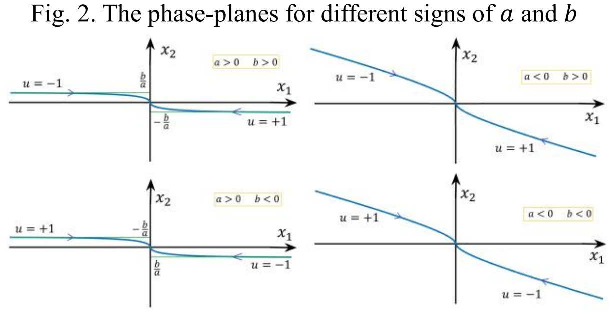


Fig. 3. Stable switching curves for different signs of a and b

According to Fig. 3, depending on a and b, a stable switching curve could be found by the following formula:

$$x_1 = \frac{1}{a}x_2 + u_{\max}\frac{|b|}{a^2}\mathrm{sgn}(x_2)\mathrm{Ln}\left(1 - \frac{a}{u_{\max}}\left|\frac{x_2}{b}\right|\right). \tag{15}$$

Then, aided by the phase plane diagram, the control signal $u(t)$ is optimally computed as follows:

$$u(t) = \begin{cases} -u_{\max}\mathrm{sgn}\big(bs(t)\big) & s(t) \neq 0 \\ -u_{\max}\mathrm{sgn}\big(bx_2(t)\big) & s(t) = 0 \end{cases} \tag{16}$$

where the switching curve $s(x_1, x_2)$ is

$$s = x_1 - \frac{1}{a}x_2 - u_{\max}\frac{|b|}{a^2}\mathrm{sgn}(x_2)\mathrm{Ln}\left(1 - \frac{a}{u_{\max}}\left|\frac{x_2}{b}\right|\right).$$

It states that the case $s(t) \neq 0$ drives the system towards the switching manifold, while the case $s(t) = 0$ with $x_2(t) \neq 0$ maintains the system on the switching surface. The control aims $(x_1(t) = x_2(t) = 0)$ would be favorably accomplished when $s(t) = 0$ and $x_2(t) = 0$ hold.

In the phase plane, it can be checked that the set $s(x_1, x_2) = 0$ or Eq. (15) constructs an invariant set. Thus the system's solutions remain on Eq. (15) for the initial conditions that start from $s(x_1, x_2) = 0$.

Unstable Region. The argument of the logarithmic function would always be positive when $a < 0$. However, for positive a, the minimum-time control problem would not have a solution when $|x_2(t_0)| > u_{\max}\frac{|b|}{a}$. For example, as far as the coefficient a is positive, either $u(t) = +u_{\max}$ combined with $x_2(t_0) = -u_{\max}\frac{b}{a}$ or $u(t) = -u_{\max}$ along with $x_2(t_0) = u_{\max}\frac{b}{a}$ is taken, applying Eq. (9) and Eq. (12), the system's solution is

$$\begin{cases} x_1(t) = x_1(t_0) + x_2(t_0)t \\ x_2(t) = x_2(t_0) \end{cases}, \quad t \geq t_0 \tag{17}$$

As a consequence, the solutions would be unbounded if both control effort $u(t)$ and initial condition $x_2(t_0)$ remain on a specified boundary set. Utilizing Eq. (9) and Eq. (12), it can be observed that the system's solutions would tend to infinity if the conditions $|x_2(t_0)| > u_{\max}\frac{|b|}{a}$ and $a > 0$ hold. In other words, the control authority is insufficient to counteract the divergent dynamics, making the time-optimal problem infeasible from such initial states.

Next, the results are discussed for the small a. To this aim, the logarithmic function $\mathrm{Ln}(1 + z)$ can be approximated as the following:

$$\mathrm{Ln}(1 + z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \frac{1}{4}z^4 + \cdots, |z| < 1. \tag{18}$$

Fig. 3 shows, for small a (i.e., $|a| \ll 1$) and $b > 0$, if $u = +u_{\max}$, then $x_2 < 0$ and the following condition would hold:

$$x_1 = \frac{1}{a}x_2 - \frac{1}{a}x_2 + \frac{1}{2}\frac{x_2^2}{bu_{\max}} - \frac{1}{3}\frac{ax_2^3}{b^2u_{\max}^2} + \cdots = \frac{1}{2bu_{\max}}x_2^2. \tag{19}$$

Similarly, if $u = -u_{\max}$, then $x_2 > 0$ and the subsequent approximation could be written:

$$x_1 = \frac{1}{a}x_2 - \frac{1}{a}x_2 - \frac{1}{2}\frac{x_2^2}{bu_{\max}} - \frac{1}{3}\frac{ax_2^3}{b^2u_{\max}^2} + \cdots = -\frac{1}{2bu_{\max}}x_2^2. \tag{20}$$

Thus the final switching surface is reduced to the following second-order function:

$$x_1 = \frac{-1}{2bu_{\max}}|x_2|x_2, \ |a| \ll 1, b > 0 \ . \tag{21}$$

Accordingly, the same results were obtained in the study [9] for $a = 0$. The exact and approximated switching curves clearly illustrate the error introduced by the series truncation. This analysis provides the necessary context for practitioners to assess the trade-off between simplicity and optimality when deploying the simplified controller. In conclusion, the phase-plane analysis has yielded a complete solution to the minimum-time stabilization problem. The optimal switching curve has been rigorously derived, partitioning the phase plane into two distinct regions. The corresponding control law, defined by the bang-bang policy $u = +u_{\max}$ for states below the switching curve and $u = -u_{\max}$ for states above it, has been proven to move the initial states to the origin in the minimum achievable time while strictly adhering to the control constraint. Eq. (15) is a switching curve in the optimal control sense, guaranteeing that once the system trajectory reaches it, no further switching is needed. Thus, this law requires at most one switching event and provides the foundational result for addressing the output regulation in the following subsection.

Regulation Problem. The output regulation would be achieved at an invariant set-point y_{ref} with the equilibrium point $\bar{x} = [y_{\text{ref}} \quad 0]^T$ and $\bar{u} = 0$. So, satisfying the constraint $|u(t)| \le u_{\max}$, it is desired to determine $u(t)$ so that the states are moved from $x(t_0) = x_0$ to $x(t_f) = \bar{x}$ in the minimum-time. Defining $\xi(t) = x(t) - \bar{x}$, we have:

$$\begin{cases} \dot{\xi}_1 = \xi_2 \\ \dot{\xi}_2 = a\xi_2 + bu \end{cases} \tag{22}$$

Without loss of generality, it is equivalent to moving the state from $\xi_0 = x_0 - \bar{x}$ to the origin. Thus, substituting $x_1 = \xi_1 + y_{\text{ref}}$ and $x_2 = \xi_2$, a stable switching manifold is

$$\begin{aligned} x_1 &= y_{\text{ref}} + \frac{1}{a} x_2 \\ &\quad + u_{\max} \frac{|b|}{a^2} \operatorname{sgn}(x_2) \operatorname{Ln}\left(1 - \frac{a}{u_{\max}} \left|\frac{x_2}{b}\right|\right). \end{aligned} \tag{23}$$

Then the optimal control signal $u(t)$ would be found as

$$u(t) = \begin{cases} -u_{\max} \operatorname{sgn}\big(bs(t)\big) & s(t) \ne 0 \\ -u_{\max} \operatorname{sgn}\big(bx_2(t)\big) & s(t) = 0 \end{cases} \tag{24}$$

where

$$s = x_1 - y_{\text{ref}} - \frac{1}{a} x_2 - u_{\max} \frac{|b|}{a^2} \operatorname{sgn}(x_2) \operatorname{Ln}\left(1 - \frac{a}{u_{\max}} \left|\frac{x_2}{b}\right|\right).$$

Remark 1. The control signal contains instantaneous fluctuations around the equilibrium point. In the control law (24), to avoid chattering phenomena when $s(t) = 0$, the sign function can be approximated with a smoothed one like $\text{sat}(.)$, $\tanh(.)$, $\tan^{-1}(.)$ and the other ones. However, smoothing the sign function would fundamentally lead to sub-optimality.

Remark 2. The proposed bang-bang controller acts as a finite-time convergent control law, ensuring robustness to infinitesimal disturbances during the reaching phase. Accordingly, the system's states are driven to the origin in a pre-computable finite settling-time.

Remark 3. In the proposed control law, the computational burden associated with the real-time calculation of the logarithmic function may be a challenging issue. Although this imposes a greater load than a linear controller, it remains feasible for standard hardware, a fact demonstrated by its prevalence in other sophisticated control algorithms. To further mitigate this cost in resource-limited environments, the option of pre-computing the switching curve into a lookup table is presented, trading a marginal amount of memory for a deterministic and fast execution time.

Remark 4. The proposed method develops a framework to design an optimal control law for the second-order LTI systems with no uncertain part and a hard limitation on the control input. The solving technique is the phase plane which is helpful for the second-order models. Thus, the problem formulation is dedicated to the second-order systems. However, it may be generalized to the high-order LTI control systems. Moreover, deriving an exact (closed form) solution would have some extra complexities in the nonlinear second-order systems.

The phase-plane analysis has yielded a complete and exact solution to the minimum-time control for the constrained second-order model. The optimal control law, defined by the switching curve, has been rigorously derived and proven to possess a bang-bang structure, requiring at most one switching event to direct the initial states to the origin in minimum time. Furthermore, by defining the error state, this stabilization law is directly applicable to the output regulation problem, enabling time-optimal set-point tracking. The following section will demonstrate the implementation and efficacy of this control law through a numerical simulation.

IV. Numerical Simulation

To confirm the theoretical derivations and demonstrate the practical performance of the designed time-optimal control law, this section presents a comprehensive numerical example. A specific second-order system, representative of a typical robotic joint, is simulated under the derived bang-bang control policy. The response of the system's states in the phase portrait, the time evolution of the control effort, and the output regulation performance is analyzed in detail. The results serve to visually confirm the existence and properties of the optimal switching curve and, most importantly, to verify that the system achieves exact regulation in the minimum achievable time while strictly adhering to the control constraint. A robot manipulator with a single arm is discussed later. To this purpose, consider a robotic model described by Eq. (3) with $u_{\max} = 3$, $a = 1$ and $b = 2$. The initial conditions of the position and velocity are $x_1(0) = -3$ and $x_2(0) = 5$.

The integration time step is one millisecond, the solver is the Runge–Kutta method, and the duration of the simulation is 20 seconds. In this example, the reference is the following piece-wise constant signal:

$$y_{\mathrm{ref}}(t)=\begin{cases}2 & 0\le t<8\\ 3 & 8\le t<12\\ 0 & 12\le t<20\end{cases} \quad (25)$$

It is desired to track reference $y_{\mathrm{ref}}(t)$ as shown in Fig. 4. To avoid the chattering phenomena when $s(t)=0$, the function $\mathrm{sign}(bx_2)$ is approximated by $\tanh(100x_2)$. Applying the suggested procedure, the optimal control law (24) is realized with smoothed ones as follows:

$$u(t)=\begin{cases}-u_{\max}\mathrm{sgn}\big(s(t)\big) & s(t)\neq 0\\ -u_{\max}\tanh\big(100x_2(t)\big) & s(t)=0\end{cases} \quad (26)$$

where $s=x_1-y_{\mathrm{ref}}-x_2-6\mathrm{sgn}(x_2)\mathrm{Ln}\left(1-\frac{|x_2|}{6}\right)$.

To show the effectiveness of the proposed technique, the system behavior is evaluated with the LQR method. Such a state feedback optimal controller would be realized as

$$u(t)=-(x_1(t)-y_{\mathrm{ref}})-1.5x_2(t). \quad (27)$$

The system's output is shown in Fig. 4, while the second state is illustrated in Fig. 5. It is verified that zero steady-state error is achieved via both controllers. Moreover, corresponding to $y_{\mathrm{ref}}(t)=2$, the minimum time is 2.5545 seconds. The state trajectories of the robot in the phase-plane are seen in Fig. 6. The control effort $u(t)$ and the switching curve $s(t)$ are plotted in Fig. 7. Additionally, to show the effectiveness of the suggested control method, during interval $0<t<8$, some quantitative metrics like settling-time, peak output and peak control efforts are calculated and provided in Table 1.

Table 1. Comparison of some control indexes

Index	Proposed Method	LQR
Settling-Time [sec]	2.5545	5.6280
Peak Output [m]	2.7559	3.1055
Peak Control Effort [m/sec^2]	3	0.5019

Applying the proposed minimum-time controller, the output $y(t)$ follows the set-point $y_{\mathrm{ref}}(t)$ during the start and the brake phases. Compared with similar control techniques, the simulation results conclusively validate the theoretical framework developed in Section 3.

The state trajectory in the phase plane was observed to seamlessly follow the predicted optimal switching curve, switching control effort at the precise theoretical point to guide the system directly to the origin. The control input maintained its bang-bang structure, operating strictly within the defined saturation limits of $\pm u_{\max}$ throughout the process. Most significantly, the system achieved perfect regulation at a time that aligns with the theoretical minimum, as no other admissible control signal could have completed the transition faster. This numerical example effectively demonstrates the practical applicability and optimal performance of the phase-plane derived control law for the constrained second-order systems.

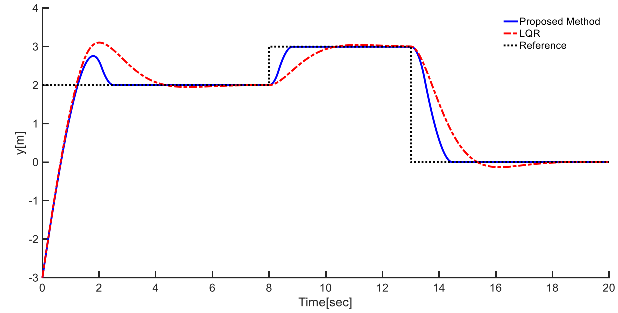


Fig. 4. The system's output and reference signal

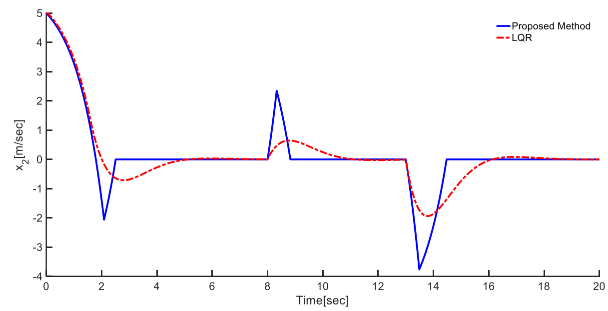


Fig. 5. The second state of the system

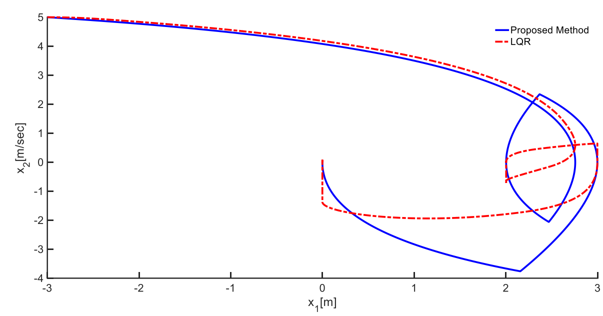


Fig. 6. The state trajectory in the phase-plane

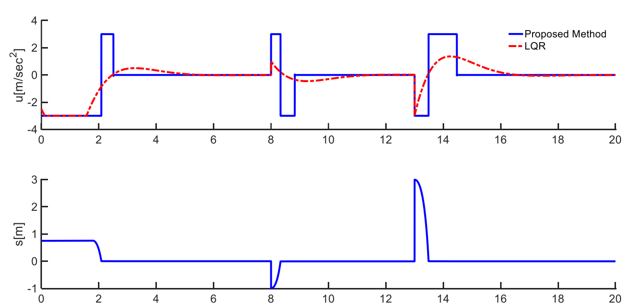


Fig. 7. The control effort $u(t)$ and the switching curve $s(t)$

To analyze the robustness against perturbations, new simulations explicitly test the controller's performance under both external disturbances and parametric model uncertainties. To this purpose, a torque disturbance with an amplitude of 2 is applied during $4<t<5$. Moreover, in the control law, $+20\%$ increment has been made in the parameters a and b. The output and input of the system are presented in Figs. 8 and 9. Additionally, to check the controller's robustness under external disturbance and model uncertainty, some quantitative control metrics during time interval $0 < t < 8$ are numerically calculated and reported in Table 2.

Table 2. Robustness of the proposed controller

Index	Nominal Case	Uncertain Case
Settling-Time [sec]	2.5545	2.8780
Peak Output [m]	2.7559	3.1111

Compared with the nominal case, the quantitative and figurative results demonstrate the controller's inherent robustness properties and quantify the resulting performance degradation such as the increase in settling-time.

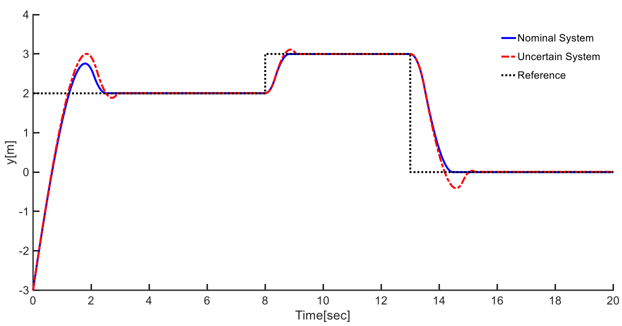


Fig. 8. The output and reference under uncertain terms

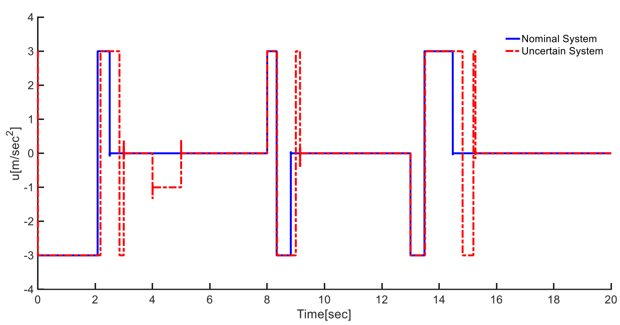


Fig. 9. The control effort $u(t)$ under uncertain terms

V. Conclusion

This paper has successfully derived and demonstrated an exact minimum-time control law for a class of second-order LTI systems subjected to control effort constraints. By employing a phase-plane analysis approach, the time-optimal control was translated into a geometric problem of finding the optimal switching curve that drives the states of the system to the origin in the shortest time. The core of the strategy was first established by solving the stabilization problem, proving that the optimal control policy is a bang-bang control law that switches between the maximum allowed control efforts. This policy was then rigorously extended to solve the output regulation, guaranteeing that the system reaches its desired set-point in the theoretical minimum time. The practical applicability of the theoretical findings was conclusively validated through a numerical example. The simulation results clearly illustrated the controller's performance, showcasing its ability to achieve precise regulation exactly as predicted by the phase-plane analysis, with no overshoot and in the fastest achievable time given the system's constraints. In summary, this work provides an exact and practically verifiable solution to the classic minimum-time control problem. The clarity of the phase-plane method offers significant intuitive insight into the structure of the optimal control law. Future work will focus on extending this approach to higher-order systems and exploring its robustness in the presence of model uncertainties and external disturbances.

Reference

[1] D. E. Kirk, Optimal control theory: an introduction. Courier Corporation, 2004.

[2] H. M. Y. Naeem, A. I. Bhatti, Y. A. Butt, Q. Ahmed, and X. Bai, "Energy Efficient Solution for Connected Electric Vehicle and Battery Health Management Using Eco-Driving Under Uncertain Environmental Conditions," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 4, pp. 4621-4631, 2024.

[3] X. Bai, H. Jiang, B. Zhang, and Z. Wu, "An Improved Multi-Population Genetic Algorithm for Multi-Robot Task Assignment With Complex Precedence Constraints," in International Conference on Autonomous Unmanned Systems Singapore, 2025, pp. 589-601.

[4] R. Bellman, I. Glicksberg, and O. Gross, "On the "bang-bang" control problem," Quarterly of Applied Mathematics, vol. 14, no. 1, pp. 11-18, 1956.

[5] X. Bai et al., "Efficient Hybrid Multi-Population Genetic Algorithm for Multi-UAV Task Assignment in Consumer Electronics Applications," IEEE Transactions on Consumer Electronics, vol. 71, no. 2, pp. 2395-2406, 2025.

[6] H. M. Y. Naeem, Q. Ahmed, Y. A. Butt, A. I. Bhatti, L. Qiu, and X. Bai, "Optimizing Energy and Battery Health Using Route, Allocation, and Velocity Planning for a Multielectric Vehicle System," IEEE Transactions on Transportation Electrification, vol. 11, no. 5, pp. 12618-12632, 2025.

[7] L. S. Pontryagin, Mathematical theory of optimal processes. London: Taylor & Francis, 2018.

[8] L. Sonneborn and F. Van Vleck, "The bang-bang principle for linear control systems," Journal of the Society for Industrial and Applied Mathematics: Control, vol. 2, no. 2, pp. 151-159, 1964.

[9] F. L. Lewis, D. Vrabie, and V. L. Syrmos, Optimal control. John Wiley & Sons, 2012.

[10] H. C. Lim, "Classical approach to bang-bang control of linear processes," Industrial & Engineering Chemistry Process Design and Development, vol. 8, no. 3, pp. 334-342, 1969.

[11] D.-S. Choi, S.-J. Kim, and I.-J. Ha, "A phase-plane approach to time-optimal control of single-DOF mechanical systems with friction," Automatica, vol. 39, no. 8, pp. 1407-1415, 2003.

[12] V. Ghaffari, "The exact solution of min-time optimal control problem in constrained LTI systems: a state

transition matrix approach," AUT Journal of Modeling and Simulation, vol. 51, no. 2, pp. 103-110, 2019.

[13] V. Ghaffari, "An analytical approach to min-time control of discrete-time linear systems with constrained input," International Journal of Industrial Electronics Control and Optimization, vol. 8, no. 1, pp. 37-44, 2025.

[14] K. Shin and N. McKay, "Minimum-time control of robotic manipulators with geometric path constraints," IEEE Trans on Automatic Control, vol. 30, no. 6, pp. 531-541, 1985.

[15] J. E. Bobrow, S. Dubowsky, and J. S. Gibson, "Time-optimal control of robotic manipulators along specified paths," The international journal of robotics research, vol. 4, no. 3, pp. 3-17, 1985.

[16] D. Verscheure, B. Demeulenaere, J. Swevers, J. De Schutter, and M. Diehl, "Time-optimal path tracking for robots: A convex optimization approach," IEEE Transactions on Automatic Control, vol. 54, no. 10, pp. 2318-2327, 2009.

[17] R. C. Loxton, K. L. Teo, V. Rehbock, and K. F. C. Yiu, "Optimal control problems with a continuous inequality constraint on the state and the control," Automatica, vol. 45, no. 10, pp. 2250-2257, 2009.

[18] V. Azhmyakov, R. Rodriguez Serrezuela, A. M. Rios Gallardo, and W. Gerardo Vargas, "An approximations based approach to optimal control of switched dynamic systems," Mathematical Problems in Engineering, vol. 2014, no. 1, p. 795207, 2014.

[19] E. Ferrentino, A. Della Cioppa, A. Marcelli, and P. Chiacchio, "An evolutionary approach to time-optimal control of robotic manipulators," Journal of Intelligent & Robotic Systems, vol. 99, no. 2, pp. 245-260, 2020.

[20] M. Tanaskovic, L. Fagiano, and V. Gligorovski, "Adaptive model predictive control for linear time varying MIMO systems," Automatica, vol. 105, pp. 237-245, 2019.

[21] H. Lu, Q. Zong, S. Lai, B. Tian, and L. Xie, "Real-time perception-limited motion planning using sampling-based MPC," IEEE Transactions on Industrial Electronics, vol. 69, no. 12, pp. 13182-13191, 2022.

[22] J. Zhao and M. Gan, "Finite-horizon optimal control for continuous-time uncertain nonlinear systems using reinforcement learning," International Journal of Systems Science, vol. 51, no. 13, pp. 2429-2440, 2020.

[23] Y. Liu, Z. Chen, and B. Yao, "Online optimization-based time-optimal adaptive robust control of linear motors with input and state constraints," IEEE/ASME Transactions on Mechatronics, vol. 29, no. 4, pp. 3157-3165, 2024.

[24] Y. Zhang and Z. Xiang, "Prescribed-time optimal control for a class of switched nonlinear systems," IEEE Transactions on Automation Science and Engineering, vol. 22, pp. 3033-3043, 2024.

[25] J. Ye, L. Hao, and H. Cheng, "Multi-objective optimal trajectory planning for robot manipulator attention to end-effector path limitation," Robotica, vol. 42, no. 6, pp. 1761-1780, 2024.

[26] J. L. J. Pereira, G. A. Oliver, M. B. Francisco, S. S. Cunha Jr, and G. F. Gomes, "A review of multi-objective optimization: methods and algorithms in mechanical engineering problems," Archives of Computational Methods in Engineering, vol. 29, no. 4, pp. 2285-2308, 2022.

[27] J. Son, H. Kang, and S. H. Kang, "A review on robust control of robot manipulators for future manufacturing," International Journal of Precision Engineering and Manufacturing, vol. 24, no. 6, pp. 1083-1102, 2023.

[28] X. Bai, B. Li, I. Ullah, Z. Wu, S. Basheer, and A. K. Bashir, "Energy-efficient routing for IoT-enabled multi-truck multi-drone pickup and delivery systems," Applied Energy, vol. 400, p. 126546, 2025.

[29] P. Ouyang, J. Acob, and V. Pano, "PD with sliding mode control for trajectory tracking of robotic system," Robotics and Computer-Integrated Manufacturing, vol. 30, no. 2, pp. 189-200, 2014.

Valiollah Ghaffari received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Shiraz University, Shiraz, Iran, in 2006, 2009, and 2014, respectively. He is currently a Professor with the Department of Electrical Engineering, Persian Gulf University. His research interests include robust control, nonlinear control, adaptive control, and model predictive control.

IECO International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Expanding Generator Flexibility Regions in Renewable-Rich Grid via Optimal Transmission Switching and Dynamic Line Rating

Ali Beiranvand [1] | MahmoodReza Shakarami [2] | Meysam Doostizadeh [3]

Faculty of Engineering, University of Lorestan, Khorram Abad, Iran.[1,2,3]
Corresponding author's email: Doostizadeh.m@lu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 25-November-2025
Received in revised form: 07-February-2026
Accepted: 13-February-2026
Published online: 23-Sep-2026

Keywords:
Dynamic line rating,
Flexibility region,
Optimal transmission switching,
Renewable power generation,
Radar Scanning algorithm.

ABSTRACT

Today, power system operators face two major challenges: (1) determining the flexibility region for generators due to the increasing penetration of renewable generation in power systems, and (2) the rise in ambient temperature. Structural limitations in a power network and changes in its topology significantly affect the generator's flexibility region. In this paper, in addition to describing the structure of the flexibility region, the variability of these regions due to changes in the network topology is thoroughly demonstrated. For this purpose, the problem of determining the flexibility region of the generators (GFR) is integrated with the Optimal Transmission Switching (OTS) and Dynamic Line Rating (DLR) problems to develop the flexibility region and maximize the use of renewable power generation. To reduce the computational time and increase the effectiveness of the proposed method, an innovative method (referred to as Radar Scanning) is used to categorize the data related to the changes in the output power of renewable generation units. To evaluate the proposed method, the IEEE 30-bus and 118-bus power systems are employed, and the numerical results obtained from the simulation show that: first, OTS and DLR can enhance and improve the flexibility region of the generators in a power system; and second, by employing the proposed algorithm, the flexibility region of 30- and 118-bus power systems are determined by examining only a limited number of data (less than 30%), resulting in a 92% and 89% reduction in computational time, respectively.

NOMENCLATURE

DLR, DTR	Dynamic Line/ Thermal Rating	M , Ms	positive number related to the Big M method
DNE	DO-Not-Exceed limit	P_i	Base case output power of conventional unit *i*
DR	Dispatchable Region	P_{Dj}	Demand load on bus *j*
GFR	Generator Flexibility Region	R_i^+, R_i^-	Ramp-up/down limit of conventional unit *i*
LP	Linear Program	$\omega_k^e, \Delta\omega_k$	Forecasted and Uncertain output of renewable power generation unit *k*
MILP	Mixed-Integer linear program	π_{kl}	Power Transfer Distribution Factors (PTDF)
ODPs	Observed Data Points	P_i^{min}, P_i^{man}	The min/max generation of conventional unit *i*
OPF	Optimal Power Flow	H, h	matrix corresponding to the GFR
OTS	Optimal Transmission Switching	N_{sw}	Maximum number of lines for switching
PABs	Potentially Active Boundaries	θ_{max}	Maximum value of voltage angle in each bus
SOPs	soft open points method	P_i^+, P_i^-	A positive/negative change in the output of the conventional unit *i*
STR	Static Thermal Rating	y	value corresponding to P_i^+ and P_i^-
A, B, C, b0	Coefficient matrix of power system	Z_l	Binary variable for line state (1=closed,0=open)
F_l,B_l	Capacity/Susceptance of transmission line L	S^+, S^-	Slack variables

I. Introduction

With the beginning of the 21st century, due to the challenges caused by air pollution and the increasing cost of fossil fuels, the use of renewable resources such as wind and solar power has been increasing [1], [2], [3]. The increased use of renewable energy, due to its variability and high uncertainty (particularly wind power) along with rising ambient temperature, creates challenges for power systems and affects their flexibility, security and reliability.

To carry out preventive and corrective measures in the presence of renewable power generation units and to adapt the power system to them, many studies have been conducted in recent years. To operate power systems reliably and economically with renewable power generation, proposed methods often focus on adapting the power system by limiting the amount of renewable power generation. One such method is the "Do-Not-Exceed" concept, which defines the maximum range of renewable power generation that the power system can accommodate while maintaining security and reliability with minimal operational challenge [4]. The concept of the admissible region for renewable power generation is another method used in [5], [6] to assess power system risk. Another related concept is the Dispatchable Region (DR) of renewable power generation, which refers to the amount of power that the power system can tolerate. In other words, the DR represents the wind or solar power output that the power system can adapt to under given operating conditions [7], [8]. According to the research presented in [9], instead of examining all possible dispatchable regions, only the output data from renewable power generation units needs to be evaluated. This approach significantly reduces the number of boundaries that must be defined for the dispatchable region, as some theoretical boundaries will never be activated by actual renewable power outputs, making their determination unnecessary. Therefore, the concept of potentially active boundaries (PABs) was introduced.

In the research conducted in over recent years, most authors have focused on determining the flexibility and security regions of the power system in the presence of renewable sources. For example, [10] investigated problems in active distribution networks caused by the increased penetration of distributed generators (DGs) and proposed solutions to enhance network flexibility using control measures. Reference [11] explored improving flexibility and, consequently, the reliability and economy of power networks using the Soft Open Points (SOPs) method. Reference [12] examined the effect of transmission line switching on the flexibility of networks with wind power generation and showed that transmission switching can increase the network's flexibility with respect to renewable sources. Reference [13] described the security region of the power system in response to increasing renewable resource injection.

The flexibility region of generators is limited by network equipment, such as the thermal capacity of transmission lines and the reserve capacity of conventional generators. The injection of power from renewable energy sources into the power system buses can lead to overloading in the lines connected to those buses or elsewhere in network.

According to past studies, power transmission limitations in power system lines can be improved through two methods: optimal transmission switching (OTS) and the dynamic line rating determination system (DLR), also known as dynamic thermal rating (DTR) [14], [15]. The concept of OTS was introduced by [16]. This approach combines the OTS problem with the optimal power flow (OPF) problem, enabling to the relief or reduction of power system congestion and overload while maintaining network reliability at a desirable level. Past works have utilized OTS for various purposes, including achieving economic benefits, reducing operating costs, relieving line congestion, mitigating overload, serving as a corrective mechanism for voltage violations, and improving power system reliability [17].

Typically, most studies related to load flow or economic dispatch use the static thermal rating (STR), a specified and fixed value determined by assuming the worst-case weather conditions along a transmission line route [18]. However, as noted in [19], [20], if the atmospheric and environmental conditions around each line or conductor are known, the instantaneous and subsequently the dynamic thermal capacity can be determined. Studies indicate that implementing DLR can increase transmission line capacity by 15-50% in many cases and up to 150% of its STR capacity in others [21]. Furthermore, utilizing DLR capacity can reduce network congestion [22], enhance power system reliability [23], and consequently, defer or prevent the construction of new lines [24]. Reference [25] simultaneously examines the impact of using the dynamic capacity of lines and transformers to increase renewable generation injection, showing that using transformer dynamic capacity requires greater care due to its impact on reducing their useful life. Reference [26] comprehensively examines power grid transmission capacity optimization using DLR, demonstrating that this solution can mitigate the challenges associated with increased renewable resource usage .

A review of the literature reveals that prior research has predominantly focused on the mathematical determination and definition of power system flexibility or dispatchable regions. However, the impact of structural changes within the power system on these generator flexibility regions has been largely unexplored. Structural changes in a power system can include line exits, changes in the number of generation units, changes in the thermal capacity of lines, and similar factors, which can cause changes in the size of the flexibility region.

Furthermore, existing methodologies lack an efficient algorithm for processing and classifying the output data from renewable sources. This shortcoming increases computational complexity and computational time, rendering previous approaches less practical for large-scale network analysis.

Motivated by the global imperative to integrate renewable resources and address challenges such as rising ambient temperature, this paper investigates a data-driven approach for determining flexibility regions. Specifically, it examines how these regions are affected by network topology changes via OTS and capacity adjustments via DLR. A key contribution of this work is the introduction of an algorithm to identify the most impactful line for DLR deployment—a step that not been addressed in earlier works—to maximize the expansion of the system's flexibility region.

In this paper, a new formulation using a data-driven method formulated as a linear programming (LP) problem with high computational efficiency for determining the generator flexibility region is introduced. Its main innovations are as follows:

1- It reduces the complexity of determining the flexibility regions of power systems using a newly proposed model, transforming the problem from a MILP to an LP problem, and increases its efficiency for analyzing large-scale systems.

2- A heuristic algorithm called "Radar Scanning" is introduced to classify and cluster the output data of renewable resources thereby increasing computational speed and reducing computation time.

3- A heuristic algorithm is introduced to identify the most effective transmission line for installing the DLR system in order to achieve the greatest impact on the network flexibility region.

II. Defining the Dispatch Problem and GFR

A) Mathematical Formulation Related to Load Flow in a Power System

Inspired by [9], the generator flexibility region (GFR) of a power system is defined as the set of all possible renewable power generation outputs under which the system can maintain secure and reliable operation within acceptable limits. The boundaries of the GFR are constrained by the physical limits of network equipment. Therefore, the aim of this section is to determine the maximum possible GFR subject to these operational conditions. If the power output of conventional units is denoted by P_i and the power output of renewable generation is denoted by ω^e, the equations describing variations in generation and transmission power are formulated as follows.

$$\sum_i (P_i + P_i^+ - P_i^-) + \sum_k (\omega_k^e + \Delta\omega_k) = \sum_j P_{Dj} \tag{1}$$

$$-F_l \le \sum_i \pi_{il}(P_i + P_i^+ - P_i^-) + \sum_k \pi_{kl}(\omega_k^e + \Delta\omega_k) - \sum_j \pi_{jl} P_{Dj} \le F_l, \ \ \forall l \tag{2}$$

$$P_i^{min} \le P_i + P_i^+ - P_i^- \le P_i^{max}, \qquad \forall i \tag{3}$$

$$0 \le P_i^+ - \le R_i^+, \qquad \forall i \tag{4}$$

$$0 \le P_i^- - \le R_i^-, \qquad \forall i \tag{5}$$

where, equation (*1*) represents the power balance. Inequality (*2*) enforces the security limits on the maximum transmission power of the lines, which must remain within the range (F_l and -F_l) under all conditions. Constraint (*3*) is the maximum and minimum power generation capacity of the conventional units. Constraints (*4*) and (*5*) govern the upward and downward reserve capacity of each conventional unit. Due to changes in renewable energy generation, it is necessary to adjust the output of conventional units according to the corresponding reserve capacity. In equations (*1*) to (*5*), the variables P^- and P^+ represent changes in the power output of conventional generation units following a change $\Delta\omega$ in renewable power generation.

B) Formulation of the Generator Flexibility Region

In general, all equality and inequality constraints from (*1*) - (*5*) can be compactly expressed in the following form.

$$AP + B\begin{bmatrix} P^+ \\ P^- \end{bmatrix} + C(\omega^e + \Delta\omega) \le b_0 \tag{6}$$

where, the coefficients from the power system constraints (*1*) - (*5*) are represented by matrices A, B, C, and the vector b_0. In equation (*6*), the variable $\begin{bmatrix} P^+ \\ P^- \end{bmatrix}$ can be substituted with a new variable y. Furthermore, since ω^e is a constant, the corresponding terms in inequality (*6*) can be moved to the right-hand side. Thus, the formulation can be rewritten as follows:

$$AP + By + C\Delta\omega \le b \tag{7}$$

where, $b = b_0 - C\omega^e$.

Finally, the generator flexibility region can be defined as follows [27]:

$$W^{GFR} = \{\Delta\omega | \exists y: By + C\Delta\omega \le b - AP\} \tag{8}$$

The region W^{GFR} defined in equation (*8*) is formed by the boundary lines or planes arising from the power system constraints given in equations (*1*) - (*5*). Research in [9] indicated that ODPs activate only a limited subset of these GFR boundaries, termed Potentially Active Boundaries (PABs). A boundary of the GFR is activated mathematically whenever a set of ODPs violates inequality (*8*). It was demonstrated in [27] that inequality (*8*) can be reformulated as follows.

$$W^{GFR} = \begin{Bmatrix} \Delta\omega | u^T(C\Delta\omega) \ge u^T(b - AP), \\ \forall u \in vert(U) \end{Bmatrix} \tag{9}$$

$$U = \{u | B^T.u = 0, \qquad -1 \le u \le 0\} \tag{10}$$

where $vert\ (U)$ denotes the set of vertices of polyhedron U, and the superscript T represents the matrix transpose. The activation condition for a boundary I of the GFR is:

$$I = \{\exists \Delta\omega \in \Omega, u^T(C\Delta\omega - b + AP) < 0\} \tag{11}$$

where Ω denotes the total observed points related to the outputs of renewable power generation.

C) Process for Determining the GFR Boundaries

The process for determining the boundaries of the GFR follows a methodology similar to the one described in reference [9] for determining the dispatchable region. The key distinction, however, is that this paper formulates the final problem as a Linear Programming (LP) model, whereas the reference method uses a Mixed-Integer Linear Programming (MILP) problem. A further significant contribution of the proposed method is its use of an innovative data categorization and ranking technique to identify and evaluate the effective data points critical for defining the GFR boundaries. This technique will be detailed in the subsequent section. The objective is to determine the active boundaries, provided that the union of the regions outside the GFR, resulting from both the current boundary and all previously identified boundaries, is either an empty set or remains mutually exclusive with any existing region. Mathematically, the activation condition for a boundary is equivalent to:

$$W_1' \bigcup \emptyset = D \rightarrow \begin{cases} if\ D = \emptyset \rightarrow PAB_1\ is\ deactive. \\ if\ D \neq \emptyset \rightarrow PAB_1\ is\ active. \end{cases}$$

$$W_i' \cap \left(\bigcup_{j=1}^{i} W_j' \right) = k$$

$$\begin{cases} k = \emptyset \rightarrow PAB_i\ is\ active. \\ k = W_j' \rightarrow PAB_j\ is\ deactive\ \&\ PAB_i\ is\ active. \\ k = W_i' \rightarrow PAB_i\ is\ deactive. \\ k = W'', W'' \neq W_i',\ W'' \neq W_j' \rightarrow PAB_i\ is\ active \end{cases} \tag{12}$$

where W_i' and W_j' represent the regions outside the GFR, separated by boundaries i and j, respectively.

The logic of Equations (12) follows that after evaluating each point of the ODPs set using the proposed algorithm, the conditions related to set D (the first two equations in (12)) are checked first. This step aims to activate effective boundaries while preventing the activation of ineffective or redundant ones.

The procedure is as follows:

a) For boundary number 1, the condition is compared with the empty set ϕ. If W_1' has no members (i.e. D=ϕ), no boundary is activated. If it has members (i.e. D≠ ϕ), then boundary PAB_1 is activated.

b) For any subsequent boundary *i* activated by the ODPs, the second part of Equation (12) (involving set k) must be examined. The set of points separated by the new boundary *i* (W_i') is compared with the union of all points outside the final flexible from previously active boundaries:

Criterion 1: If W_i' adds new points to the total excluded set, boundary *i* is accepted as a new boundary, otherwise no new boundary is created.

Criterion 2: If all points excluded by a prior boundary *j* are entirely contained within W_i' (i.e., $W_i' \cap \left(\bigcup_{j=1}^{i} W_j'\right) = W_j'$), then boundary *j* becomes ineffective and is removed and replaced by boundary *i*.

The conditions specified in (12) are used to prevent the formation or activation of ineffective boundaries for the GFR. The proposed process for determining the W^{GFR} space in this paper is as follows:

Step 1) The initial operating point values of the system (P_i and ω_k^e) are determined.

Step 2) All ODPs are represented as Ω set.

Step 3) The phase and magnitude values of all ODPs are determined in polar coordinates and sorted based on their phase value.

Step 4) The data space is divided into N_{scan} segments, and all data points are allocated to these segments based on their phase values. Within each segment, the data points are sorted in descending order of magnitude.

Step 5) The counter for the parts inside the space is set to SC_n=1.

Step 6) The highest-ranked (largest magnitude) data point within segment SC_n is assigned to $\Delta\omega_k$, represented by the counter k=1.

Step 7) The GFR (W^{GFR}) is initially assumed to be sufficiently large to contain all ODPs. This can be formulated as $W^{GFR} = \{\Delta\omega | H.\Delta\omega \geq h\}$, where matrix H and vector h are updated iteratively during the boundary determination process.

Step 8) If $\Delta\omega_k$ satisfies the condition in step 7, problem (13)-(15) is considered solved for it. In this case, the segment counter k is incremented by one. If the condition is not satisfied, the next data point with the largest magnitude will be selected from the same segment (*k = k + 1*). To determine the boundary of the GFR, the following LP problem is solved for the output data $\Delta\omega_k$.

$$\max u^T (b - AP - C\Delta\omega_k) \tag{13}$$

$$s.t.: B^T.u = 0 \tag{14}$$

$$-1 \leq u \leq 0 \tag{15}$$

Step 9) If the constraint $u^{*T}(C\Delta\omega_k) \geq u^{*T}(b - AP_i)$ is satisfied for the optimal solution u^*obtained from the problem (13)-(15), then there is no need to establish a new boundary, and the algorithm returns to step 8. However, if the constraint $u^{*T}(C\Delta\omega_k) \geq u^{*T}(b - AP_i)$ is not satisfied, the existing GFR (W^{GFR}) is updated by adding a new boundary as $W^{GFR} = W^{GFR} \cap \{u^{*T}(C\Delta\omega - b + AP) \geq 0\}$ and considering equations (13)-(15). The values of matrix H and vector **h** are updated by adding the values of $u^{*T}C$ and $u^{*T}(b - AP)$, respectively. The algorithm then proceeds to step 10.

Step 10) The condition presented in equation (12) is applied to select the active boundary and then returned to step 6.

Step 11) the value of counter k is incremented by one unit.

Step 12) If SC_n is less than the total number of segments, it adds one unit, and then the algorithm returns to step 6. Otherwise, the algorithm proceeds to step 13.

Step 13) End.

By executing the above algorithm, if any ODPs from the set Ω within a segment violate the condition $u^{*T}(C\Delta\omega_k - b + AP) \geq 0$, the corresponding boundary line will be activated.

III. Formulating the OTS in the Presence of Renewable Sources

The GFR in a power system is not rigid or fixed, as it depends heavily on the network structure. Changes in the network structure, such as taking transmission lines out of service, can cause variations in these regions. Therefore, this variable characteristic of the GFR can be utilized and developed according to the needs of the power system operators. One available corrective measure is OTS. The concept of OTS involves temporarily taking one or more transmission lines out of service at specific times to allow the power system to operate more efficiently. OTS can be implemented based on either AC Optimal Power Flow (ACOPF) or DC Optimal Power Flow (DCOPF). However, due to its simplicity, reasonable accuracy, and computational speed, DCOPF is typically preferred for research studies. Consequently, this paper employs DCOPF to assess the impact of OTS on the GFR. According to [28], [29], integrating OTS and DCOPF can be expressed as problem (16)-(23), which is a MILP problem.

$$\min \sum_{i} C_i P_i \tag{16}$$

$$-P_l + B_l(\theta_i - \theta_j) + M(1 - Z_l) \geq 0, \;\; l \in L \tag{17}$$

$$-P_l + B_l(\theta_i - \theta_j) - M(1 - Z_l) \leq 0, \;\; l \in L \tag{18}$$

$$\sum_{i \in g} P_i - \sum_{i \in D} P_{Di} = \sum_{l \in L} P_l \tag{19}$$

$$-\theta_{max} \leq \theta_i - \theta_j \leq \theta_{max} \tag{20}$$

$$P_i^{min} \leq P_i \leq P_i^{max} \tag{21}$$

$$-Z_l . F_l \leq P_l \leq Z_l . F_l \tag{22}$$

$$\sum_{l \in L} (1 - Z_l) \leq N_{sw} \tag{23}$$

where, expression (16) denotes the objective function. Inequalities (17) and (18) define the power flow on each line and its relationship to the voltage angles at the terminal buses. In these constraints, the Big-M parameter is a sufficiently large positive constant and Z_l is a binary variable representing the on/off status of line l (where $Z_l = 1$ indicates the connected line and $Z_l = 0$ indicates a disconnected line). Equation (19) enforces the balance of generation, consumption and transmission power at each bus. Constraints (20) - (22) impose limits on bus voltage angle differences, the active power output of the conventional generators, and the line transmission capacities, respectively. Constraint (23) restricts the maximum number of permitted switching operations in the system.

To adapt the OTS problem formulated in (16)-(23) for purpose of improving the GFR, several specific modifications are required. Accordingly, The proposed OTS formulation for improving the GFR in this paper is as follows.

$$\min \sum_{i \in bus} [M_s . (S_i^+ + S_i^-)] \tag{24}$$
s.t: (17), (18), (20), (22), (23)

$$\sum_{i} (P_i + P_i^+ - P_i^-) + \sum_{i} (S_i^+ - S_i^-) + \sum_{k} (\omega_k^e + \Delta\omega_k) - \sum_{j} P_{Dj} = \sum_{l \in L_{bus}} P_l \tag{25}$$

$$\sum_{i} (P_i^+ - P_i^-) \leq \sum_{k} \Delta\omega_k \tag{26}$$

The objective function of the resulting MILP problem is given in equation (24). Its purpose is to minimize the costs associated with the increasing or decreasing the power output of conventional units. The variables S_i^+ and S_i^- are slack variables similar to P_i^+ and P_i^-, introduced to ensure the problem's feasibility. The coefficient M_S, a large positive number, penalizes these slack variables to minimize their values under normal power system conditions. These slack variables measure the deviation in output of conventional units from their allowable reserve capacity limits. Equation (25) represents the DC power balance constraint at each bus, which now includes the power from renewable generators as well as the adjustments from conventional units. Inequality (26) is a new condition introduced in this paper. It states that the total adjustments (increase or decrease) in conventional generation must be less than or equal to the total change in renewable power output.

It is important to note that a power system's operating state can change with variations in renewable generation output. Finding a single optimal switching solution for all possible operating points can be computationally prohibitive or infeasible. To address this and find an optimal solution for problem (24)-(26) given a specific change in renewable output $\Delta\omega_k$, the following procedure is adopted.

1-The data points already identified during the determination of W^{GFR} by solving problem (13)-(15) are not required to be re-evaluated in the OTS problem (24)-(26). Only points outside the current W^{GFR} require assessment. The rationale is that if a feasible OTS solution exists for a point outside W^{GFR}, the constraints (13)-(15) guarantee a solution for all points inside it. Conversely, if no feasible solution exists for points outside W^{GFR}, the status of interior points becomes irrelevant for this specific $\Delta\omega_k$.

2- As stated above, since problem (24)-(26) is evaluated for all points outside W^{GFR}, the resulting binary variable Z_l (line status) are stored for every such ODP and for each step.

3- The sorted Z_l values are then analyzed to determine the optimal line(s) to switch out for the given $\Delta\omega_k$. The line or set of lines with the highest frequency of $Z_l = 0$ across the evaluated points is selected as the optimal switching action.

4- After identifying the line(s) to be taken out of service, the system matrices A, B, C, and vector b in (*1*)-(*5*) are updated according to the new network topology. The proposed algorithm from section II is then repeated for points outside the initial W^{GFR} to determine a new, enhanced flexibility region, denoted W^{GFR-S}.

It is possible that the GFR may decrease if a transmission line is taken out of service for non-optimal reasons (e.g., maintenance or outage due to a fault). Such an outage can change the power flow on the lines, potentially causing congestion on other lines and reducing the system's ability to adapt to changes in renewable power generation. Consequently, system operators must re-evaluate the GFR following any change in network topology.

IV. Effect of DLR on the GFR

Transmission lines are affected by various atmospheric conditions and environmental factors, including ambient temperature, wind speed, altitude, time of day, and angle between wind and conductor. These factors cause the thermal capacity of transmission lines to be time-varying. In power system studies, a fixed Static Thermal Rating (STR) is often used to limit the line power flow. This rating (capacity) is determined based on the worst-case conditions for the line, which can be a conservative estimate for many hours of the day or even for most times of the year. In recent years, significant research has focused on using DLR systems [17]. Various systems are now employed to determine the real-time dynamic capacity of power lines, enabling the control of current flow at any given moment [30]. Conductor temperature depends on the current passing through it. If the current exceeds the permissible limit, the conductor temperature rises. This can reduce the required clearance to the objects below and, in more severe cases, cause the conductor temperature to exceed safe limits, reducing mechanical strength and potentially leading to line damage. The detailed calculations describing how line thermal capacity varies with environmental conditions are fully explained in references [19], [20]. This paper utilizes the results from a hypothetical DLR system installation for analytical purposes.

To assess the effect of the DLR system on the GFR, the first step is to identify the transmission line that limits the GFR. For this purpose, the power flow through the lines is calculated using constraint (*2*) for the data points inside the GFR that are closest to its active boundary. The line(s) with power flow values near their static thermal capacity are then identified as the limiting line(s). Consequently, the GFR can be expanded or contracted by adjusting the power transmission capacity of these line(s).

The algorithm for evaluating the effect of changing line thermal capacity on the GFR is as follows.

Step 1) the limiting line is identified as described above.

Step 2) the thermal capacity of the limiting line is adjusted by a percentage of its static thermal rating. This adjustment can be an increase or decrease.

Step 3) Using Equations (*1*)-(*5*), the coefficients A, B, C and **b** are updated.

Step 4) the algorithm proposed in Section II is repeated for all points outside the initial W^{GFR} to determine a new region, denoted W^{GFR-L}.

V. Introducing the Innovative Algorithm for Classifying Renewable Power Output Data

Reducing computation time in large-scale power systems is primary concern. For this purpose, an innovative classification method called the "Radar Scanning" algorithm is introduced to categorize the output data of renewable units, thereby decreasing the overall calculation time.

A) Data Classification Based on Innovative Radar Scanning Algorithm

If the number of buses with renewable power generation is n_w, the analytical space of the flexibility region is n_w-dimensional. Analyzing and understanding an n_w-dimensional space, where n_w>3, is extremely complex. To simplify the analysis of renewable generation output points, a two-dimensional space is utilized. The output data from n_w different network buses are examined pairwise within a two-dimensional space to determine their positions. The two-dimensional Radar Scanning algorithm is then applied to classify the data. In an n_w-dimensional space, the number of possible two-dimensional subspaces created is given by:

$$n_{2D} = \frac{n_w!}{(n_w-2)!\,2!} \tag{27}$$

where n_{2D} represents the total number of n_{2D} subspaces established in an n_w-dimensional space. For all data points corresponding to the changes in renewable power output ($\Delta\omega_k$), their phase and magnitude values in polar coordinates are determined as follows.

$$\begin{aligned} &\Delta\omega_k = r_k \measuredangle \theta_k \;, \qquad 0 \le \theta \le 2\pi \\ &r_{ki} = \sqrt{\Delta\omega_{ki,1}^2 + \Delta\omega_{ki,2}^2} \\ &\theta_{ki} = \tan^{-1}\left(\Delta\omega_{ki,2}^2 / \Delta\omega_{ki,1}^2\right) \end{aligned} \tag{28}$$

where r is the magnitude and θ represents the phase of the data in the two-dimensional subspace i. The polar representation of a renewable generation output point is illustrated in Fig. 1. To enhance the algorithm's effectiveness, it is preferable to use the magnitude of the data point in the original n_w-dimensional space for the value of r in Equation (28). In general, increasing the number of segments improves the accuracy of the program at the cost of longer calculation time. The number of segments can be selected between 20 and 72, resulting in a segment angle ranging from 18° to 5°.

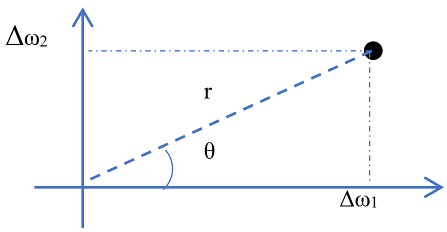


Fig. 1. Polar coordinate representation of a ODP.

The data space division by the Radar Scanning method is shown in Fig. 2. The scanning process and division of the region proceeds counterclockwise, starting from the positive horizontal axis. To facilitate a clearer understanding of the proposed innovative Radar Scanning algorithm in two-dimensional space, the steps for its application are presented sequentially as follows.

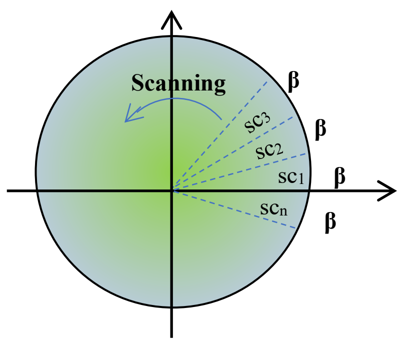


Fig. 2. Data space segmentation via the Radar Scanning algorithm.

Radar Scanning Algorithm

1- For each data point, the dominant two-dimensional subspace is determined by identifying its two components with the largest magnitudes.

2- Within each identified two-dimensional subspace i, the amplitude and phase of all points are calculated using Equation (28). It is recommended to use the point's magnitude the original n_w-dimensional space rather than its magnitude in the two-dimensional subspace.

3- All points in the subspace are sorted in ascending order based on their phase value.

*4- The two-dimensional subspace i is divided into SC_n equal angular segments (see **Fig. 2**), and points are assigned to segments according to their phase.*

5- Within each segment, the points are sorted again, this time in descending order based on their overall n_w-dimensional magnitude.

B) Process of the Innovative Radar Scanning Algorithm

Analytically, the GFR is continuous and lacks singular points. Consequently, it is not necessary to evaluate all renewable output data points. Instead, only the highest-ranked points by magnitude within each segment need to be assessed. For this purpose, the process begins with the first segment (lowest phase angle), selecting the top-ranked point (largest magnitude). If this point satisfies the condition in step 7 of the algorithm from section II-C (i.e., lies within the current GFR), the algorithm processes only this point and moves to the next segment. If the selected point does not satisfy the condition (i.e., lies outside the GFR), it is recorded as an external point, and the next highest-ranked point within the same segment is evaluated. This process repeats for all segments. The algorithm' counter-clockwise traversal of the two-dimensional subspace resembles the operation of a military radar system, hence the name "Radar Scanning".

This method ensures that all points near the GFR boundary are evaluated while avoiding unnecessary computation on low-ranking interior points that do not define the region's limits. This significantly reduces the overall computational time.

VI. Numerical Simulation Results

This section, analyzes the proposed methods and algorithms for the developing and modifying the GFR through OTS and DLR. The simulation results are compared against a base case that applies neither corrective action. All simulations are performed on the IEEE 30-bus and IEEE 118-bus test systems, with data from [31] and [9]. The parameter M is set to 20,000, and the reserve capacity of conventional units is set to 25% of their maximum output, following [27]. Simulations were executed on a PC with an Intel Dual Core E 5200 processor and 2.0 GB RAM.

1) Case 1: IEEE 30-Bus System Test

The IEEE 30-bus system comprises 30 buses and 41 transmission lines, with a total load of 189.2 MW and a maximum conventional generation capacity of 335 MW. To study the GFR in this test system, as per [9], renewable generation units with base generation capacities of 10 MW and 20 MW are installed at buses #1 and #22, respectively, resulting in a renewable penetration rate of 15.86%. The line loading under these base conditions is shown in Fig. 3. Lines 11, 28 and 32 exhibit the highest loading percentage and are most prone to congestion if the renewable power injection increases.

A dataset of 500 ODPs is used for renewable output (data-driven method). For this purpose, first, the GFR is determined for the base case without OTS or DLR.

A) GFR Determination without OTS and DLR

As outlined in section II, the GFR is determined by using the proposed method and innovative Radar Scanning algorithm applied to the ODP set. This is formulated as a LP problem ((13)-(15)), solved using the ODP data to establish a data-driven GFR. The base case GFR (i.e., without OTS and DLR) is shown in Fig. 6. The region is enclosed by a single active boundary. In this case, 72 ODPs lie outside the GFR (shown in red). The activated boundary is given by $[0 \quad -0.4864].\Delta\omega \geq [-0.1455]$. This boundary is primarily activated by increased power injection from the renewable generator at bus #22. The π coefficients for this bus and all transmission lines are shown in Fig. 4. Lines 1, 2, 6, 12, 16, 20, 32 and 33 have notably higher coefficients. As shown inFig. 5, a gradual increase in power injection at bus #22 causes line #32 (between buses 21-22) to congest first, limiting further power transfer. This activated boundary must be considered as a constraint in power flow solutions to ensure secure system operation when renewable generation changes.

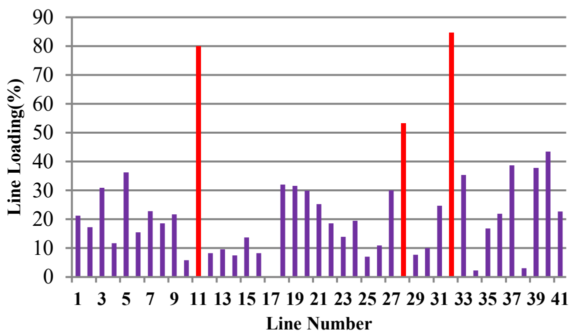


Fig. 3. Line loading percentages under base case condition.

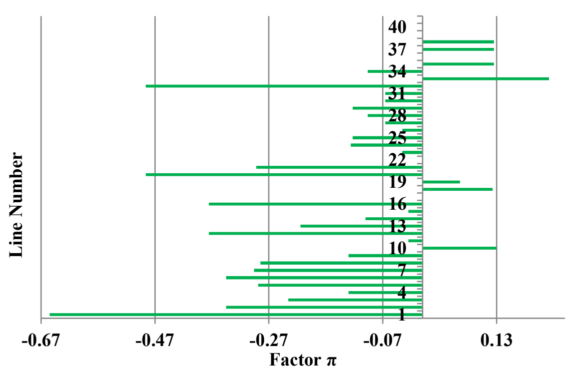


Fig. 4. π coefficients for power injection at bus 22.

It is important to note that the identification of the limiting line depends on several interconnected factors: the π coefficient, the initial line loading, and the line's capacity. This combination explains why line #32 becomes congested earlier than others. Specifically, an initial load exceeding 80%, a π coefficient of approximately 0.48, and a capacity of only 32 MVA, makes line #32 the most vulnerable to congestion. In contrast, although line #1 has the highest π coefficient (0.65), it does not congest due to its low initial loading (20%) and its significantly higher capacity (130 MVA).

B) Effect of OTS on the GFR

This section analyzes the effect of OTS on the GFR using the algorithm proposed in Section III. The MILP problem formulation in (24)-(26) is applied to ODP points located outside the base GFR to identify the optimal transmission line that should be taken out of service.

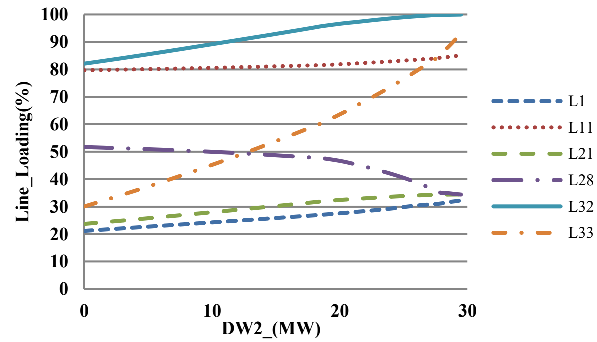


Fig. 5. Line loading progression with increased power injection at bus 22 (base case).

outside the base GFR to identify the optimal transmission line that should be taken out of service.

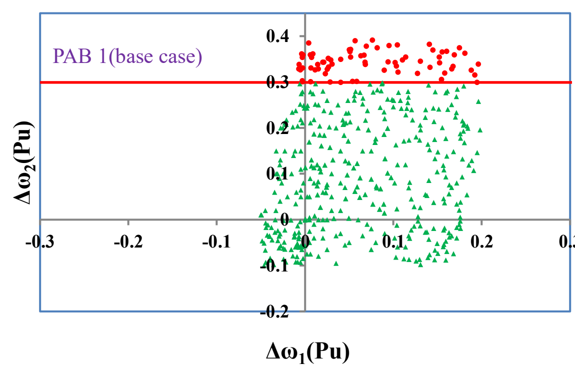


Fig. 6. Baseline GFR without corrective actions (base case).

In this system test, for the ODPs outside the GFR, line 20 (10-21) is identified as the optimal line to switch out, as it has the highest frequency of selection among all lines. TABLE I shows the selection count for each line, line 20 (10-21) has the highest value. For this analysis, the maximum number of allowed switching operations, N_{sw}, was set to 1. When N_{sw} is increased to 2, the optimal pair of lines to switch out are lines 20 (10-21) and 1 (1-2), in that order, as shown in the second column of TABLE I.

TABLE I Optimal Line Switching Frequencies for GFR Enhancement via OTS.

N_{SW}=1	N_{SW}=2
Line(No. of Selection)	Line(No. of Selection)
10-21(247)	10-21(190)
6-9(69)	1-2(175)
1-3(38)	22-24(53)
6-10(24)	23-24(40)
10-17(15)	6-8(38)

Following the switching of the selected line (20 (10-21)) and the subsequent update of the system matrices A, B, C and vector b, the GFR is recalculated using the proposed algorithm from Section II. The resulting region is shown in Fig. 7. A significant expansion of the GFR is observed after the switching action, incorporating many more ODPs. Consequently, the number of points outside the new GFR is reduced to only 9, compared to 72 in the base case. This development of the GFR through optimal switching enables the power system to accommodate a greater amount of power injection from renewable units without violating security constraints. The change in loading pattern of lines is shown in Fig. 8. As the figure indicates, line # 33 (22-24) becomes the new limiting constraint, congesting earlier than others. This outcome was predictable: switching out line 20 (10-21) fixed the power flow on the now-radial line 32 (21-22). Therefore, as renewable injection at bus #22 increases, the subsequent power flow — following the path indicated in Fig. 5— causes line 33 (22-24) to reach its thermal limit first.

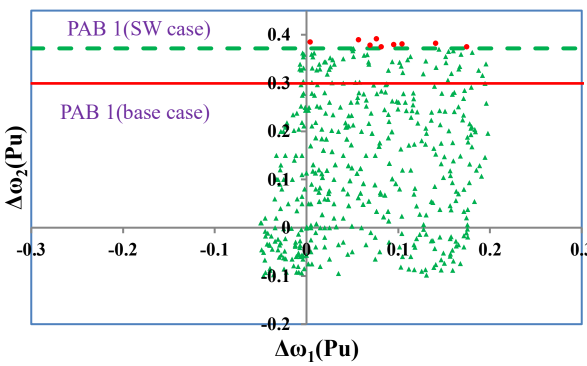


Fig. 7. Expansion of the GFR with OTS.

c) Effect of DLR on GFR

This section evaluates the impact of the Dynamic Line Rating (DLR) system on the GFR, assuming the DLR system has been optimally installed on a transmission line that limits power transfer. While the optimal placement methodology of the DLR system itself is beyond the scope of this paper, guidelines are provided in section IV. For this analysis, a DLR system is assumed to be installed on a suitable line. Its output is modeled as a percentage of the line nominal static thermal rating and incorporated into the algorithm from section IV. For the test system, the proposed algorithm identifies line 32 (21-22) as the most suitable candidate for DLR installation. To analyze the sensitivity, the line's dynamic capacity is set to 0.95 and 1.05 times its nominal static capacity, reflecting the environmental and atmospheric conditions. After updating the coefficients A, B, C and b with these adjusted capacities, the resulting GFR regions are shown in Fig. 9 and Fig. 10, respectively. The results clearly demonstrate that a reduction in transmission capacity of line 32 (21-22) contracts the GFR, whereas an increase in its capacity expands the region. Therefore, it can be concluded that at the time of connecting renewable power generations to the power system, it is necessary to carry out studies related to the thermal capacity of the lines at various times carefully so that the line does not get overloaded due to the injection of additional power in one bus of the network.

Considering the arrangement of ODPs shown in Fig. 6, if Δω increases excessively, according to Equation (*1*), a reduction in the generated power of conventional generators

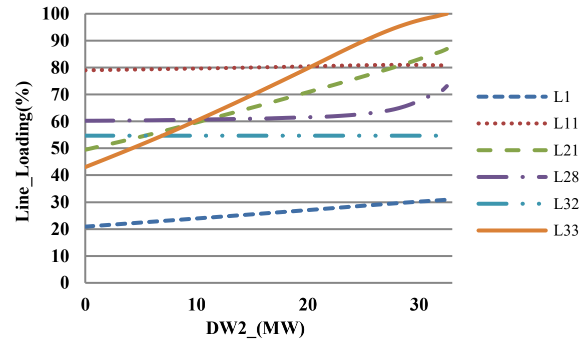


Fig. 8. Post-switching line loading analysis for increased power injection at bus 22.

is necessary to maintain power balance. If this required adjustment exceeds the reserve capacity of conventional generators defined by Equations (*3*) to (*5*), then Equation (*6*) and its dual condition, $u^T(b - AP - C\Delta\omega) \leq 0$, are violated. In addition, based on Equation (*2*), which is related to the thermal capacity of the lines, a high value of Δω, considering the π coefficients (shown in Fig. 4) and line flow dependencies, can increase the power flow in one or more lines beyond their nominal capacity. Consequently, Equations (*2*) and (*6*), along with the condition $u^T(b - AP - C\Delta\omega) \leq 0$, are violated.

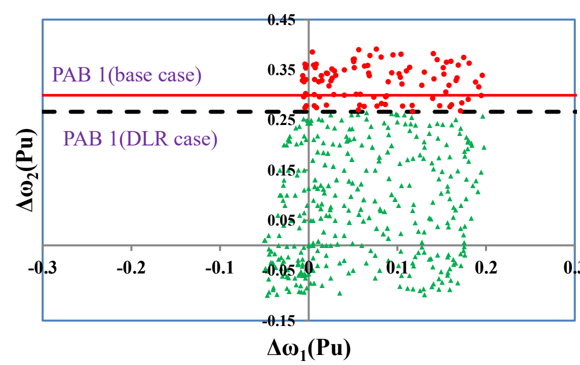


Fig. 9. GFR reduction under decreased line capacity with DLR.

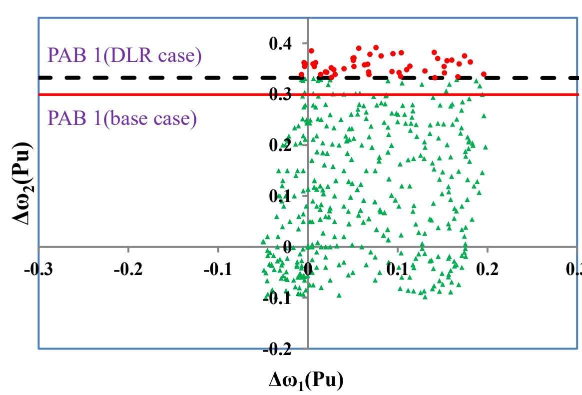


Fig. 10. GFR sensitivity to reduced line thermal limit.

Based on Inequality (*2*), it is possible to determine the capacity of the line 32 (21-22) in such a way that the active boundary that causes the most significant restriction for the GFR becomes inactive. As shown in Fig. 9 and Fig. 10, boundary 1 (when active) excludes 98 and 48 ODPs from the GFR under different capacity conditions. Applying Inequality (*2*) to the outermost excluded point yields an optimal capacity of 37.3 MVA for line 32 (21-22), which is 16.6% above its nominal static capacity. Applying this change in the problem deactivates the boundary 1, incorporates all ODP points into the GFR, and thereby expands the region. Fig. 11 shows the resulting GFR and confirms that the capacity adjustment on line 32 (21-22) has inactivated the boundary and developed the GFR.

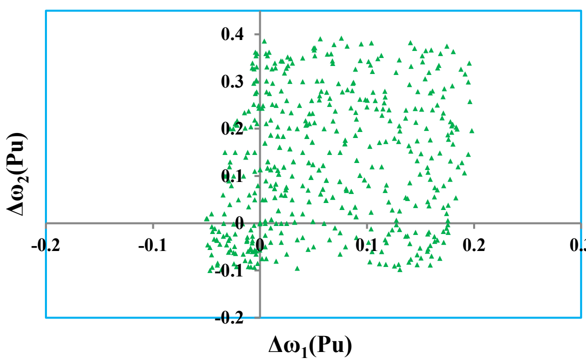


Fig. 11. GFR expansion via DLR enhancement (line 32(21-22)).

2) Case 2: IEEE 118-Bus System Test

To evaluate the proposed algorithm's scalability, it is analyzed on the IEEE-118 bus system described in reference [9] using 500 ODPs.

The IEEE 118-bus system comprises 118 buses and 185 transmission lines, with a total load of 4242 MW and a maximum conventional generation capacity of 5859.2 MW. To study the GFR in this test system, renewable generation units with base capacities of 170MW and 120MW are installed at buses #49 and #72, respectively, resulting in a renewable penetration rate of 6.84%.

The simulation results of the effect of OTS and DLR capacity utilization are shown in TABLE II. The limiting lines in this sample power system are 119 (69-77), 112 (65-68), 50 (30-38), 123 (71-72) and 120 (70-71), with initial loading of 93.28%, 27.8%, 89.54%, 37.44% and 37.7%, respectively. Optimal switching takes line 112 (65-68) out of service and, according to TABLE II, increases the GFR by about 10% compared to the base case. In this sample power system, using the algorithm proposed in Section IV, lines 120 (70-71) and 123 (71-72) are selected as the optimal lines for installing the DLR system. According to TABLE II, if the output of the DLR system increases the capacity of the mentioned lines by 15%, then the GFR increases by about 11.24%, and vice versa, if the capacity of the mentioned lines decreases by about 15%, the GFR also decreases by about 9.24%

TABLE III demonstrates that the Radar Scanning heuristic algorithm significantly enhances computational efficiency.

TABLE II Simulation results for the IEEE 118-bus test system.

Case	Number of ODPs inside the GFR	Increase in GFR compared to the base case (%)
Base	249	0
OTS	273	9.64
DLR (+15%) (increase mode for Lines 120 and 123)	277	11.24
DLR (-15%) (decrease mode for Lines 120 and 123)	226	-9.24

To evaluate and compare the performance of the proposed algorithm, its results are benchmarked against existing methods in TABLE III. The table clearly demonstrates that the proposed algorithm is significantly faster than previous approaches in both small- and large-scale systems, making it particularly suitable for large-scale applications.

TABLE III Performance comparison of the proposed algorithm with prior algorithms.

System Test	Number of ODPs	Computational time (s)		
		Without using the algorithm	With using the proposed algorithm	The proposed-nPrl algorithm in [9]
IEEE-30 Bus	250	17.52	0.9425	1.6292
	500	36.44	3.0921	8.3078
IEEE-118 Bus	250	44.628	6.7112	11.8062
	500	96.183	10.5783	29.378

By classifying data based on phase and amplitude, it reduces the calculation time for determining the GFR by approximately 92% for the 30-bus system (from 36.44 to 3.09 seconds) and 89% for the 118-bus system (from 96.183 to 10.5783 seconds), compared to not using the algorithm. In addition, the calculation time for determining the GFR with the method proposed in this paper for both the 30-bus and 118-bus test systems is much less than the method proposed in reference [9]. In the case of using 500 ODPs, the algorithm proposed in this paper reduces the computation time in the 30-bus test system by about 63% compared to the computation time of the method proposed in reference [9] (3.0921 seconds versus 8.3078 seconds). This is also true for the larger system, i.e. 118 buses, and the proposed algorithm reduces the computation time by about 64% compared to the proposed method in the reference (10.5783 seconds versus 29.378 seconds). These results, summarized in TABLE III, strongly indicate the superior performance of the proposed algorithm for determining power system flexibility regions and optimizing the utilization of installed renewable generation capacity, especially in large-scale applications.

It is important to note that the time required to perform the GFR calculation and the accuracy of the result depend partly on the number of segments in the Radar Scanning (SC_n). Fig. 12 shows the computation time and solution accuracy, number of points outside the GFR), as a function of the segment angle (the inverse of the number of segments). As can be seen from Fig. 12, the relationship between accuracy and time is inversely proportional. The greater the number of segments in the Radar Scanning method (i.e., the smaller the segment angle), the greater the solution accuracy, but the computation time also increases, and vice versa. Therefore, it is necessary to select a suitable trade-off between accuracy and speed by finding the optimal value for the number of segments in the Radar Scanning algorithm, as shown in Fig. 12.

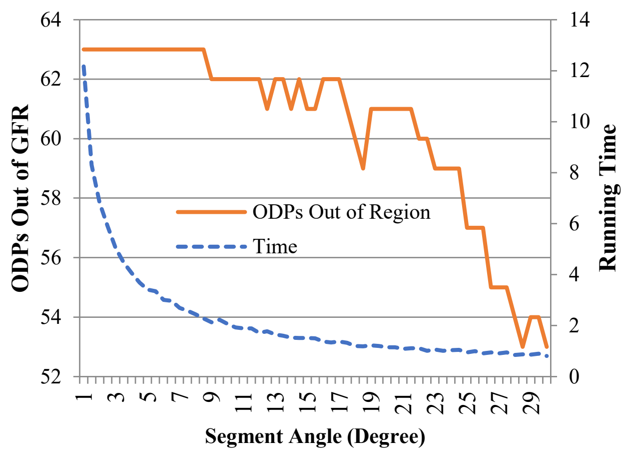


Fig. 12. Computational performance of the Radar Scanning algorithm for the 30-bus system

VII. Conclusion

This paper has focused on analyzing changes in the flexibility region of the generators (GFR) resulting from modifications to the network structure, specifically through optimal transmission line switching and thermal line capacity adjustments. The GFR was defined based on variations in renewable generation outputs, and two distinct methods were proposed to develop and expand this region: (1) Optimal Transmission Switching (OTS) and (2) utilization of Dynamic Line Rating (DLR). For each method, an iterative algorithm was introduced to efficiently solve associated MILP and LP problems. To reduce the computational time and enhance the effectiveness of the proposed method, an innovative Radar Scanning algorithm was introduced. This method classifies renewable generation output so that only the most critical points are evaluated, making the whole process much faster. The computational efficiency was further improved by restricting the OTS and DLR evaluations to data points lying outside the initial GFR region. The application of proposed algorithms demonstrated that OTS can expand the GFR, enabling greater permissible injection from renewable sources into the network. Similarly, by applying the DLR, which represents a percentage of its static capacity, the GFR also changes, so that with the increase of the limiting line thermal capacity, the GFR region expands, and with the decrease of this thermal capacity, the GFR contracts.

Several important aspects remain for future work. These include integrating the cost coefficients of conventional generation units to economically optimize the adjustments P^{+} and P^{-}, as well as incorporating full AC power flow constraints into the proposed algorithms. Such measures allow for a more comprehensive and practical assessment of the solutions resulting from the OTS and DLR implementation. Also, the integrated consideration of DLR and OTS within a unified framework offers a valuable avenue for future work to further enhance the flexibility region.

References

[1] C. Wang, Z. Gong, Y. Liang, W. Wei, and T. Bi, "Data-driven wind generation admissibility assessment of integrated electric-heat systems: A dynamic convex hull-based approach," *IEEE Transactions on Smart Grid,* vol. 11, no. 5, pp. 4531-4543, Sep. 2020, doi: https://doi.org/10.1109/TSG.2020.2993023.

[2] Z. Shen, W. Wei, T. Ding, Z. Li, and S. Mei, "Admissible region of renewable generation ensuring power flow solvability in distribution networks," *IEEE Systems Journal,* vol. 16, no. 3, pp. 3982-3992, Sept. 2022, doi: https://doi.org/10.1109/JSYST.2021.3138908.

[3] Y. Lei, B. Li, J. Meng, Y. Su, A. Li, and F. Liu, "The Changeable Region of Wind Power Considering Its Uncertainty," in 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Sep. 2020, pp. 1-5.

[4] J. Zhao, T. Zheng, and E. Litvinov, "Variable resource dispatch through do-not-exceed limit," IEEE *Transactions on Power Systems,* vol. 30, no. 2, pp. 820-828, Mar. 2014, doi: https://doi.org/10.1109/TPWRS.2014.2333367.

[5] C. Wang, F. Liu, J. Wang, W. Wei, and S. Mei, "Risk-based admissibility assessment of wind generation integrated into a bulk power system," *IEEE Transactions on Sustainable Energy,* vol. 7, no. 1, pp. 325-336, Nov. 2015, doi: https://doi.org/10.1109/TSTE.2015.2495299.

[6] Y. Zhang, X. Han, B. Xu, M. Wang, P. Ye, and Y. Pei, "Risk–based admissibility analysis of wind power integration into power system with energy storage system," *IEEE Access,* vol. 6, pp. 57400-57413, Sep. 2018, doi: https://doi.org/10.1109/ACCESS.2018.2870736.

[7] W. Wei, F. Liu, and S. Mei, "Dispatchable region of the variable wind generation," *IEEE transactions on power systems,* vol. 30, no. 5, pp. 2755-2765, Sep. 2014, doi: https://doi.org/10.1109/TPWRS.2014.2365555.

[8] Y. Liu, Z. Li, Q. Wu, and H. Zhang, "Real-time dispatchable region of renewable generation constrained by reactive power and voltage profiles in AC power networks," *CSEE Journal of Power and Energy Systems,* vol. 6, no. 3, pp. 528-536, Sep. 2019, doi: https://doi.org/10.17775/CSEEJPES.2019.01620.

[9] Y. Liu, Z. Li, W. Wei, J. Zheng, and H. Zhang, "Data-driven dispatchable regions with potentially active boundaries for renewable power generation: concept and construction," *IEEE Transactions on Sustainable Energy,* vol. 13, no. 2, pp. 882-891, Apr. 2021, doi: https://doi.org/10.1109/TSTE.2021.3138125.

[10] H. Ji, C. Wang, P. Li, G. Song, H. Yu, and J. Wu, "Quantified analysis method for operational flexibility of active distribution networks with high penetration of distributed generators," *Applied Energy,* vol. 239, pp. 706-

714, Apr. 2019, doi: https://doi.org/10.1016/j.apenergy.2019.02.008.
[11] J. Xiao, G. Zu, Y. Wang, X. Zhang, and X. Jiang, “Model and observation of dispatchable region for flexible distribution network,” *Applied Energy,* vol. 261, pp. 114425, Mar. 2020, doi: https://doi.org/10.1016/j.apenergy.2019.114425.
[12] H. Huang, M. Zhou, S. Zhang, L. Zhang, G. Li, and Y. Sun, “Exploiting the operational flexibility of wind integrated hybrid AC/DC power systems,” *IEEE Transactions on Power Systems,* vol. 36, no. 1, pp. 818-826, Jan. 2021, doi: https://doi.org/10.1109/TPWRS.2020.3014906.
[13] W. Dai, Z. Yang, J. Yu, K. Zhao, S. Wen, W. Lin, and W. Li, “Security region of renewable energy integration: Characterization and flexibility,” *Energy,* vol. 187, pp. 115975, Nov. 2019, doi: https://doi.org/10.1016/j.energy.2019.115975.
[14] K. Wu, L. Wang, H. Ha, and Z. Wang, “Dynamic line rating and optimal transmission switching for maximizing renewable energy sources injection with voltage stability constraint,” *Applied Energy,* vol. 378, pp. 124651, Jan. 2025, doi: https://doi.org/10.1016/j.apenergy.2024.124651.
[15] K. Wu, L. Wang, H. Ha, and M. Zhang, "Maximizing Wind Power Injection by Transmission Switching and DLR with Voltage Stability Constraints." Dec. 2024, pp. 659-664, doi: https://doi.org/10.1109/ICPSAsia61913.2024.10761389.
[16] E. B. Fisher, R. P. O'Neill, and M. C. Ferris, “Optimal transmission switching,” *IEEE transactions on power systems,* vol. 23, no. 3, pp. 1346-1355, Aug. 2008, doi: https://doi.org/10.1109/TPWRS.2008.922256.
[17] M. Numan, M. F. Abbas, M. Yousif, S. S. Ghoneim, A. Mohammad, and A. Noorwali, “The role of optimal transmission switching in enhancing grid flexibility: A review,” *IEEE Access,* vol. 11, pp. 32437-32463, Mar. 2023, doi: https://doi.org/10.1109/ACCESS.2023.3261459.
[18] E. Fernandez, I. Albizu, M. Bedialauneta, A. Mazon, and P. T. Leite, “Review of dynamic line rating systems for wind power integration,” *Renewable and Sustainable Energy Reviews,* vol. 53, pp. 80-92, Jan. 2016, doi: https://doi.org/10.1016/j.rser.2015.07.149.
[19] "IEEE Standard for Calculating the Current-Temperature Relationship of Bare Overhead Conductors", IEEE Std 738-2023 (Revision of IEEE Std 738-2012), 1-56, 2023, doi: https://doi.org/10.1109/IEEESTD.2023.10382442.
[20] "Guide for thermal rating calculations of overhead lines", T. B. CIGRE, CIGRÉ, Paris, Dec. 2014, doi:
[21] D. Douglass, W. Chisholm, G. Davidson, I. Grant, K. Lindsey, M. Lancaster, D. Lawry, T. McCarthy, C. Nascimento, and M. Pasha, “Real-time overhead transmission-line monitoring for dynamic rating,” *IEEE Transactions on Power Delivery,* vol. 31, no. 3, pp. 921-927, Jun. 2014, doi: https://doi.org/10.1109/TPWRD.2014.2383915.
[22] M. M. Esfahani, and G. R. Yousefi, “Real time congestion management in power systems considering quasi-dynamic thermal rating and congestion clearing time,” *IEEE Transactions on Industrial Informatics,* vol. 12, no. 2, pp. 745-754, Apr. 2016, doi: https://doi.org/10.1109/TII.2016.2530402.
[23] J. Teh, and C.-M. Lai, “Reliability impacts of the dynamic thermal rating system on smart grids considering wireless communications,” *IEEE Access,* vol. 7, pp. 41625-41635, Mar. 2019, doi: https://doi.org/10.1109/ACCESS.2019.2907980.
[24] M. Numan, D. Feng, F. Abbas, U. Rahman, and W. A. Wattoo, “Impact assessment of a co-optimized dynamic line rating and transmission switching topology on network expansion planning,” *International Transactions on Electrical Energy Systems,* vol. 30, no. 8, pp. e12457, Jun. 2020, doi: https://doi.org/10.1002/2050-7038.12457.
[25] A. Bagheri, “Optimal transformer loss of life management in day-ahead scheduling of wind-rich power systems at the presence of dynamic ratings,” *Electric Power Systems Research,* vol. 231, pp. 110259, Jun. 2024, doi: https://doi.org/10.1016/j.epsr.2024.110259.
[26] N. H. Abas, M. Z. A. Ab Kadir, N. Azis, J. Jasni, N. F. Ab Aziz, and Z. M. Khurshid, “Optimizing grid with dynamic line rating of conductors: A comprehensive review,” *IEEE access,* vol. 12, pp. 9738-9756, Jan. 2024, doi: https://doi.org/10.1109/ACCESS.2024.3352595.
[27] W. Wei, F. Liu, and S. Mei, “Real-time dispatchability of bulk power systems with volatile renewable generations,” *IEEE Transactions on Sustainable Energy,* vol. 6, no. 3, pp. 738-747, Jul. 2015, doi: https://doi.org/10.1109/TSTE.2015.2413903.
[28] K. W. Hedman, R. P. O'Neill, E. B. Fisher, and S. S. Oren, “Optimal transmission switching—sensitivity analysis and extensions,” *IEEE Transactions on Power Systems,* vol. 23, no. 3, pp. 1469-1479, Aug. 2008, doi: https://doi.org/10.1109/TPWRS.2008.926411.
[29] J. D. Fuller, R. Ramasra, and A. Cha, “Fast heuristics for transmission-line switching,” *IEEE Transactions on Power Systems,* vol. 27, no. 3, pp. 1377-1386, Aug. 2012, doi: https://doi.org/10.1109/TPWRS.2012.2186155.
[30] "Increasing Capacity of Overhead Transmission Lines–Needs and Solutions", T. B. CIGRE, CIGRÉ, Paris, 2010, doi:
[31] R. D. Zimmerman, C. E. Murillo-Sánchez, and R. J. Thomas, “MATPOWER: Steady-state operations, planning, and analysis tools for power systems research and education,” *IEEE Transactions on power systems,* vol. 26, no. 1, pp. 12-19, Feb. 2010, doi: https://doi.org/10.1109/TPWRS.2010.2051168.

Appendix A

Determination of the Dominant two-Dimensional Subspace

This appendix formalizes the criterion for assigning a multi-dimensional ODP to its dominant two-dimensional subspace within the Radar Scanning algorithm framework. The dominant subspace for a given ODP is defined as the plane spanned by the two coordinate axes corresponding to its components with the largest absolute magnitudes.

Formal Description: Let an ODP be represented by an n-dimensional vector W= $(\omega_1, \omega_2, \ldots, \omega_n)$. The algorithm identifies the indices $\{i, j\}$ of the two components with the largest magnitude values, such that $|\omega_i| \geq |\omega_j| \geq |\omega_k|$ for all $k \notin \{i, j\}$. The dominant subspace for this ODP is then defined by the orthogonal axes ω_i and ω_j. The ODP is subsequently projected onto this (ω_i, ω_j) plane for all subsequent analysis and clustering operations within the Radar Scanning method.

Illustrative Example: Consider a system analyzed in a 3-dimensional parameter space with axes ($\omega_1, \omega_2, \omega_3$). For an ODP W= (2.5, 2.7, 2.2), the magnitude values are $|\omega_1| = 2.5$, $|\omega_2| = 2.7$ and $|\omega_3| = 2.2$. The ranking of magnitudes yields $|\omega_2| > |\omega_1| > |\omega_3|$. Consequently, the dominant subspace for this point is defined by the axes ω_2 and ω_1. All geometric and clustering calculations for this ODP within the Radar Scanning algorithm are performed within this (ω_1, ω_2) plane. This deterministic assignment is performed for every ODP in the dataset. The collective result is an efficient partition of the high-dimensional data across relevant 2D subspaces, which is fundamental to the computational efficiency of the proposed algorithm. A complete numerical demonstration for a sample dataset is provided in Table A-1.

Table A-1: Example of Determining the Dominant 2D Subspace for ODPs.

ODP	First output amplitude value (ω_1 axis)	Second output amplitude value (ω_2 axis)	thirst output amplitude value (ω_3 axis)	the dominant 2-dimensional space
1	2.2	2.7	2.5	$(, \omega_3\omega_2)$
2	5	9	10	(ω_2, ω_3)
3	-5	0	3	(ω_1, ω_3)
4	4	8	9	(ω_2, ω_3)
5	4	4	2	$(, \omega_2\omega_1)$

Once each ODP is assigned to its dominant 2D subspace (e.g., as detailed in Table A-1), the algorithm performs all data clustering and flexibility region determination within these subspaces. This approach reduces the problem's dimensionality, leading to a significant decrease in computational load and processing time while maintaining the solution precision.

Ali Beiranvand received the B.S. and M.S. degrees in electrical engineering from Shahid Chamran University of Ahvaz, Ahvaz, Iran, in 2009 and 2012, respectively. He is currently pursuing the Ph.D. degree in Power engineering at Lorestan University, KhoramAbad, Iran. His research interests include power system optimization, integration of renewable energy resources, Dynamic Line Rating systems and design of overhead transmission lines.

MahmoodReza Shakarami was born in KhorramAbad, Iran in 1972. He received his M.Sc. and Ph.D. degree in Electrical Engineering from the Iran University of Science and Technology, Tehran, Iran, in 2000 and 2010, respectively. He is currently an Associate Professor in Electrical Engineering Department of Lorestan University, KhorramAbad, Iran. His current research interests include power system dynamics and stability, FACTS devices, and distribution systems.

Meysam Doostizadeh received the MSc and Ph.D. degrees in Electrical Engineering from the University of Tehran, Tehran, Iran, in 2012 and 2016, respectively. He is currently an Associate Professor with the Faculty of Engineering, Lorestan University, KhorramAbad, Iran. His research interests include power system optimization, electricity markets, smart grid technologies, and integration of renewable energy into power systems.

IECO

This page intentionally left blank.

International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Comparative Study of AI Models for Multi-Level Optimization of External Lightning Protection Systems in Photovoltaic Stations

Hamid Reza Sezavar

Department of Electrical and Computer Engineering, Qom University of Technology, Qom, Iran.
Corresponding author's email: sezavar@qut.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 15-July-2025
Received in revised form: 30-August-2025
Accepted: 20-September-2025
Published online: 23-Sep-2026

Keywords:
ELPS,
Grounding electrode,
Lightning protection,
PV station,
Multi-stage AI.

ABSTRACT

This paper presents a comparative study on the application of artificial intelligence (AI) for optimizing External Lightning Protection Systems (ELPS) in photovoltaic power (PV) plants. The research addresses the critical need for advanced protection systems in solar installations, which are particularly vulnerable to lightning strikes due to their expansive outdoor configurations. Through a detailed comparative analysis, the study evaluates multiple AI approaches, including metaheuristic algorithms and machine learning models. The investigation reveals that metaheuristic algorithms often have lower accuracy compared to modern AI techniques. All comparisons are based on a multi-level optimization framework, systematically addressing air termination design, grounding system configuration, and overall system integration. The results show superiority in sensitivity analysis in the transformer model. Compared to other models, the random forest (RF) model, along with the artificial neural network (ANN) model, has a higher speed in data analysis. However, physics-informed neural networks (PINN) achieve remarkable improvements, delivering 93% protection coverage with only 3.2% grounding error while significantly reducing design convergence times.

NOMENCLATURE

A_p	Protection area provided by termination rods	r	Radius of the rolling sphere
A_{pv}	Total area of the PV power plant	r_p	Protection radius of an air termination pole
d_e	Diameter of a ground electrode	R_e	Resistance of a single ground electrode
D	Distance between the electrodes	R_{e_0}	Ground resistance while the soil moisture is zero
h_p	Height of an air termination pole	R_t	Total resistance of the entire grounding system
I_{max}	Maximum lightning current	T_d	Number of thunderstorm days per year
k	Grounding electrode spacing factor	V_{max}	Maximum lightning voltage
L_e	Length of a ground electrode	α	Protection angle
N_e	Number of ground electrodes	ρ	Soil resistivity
N_p	Number of air termination poles	ψ	Protection Level
m	Moisture percentage	σ	Empirical constant depending on the soil conditions

I. Introduction

A. Motivation

Lightning occurs when the imbalance in electrical charges between clouds or between clouds and the ground becomes too great, leading to a sudden discharge [1, 2]. The basic principle of lightning protection is to create controlled paths that can safely conduct the resulting high current to the ground. While arresters effectively control lightning currents in power systems [3], installations at risk require additional protection. This is why comprehensive external lightning protection systems (ELPS) are implemented. These types of protection systems can provide complete protection in addition to arresters [4, 5]. An adequately designed ELPS consists of three critical components that work together. Aerial terminals prevent lightning strikes, down conductors conduct the current downward, and an extensive grounding system safely dissipates the energy to the ground [6]. This rapid deflection serves two critical purposes: it prevents equipment damage from surges while significantly reducing the risk of damaging electromagnetic induction in adjacent systems [7]. While these protection systems are essential for electrical infrastructure, they are equally critical for any structure vulnerable to lightning, including commercial buildings, industrial facilities, and especially solar power installations [8]. Solar farms present unique challenges in lightning protection due to their outdoor locations. Their vastness makes them prime targets for lightning. When lightning strikes an unprotected solar array, it affects all PV elements. PV panels can be destroyed, inverters can be damaged, and connection systems can melt. For these reasons, a thorough lightning risk assessment and properly engineered protection systems are not only recommended for solar installations but are critical for financial and operational sustainability [9].

B. Literature review

Recent advances in ELPS modeling have shifted from traditional approaches to advanced AI-based techniques. Much research has been conducted on computational methods for optimizing lightning protection systems. For example, in [10, 11] used genetic algorithms (GA) to model ground electrode configurations, while another study [12] used random forest classification [13] to analyze weather patterns for lightning prediction. Simulation-based approaches have also been valuable, and mathematical algorithms have been applied to design efficient grounding systems [14]. This field has seen interesting advances in optimization algorithms. Studies [15-19] have compared particle swarm optimization (PSO), genetic algorithms, and hybrid HGA-PSO methods for grounding system design. Recently, neural network approaches have emerged as promising alternatives. In [20] introduced a multi-stage neural network model that proposed innovative ELPS configurations – using fewer but taller poles and shorter ground electrodes compared to traditional ATP/EMTP methods. Beyond electrical performance, contemporary research has begun to examine practical installation factors.

The effects of protection poles' shading on solar panels have become an important design consideration. A case study [21] of an 8 MW solar power plant demonstrated a successful balance of concerns related to shading, tower height, and current discharge capacity. Comprehensive analyses [22] have also confirmed that passive ELPS isolated with galvanized grounding provides optimal protection for photovoltaic installations. Comparison of algorithms has yielded valuable insights. When the NSGA and MOPSO methods [23] were tested for distribution system protection, the results showed that NSGA exhibited superior performance in both grounding optimization and overall system reliability enhancement. Parallel research [24, 25] specifically focused on minimizing the impact of lightning rod shading while maintaining protection effectiveness—a critical consideration for solar farm layouts.

C. Contributions

According to the literature review, ELPS design studies have moved towards the use of artificial intelligence (AI) models. Where traditional methods rely on empirical calculations and manual iterations, modern research increasingly employs both meta-heuristic optimization techniques and machine learning models to solve this complex engineering challenge. While meta-heuristic methods have dominated early applications, machine learning approaches are now demonstrating remarkable potential when supplied with comprehensive training datasets that capture real-world variability. Perhaps less than machine learning models are used for optimization, but with appropriate training data, a suitable system can be designed to optimize and estimate ELPS parameters. Therefore, considering the gap in optimization studies based on artificial intelligence, this article has reviewed and compared artificial intelligence models that can be used for optimization in ELPS design. It is worth mentioning that multi-level optimization models are used to estimate parameters in a step-by-step manner. This comparative investigation evaluates artificial intelligence models across the full spectrum of design conditions encountered in practice. The analysis considers how different models respond to variations in local weather patterns, including thunderstorm frequency, site-specific photovoltaic array characteristics, soil conductivity properties, and international protection standards. As a result, the best model can be selected for design considering all conditions.

D. Paper organization

The rest of this paper is structured as follows: Section 2 details the External Lightning Protection System that would cover air termination parameters and grounding system design. Section 3 introduces the proposed multi-level optimization framework in ELPS. Section 4 presents the results and comparative analysis, and discussion. Finally, Section 5 summarizes the conclusions and discusses implications for future research.

II. External Lightning Protection System

ELPS provides power system protection at the desired level in an outdoor environment. ELPS consists of two parts: air terminal and ground connection. The parameters of the air terminal include the height of the poles (*hp*) and the number (*Np*). The coverage area is estimated based on the pole protection radius (*Rp*). In addition to these parameters, the parameters of the ground system include the number of electrodes (*Ne*), electrode length (*Le*), and electrode diameter (*de*). Electrical indices, including the resistance value of each electrode (*Re*) and the total resistance (*Rt*), are calculated using these parameters. Considering the importance of using ELPS in outdoor solar power plants, it is valuable to design the optimal parameters according to the constraints of a PV power plant.

A. Air Termination Parameters

The first component of an ELPS system that is usually struck by lightning is the aerial terminal. The installation height of this terminal (*hp*) plays a decisive role in the efficiency of the protection system, as it directly affects the size of the protection radius. Based on the accepted principle that lightning strikes the highest point in an area, as the installation height increases, the radius of the area covered by the protection increases, while the corresponding protection angle decreases. This relationship is shown in Equation (1)[7]:

$$r_p = h_p \times tan\alpha \tag{1}$$

In this regard, α represents the protection angle. As can be seen, although increasing the height increases the area covered by the protection, it also decreases the protection angle. For this reason, instead of increasing the height, a more efficient option is to increase the number of protection rods. Proper distribution and dispersion of the protection rods provides a better solution than increasing the height. Other factors also play a vital role in the design of this system, including the required protection level and the frequency of lightning days in the area. The average number of days with lightning in an area can be an effective indicator for determining the optimal number of protection rods required to achieve the desired protection level. An increase in this average means an increase in the probability of sudden lightning strikes, and as a result, increases the necessity of deploying the ELPS system in various areas. To calculate the radius of protection (*rp*) and measure the area covered, the standard "rolling sphere" method is used. The geometric performance of this method is illustrated in Figure 1. In accordance with the IEC standard and using the radius of the rolling sphere (*r*) whose values are given in Table I, the value of *rp* can be obtained through the following equation: [26, 27]:

$$r_p = \sqrt{h_p(2r - h_p)} \tag{2}$$

therefore, the protection area (A_p) can be calculated as follows:

$$A_p = \pi {r_p}^2 \tag{3}$$

B. Grounding Parameters

The grounding system, as a vital component of the ELPS, plays a crucial role in transferring the fault current to the ground. The resistance value of the grounding electrode depends on factors such as the length of the rod and the distance between the electrodes. This resistance can be calculated using the following equation [28]:

$$R_e = \frac{\rho}{2\pi L_e}\left(Ln\left(\frac{4L_e}{d_e}\right) - 1\right) \tag{4}$$

where ρ is the specific resistance of the soil. Considering the importance of the number of electrodes, the total resistance is calculated with the number of electrodes as follows [29]:

$$R_t = \frac{R_e}{N_e}\left(1 + \frac{N_e - 1}{2}k\right) \tag{5}$$

where k is the space factor and can be calculated according to the following equation [28-30]:

$$k = \frac{1}{\ln\left(\frac{D}{d_e}\right)} \tag{6}$$

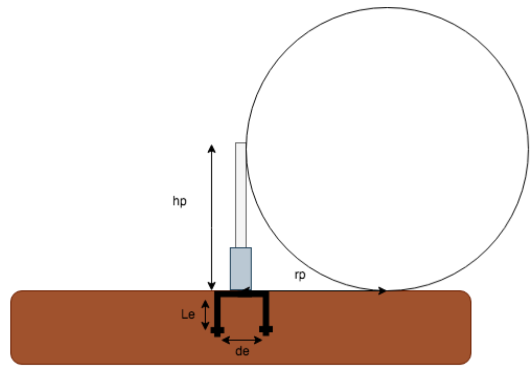


Fig. 1. Schematic view of rolling sphere method geometry.

TABLE I. IEC standard lightning protection level classification.

Protection levels	**radius of rolling sphere (m)**	**Description**
1	20	Highly critical structures
2	30	Important infrastructure
3	45	Standard commercial structures
4	60	Low-risk structures

where D is the distance between the electrodes. The parameters of the number of electrodes, their length, diameter, and distance between them play a significant role in eliminating the fault caused by lightning. The specific resistance of the soil based on [31] can have different values according to the humidity and climate of the region. However, from an electrical perspective, the maximum current during a fault should be reasonably correlated with the resistance value of the electrodes. Optimal resistance values ensure the effective conduction of electrical current in the soil environment, allowing for safe fault clearance. According to [32], the maximum current that the ELPS is able to protect, can be approximated through the following:

$$I_{max} = \sqrt[0.75]{r} \tag{7}$$

Based on the protection level, the value I_{max} is calculated. According to [33], based on the maximum lightning voltage in an area, the grounding resistance relationship can be expanded as follows:

$$R_e = R_{e_0} e^{-\sigma m} \tag{8}$$

where R_{e_0} is earth resistance while the soil moisture is zero, m is the moisture percentage, and σ is an empirical constant depending on the soil conditions. In order to simplify the calculation of the electrode resistance, the following equation can be proposed to estimate it based on electrical parameters:

$$R_e = \frac{V_{max}}{I_{max}} \times e^{-\sigma m} \tag{9}$$

Therefore, considering equations (7) and (9) and also the relationship with them, equations (4) and (5), the number of electrodes (N_e), their length (L_e), and their diameter (d_e) can be calculated based on the maximum protective current [32, 34, 35].

III. Multi-Level Optimization in the ELPS

According to the different parts of ELPS that are explained in the previous section, its design includes several interconnected parameters. In the first stage, the goal is to design the optimal air termination parameters. *Np* and *hp* are calculated according to the input data based on Equations (1) and (2). The input data include the PV area and PV parameters, the protected area, and the number of thunderstorm days. According to [20], there is a relationship between *Td* and the optimal number of *Np*. This relationship is obtained according to the training data of the first stage neural network by the artificial intelligence model and is considered in the optimal design. Finally, the defined input variables are considered to produce the *Np* and *hp* outputs, and the following equation is defined:

$$\begin{aligned} &N_p, h_p = f_{AI1}(T_d, A_p, \Psi) \\ &subject\ to: r_p = \begin{cases} hp \cdot tan\alpha \rightarrow (Eq.1) \\ \sqrt{hp(2r - hp)} \rightarrow (Eq.2) \end{cases} \end{aligned} \tag{10}$$

where Ψ denotes the protection level (IEC standard), and *r* is the rolling sphere radius from Table I. In Equation (10), the model *fAI1* predicts the optimal number of air termination poles (*Np*) and their heights (*hp*) using inputs such as thunderstorm days (*Td*), protected PV area (*Ap*), soil resistivity (ρ), and the protection level (Ψ). The protection radius (*rp*) is derived via either the empirical tangent method or the rolling sphere method (IEC 62305), ensuring compatibility with site-specific requirements. This replaces traditional heuristic approaches with a data-driven model trained on historical lightning strike data and PV plant configurations. The objective of the first optimization level is to determine the optimal configuration of the air termination system. The primary goal is to minimize both the number and height of protection poles. The design variables for this level are formally defined as the vector $X_1 = [Np, hp]$. This optimization is subject to several critical constraints. First, the total area of the photovoltaic plant (*Apv*) must be completely contained within the total protected area, governed by the inequality $Apv\sim \leq Np \times \pi \times rp^2$. Second, the protection radius for each pole must be derived strictly in accordance with the IEC 62305 standard, based on the selected protection level (Ψ) and its corresponding rolling sphere radius, as specified in Table I. The governing equation for the protection radius, Equation 2, ensures the design adheres to the standardized rolling sphere method. The outputs of the first stage are considered as the input of the second stage. Also, according to the calculation of *hp* and *Np* and Table I, the ground resistance can be calculated using Equations (7) to (9). This resistance can be measured based on the outputs of the first stage, including *hp* and *Np*, using Table I to calculate *Re* and finally using the maximum lightning voltage and current. In this study, the maximum lightning voltage is assumed to be constant. The following formula is also designed for optimization in the second stage to calculate *Re*:

$$\begin{aligned} &Re = f_{AI2}(h_p, N_p, V_{max}, I_{max}) \\ &with: I_{max} = {}^{1/4}\sqrt{r} (from\ Eq.\ 7) \end{aligned} \tag{11}$$

The second level focuses on defining the performance requirement for the grounding system. The key variable for this stage is $X_2 = [Re]$. It must be sufficiently low to ensure that the voltage rise during a maximum fault current does not exceed the system's insulation withstand capability. This fault current level (*Imax*) is naturally linked to the protection level. It is calculated via the governing Equation 7. The target resistance is therefore not a fixed value but a function of the outputs from first stage and the chosen protection level.

In the third stage, the output of the second stage is considered as an input. Considering the soil resistivity (ρ), the desired total resistance (*Rt*), and the electrode spacing coefficient (*k*), the number of electrodes (*Ne*) and the length of the electrodes (*Le*) can be calculated based on equations 4 to 6. In this case, the optimization equation of the third stage is defined as follows:

$$\begin{aligned} &N_e, L_e, D = f_{AI3}(R_e, \rho, k), \\ &constrained: R_t = \left(\frac{N_e}{R_e}\right)\left(1 + \frac{(N_e - 1)}{(2ln(D/d_e))}\right) \end{aligned} \tag{12}$$

In the third stage, according to the output of the second stage, equation 4 is formed. This is the first optimization constraint in this stage. Then, equation 5 is developed according to the total resistance requirement, and the second constraint is also established. This constraint, along with Equation 6, is actually the basic constraint of this stage. Therefore, using this constraint, the optimization model can be run, and the best values for RE and LE can be calculated.

Table II is a representation of some of the data used to train the neural network used in this study. The data in Table II is a hybrid dataset, combining simulated parameters in ATP/EMTP with relationships derived from the cited literature and standards. This data has been collected based on [35, 36] studies. The machine learning optimization process is a multi-level approach that separates the design of all parameters into three sections. First, various AI models are trained on an extensive historical dataset that matches site conditions—like the number of local storm days, the size of the solar farm, and soil type—to their optimal protection system parameters. Once trained, this AI assistant is used in a practical, three-step design process. In the first step, it takes the new site's conditions and instantly predicts the best number and height for the lightning poles. The results from this step are then passed to a second AI model, which calculates exactly how low the electrical resistance of the grounding system must be to handle the energy from a strike safely. Finally, a third AI model uses that resistance target to design the physical grounding system itself, specifying the number, length, and diameter of the ground rods needed.

TABLE II. Neural network training dataset specifications

Td	*Apv*	*Ap*	*ρ*	*Rs*	*Rt*	*k*
130	5369	6563	30	7.32	1.85	0.13
102	20164	21737	300	46.79	12.70	0.18
109	17984	22269	150	18.52	5.51	0.13
124	10438	13769	400	88.70	48.02	0.17
127	31533	34281	1000	139.74	75.15	0.15
114	25262	25850	100	12.48	3.17	0.13
115	14500	18000	50	9.45	2.50	0.14
108	22000	24000	200	25.30	7.80	0.16
121	8500	12000	250	65.20	32.10	0.19
132	42000	45000	800	120.50	65.30	0.14
105	18000	21000	120	15.75	4.20	0.12
118	32000	35000	600	95.40	52.60	0.17
126	12500	16000	350	72.80	40.50	0.18
111	19500	23000	180	22.10	6.90	0.15
129	38000	41000	900	130.20	70.80	0.16
107	16500	19000	90	11.20	3.05	0.13

In general, what distinguishes multi-level optimization based on artificial intelligence from traditional models is the unique ability to find optimal points concerning suitable training data. If the data is selected appropriately, this type of optimization can bridge the gap between heuristic and meta-heuristic models such as genetic, PSO, and other such models. In this study, an attempt has been made to examine various artificial intelligence models for the purpose of multi-level optimization and to present a comprehensive comparative study. Each of these models has its strengths and weaknesses in the design of ELPS in a PV system. Finally, considering the different inputs, a suitable model can be designed for the design of an ELPS suitable for installation in a PV power plant.

Figure 2 shows the proposed scheme for optimizing the design of a lightning protection system. The model breaks down a significant, complex problem into three smaller, easier-to-understand parts. The first step is to design the lightning rods. The AI uses information about the weather, the size of the solar farm, and safety standards to decide on the number of rods needed and their height. This solves the problem of protecting the area from lightning strikes. The second step uses the results of the first step to calculate the target of the grounding system. This step determines how low the electrical resistance should be to handle lightning energy safely. This step connects the design of the above-ground rods to the underground electrodes. The final step is to design the grounding system itself. Using the resistance target from the second step, the AI calculates how many ground rods are needed, how long they should be, and how wide they should be. This ensures that the lightning current can be safely absorbed into the ground.

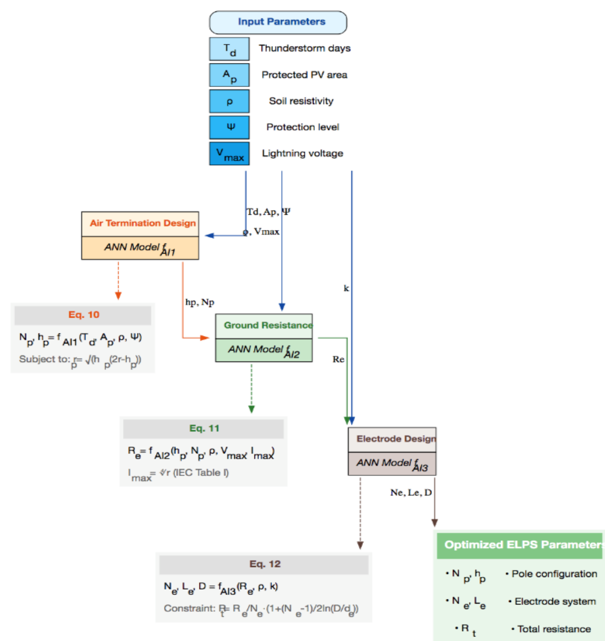


Fig. 2. Flowchart of multi-level optimization framework.

IV. Results and Discussion

In order to analyze the presented model, according to the flowchart of Figure 2, 400 test data samples based on Table II have been examined. The PINN, RF, Transformer [37, 38] , and artificial neural network (ANN) [39, 40] models have been implemented for AI purposes. The GA, PSO, and HGA-PSO models have been changed to metaheuristic models [41, 42].

A. Performance Analysis

Optimization of ELPS for PV power plants requires a careful examination of computational efficiency and protection effectiveness. A comparative analysis between traditional metaheuristic approaches and modern AI-based methods is investigated in three criteria: convergence time, protection score, and ground fault. The results of these investigations are shown in Table III.

In the first part, the convergence time is investigated based on the same data. Genetic algorithms (GA) show the slowest approach with 142 seconds, while particle swarm optimization (PSO) shows a 31% improvement with 98 seconds. The combined HGA-PSO method has a more suitable improvement with convergence in 76 seconds due to the integration of the effective points of both previous algorithms. However, AI models show superior efficiency, with RF providing the fastest results in only 18 seconds. The ANN model shows an average result of 22 seconds. Physics-based neural networks require a slightly longer time of 28 seconds due to their inherent physical limitations, while transformers come in at 34 seconds. This significant time reduction – up to 88% compared to GA – allows for rapid design iterations and real-time adjustments as environmental conditions change.

TABLE III. Comparative performance metrics of optimization methods

Method	*Convergence Time*	*Protection Score*	*Grounding Error*
GA	142 sec	82%	12.3%
PSO	98 sec	85%	9.7%
HGA-PSO	76 sec	88%	7.1%
ANN	**22 sec***	91%	5.9%
PINN	28 sec	**93%**	**3.2%**
Transformer	34 sec	90%	4.1%
RF	**18 sec***	89%	6.7%

The *Protection Score* is a quantitative metric that evaluates the effectiveness of the air termination system design. It is calculated as the percentage of the total photovoltaic plant area (A_{pv}) that is successfully covered by the protective zones of the lightning poles, as defined by the rolling sphere method. Protection effectiveness shows a steady improvement from traditional to modern methods. GA provides a baseline protection coverage of 82%, while PSO and HGA-PSO achieve 85% and 88%, respectively. AI methods outperform these results, with ANNs achieving 91% and PINNs providing the highest protection score of 93%. The 11% difference between GA and PINN represents a significant increase in system reliability, potentially preventing costly lightning damage to sensitive PV components. The transformer models perform slightly below PINNs at 90%, while the RF maintains a remarkable protection score of 89%.

The *Grounding Error* measures the accuracy of the grounding system design predicted by the optimization algorithm. It is defined as the relative difference between the algorithm's predicted total grounding resistance ($R_{t\ predicted}$) and the target resistance ($R_{t\ target}$) calculated as necessary for safety in Level 2. Optimizing the grounding system is particularly challenging for traditional methods. GA exhibits an error rate of 12.3% in ground resistance calculations, while PSO reduces this to 9.7%. The combined HGA-PSO approach achieves an error rate of 7.1%, demonstrating the benefits of hybrid optimization strategies. AI methods show significant improvements in this critical parameter – ANN reduces errors to 5.9%, while PINN achieves an exceptional error rate of 3.2% through its physics-constrained architecture. This 74% error reduction compared to GA translates into more reliable fault current dissipation and improved system safety. The transformer models maintain their strong performance with an error of 4.1% and the RF with 6.7%.

Figure 3 provides a representation of the performance characteristics of the different ELPS optimization methods. The scatterplot divides the solution space into two distinct regions: metaheuristic methods (GA, PSO, HGA-PSO) cluster in the lower right quadrant and show slower convergence times and moderate accuracy, while AI-based approaches (ANN, PINN, Transformer, RF) dominate the upper left quadrant and achieve faster computations and superior protection scores. Notably, the position of PINN emerges as the most balanced solution, offering near-optimal performance in both dimensions. This visualization effectively demonstrates how hybrid approaches such as HGA-PSO act as transitional methods between the two clusters. At the same time, pure AI implementations push the Pareto frontier toward the ideal starting point of instantaneous and perfect accuracy.

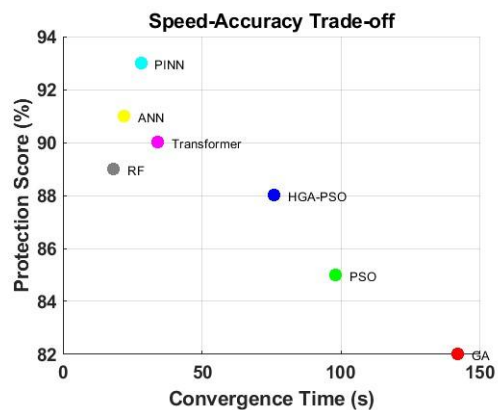


Fig. 3. Scatterplot of method performance

B. Comparative Analysis of AI

The performance benefits of AI methods are particularly evident when examining the trade-off between computational speed and accuracy. While meta-heuristic approaches require transaction speed for accuracy, AI models achieve both faster results and superior conservation metrics. PINNs emerge as the most balanced solution, combining physical fidelity with computational efficiency. Their ability to directly encode the fundamental equations in the network architecture ensures compliance with international standards while maintaining a practical design timeline.

Figure 4 shows the three-dimensional visualization critical for the performance trade-offs of different AI approaches. The Figure reveals that while RF offers the lowest computational complexity at 0.7 MFLOPs, making it suitable for edge devices and rapid prototyping, it achieves only 89%

protection with a relatively high 6.7% grounding error. At the other extreme, Transformer models demand 2.5 MFLOPs of computation but deliver only marginally better protection, about 90%, than simpler methods. The PINN method emerges as the most balanced solution, achieving the highest protection score with 93% and 1.8 MFLOPs computational requirements and a 3.2% grounding error. This performance of PINN is better than others because it can add fundamental physical equations directly into its architecture. So, it can solve solutions that comply with IEC standards based on Table I while maintaining computational efficiency. The ANN represents a middle ground, offering 91% protection at 1.2 MFLOPs, but with 5.9% grounding errors, it has a higher error than PINN. The results demonstrate that computational power (Transformers) does not guarantee optimal results, while PINN achieve better accuracy with reasonable computational demands. Therefore, for practical implementation, PINN should be the default choice. In comparison, RF remains applicable for quick assessments in resource-constrained environments.

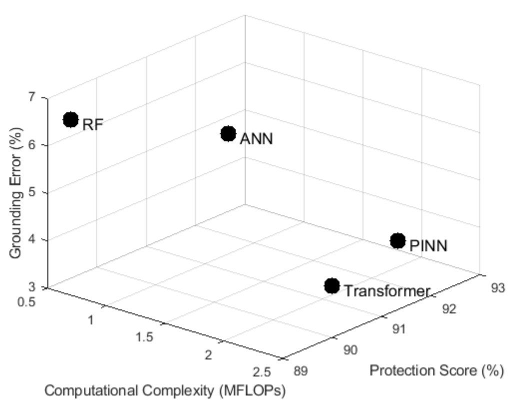


Fig. 4. 3D visualization of computational efficiency trade-off

Figure 5 illustrates that the radar chart provides a multidimensional comparison of four AI approaches for optimizing ELPS. The visualization demonstrates that PINN offer the most balanced performance profile, excelling particularly in grounding accuracy and protection coverage while maintaining reasonable computational efficiency. Its uniform shape on the sides indicates better performance than other models. In contrast, the RF model shows a sharply spiked profile, highlighting its exceptional computational speed and FLOP efficiency but significantly poorer performance in grounding system design. The Transformer model presents an interesting middle ground, with decent protection capabilities but higher computational demands, while ANN demonstrates competent all-around performance without any standout weaknesses. The radial format effectively communicates how each approach makes distinct trade-offs between physical fidelity and computational efficiency, with PINNs emerging as the most robust solution for comprehensive lightning protection system design in PV applications.

Figure 6 shows the performance of various models under the same data to estimate ELPS parameters. In this figure, the learning convergence of different AI approaches for ELPS design is presented. The logarithmic scale loss values show that PINN achieves the fastest convergence and the lowest final error. This demonstrates the effectiveness of incorporating physical constraints into the training process. While RF shows fast initial convergence, it reaches a steady state at a relatively high error level. The Transformer model shows slower initial progress but eventually outperforms the ANN in final accuracy, indicating that its attention mechanisms become increasingly effective with further training cycles. The ANN shows continuous improvement but cannot match the final accuracy of PINN, highlighting the value of physics-based regularization.

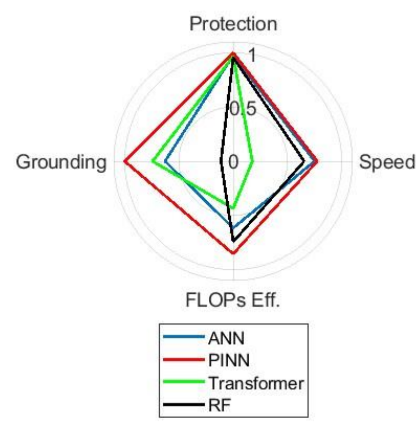


Fig. 5. Radar chart of AI model capabilities

The parameter sensitivity analysis in Figure 7 reveals fundamental differences in how each AI model prioritizes input variables for lightning protection system design. PINN demonstrates the most physically intuitive sensitivity profile, showing a strong dependence on soil resistivity with a value of 0.45. This contrasts sharply with RF, which overweight the PV area with a value of 0.35, while underweighting soil conditions with a value of 0.25, explaining its poorer grounding accuracy. The Transformer model shows balanced but somewhat diffuse sensitivity across parameters, while ANN's moderate sensitivity to protection level with a value of 0.35 reflects its data-driven approach. Notably, PINN's low sensitivity to thunder days, with a value of 0.08, suggests robust performance across varying weather conditions. Generally, the patterns show the sensitivity analysis of each model and the effect of each parameter in models. PINN's physics-aware architecture naturally emphasizes the most physically significant parameters, while data-driven methods like RF and ANN tend to overemphasize more readily measurable but less critical inputs. This is while the transformer model is less sensitive to data and can be considered a good advantage of this model. Figure 8 shows the failure probability comparison. Based on the results, PINN has high reliability for lightning protection systems, with its 5.41% failure rate representing an improvement over Transformer and ANN models. The Transformer's intermediate 9.95% failure rate suggests its attention mechanisms provide some robustness benefits over conventional ANN. Furthermore, PINNs have a remarkable reduction compared to RF. These field validation results directly reflect each method's underlying architecture. In fact, PINN's physics-constrained design prevents unphysical solutions that could lead to protection failures, while RF's data-driven approach proves particularly vulnerable to edge cases not represented in training data.

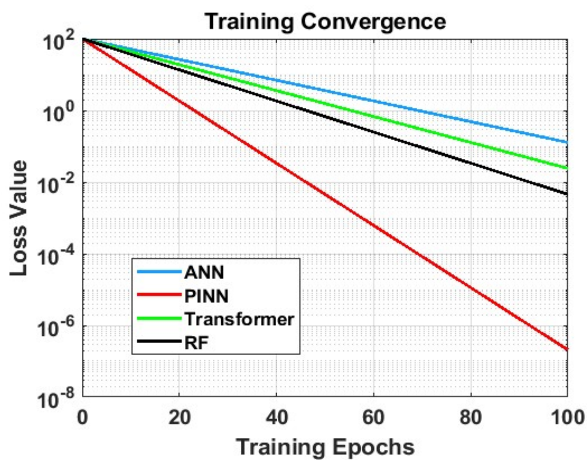


Fig. 6. Learning convergence curves results

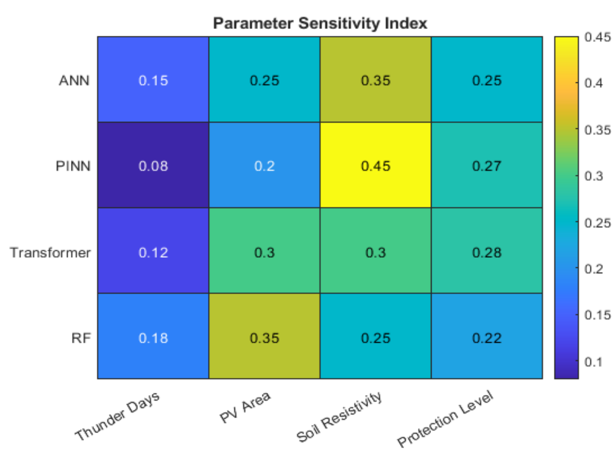


Fig. 7. Parameter sensitivity analysis of AI models

To strengthen the validation, new data are generated through simulation based on Table 2. A comparative study is conducted using the new data, and the results are analyzed and compared. The results are shown in Figures 9-11. Figure 9(a) shows the lightning protection coverage of each AI model. The PINN model achieves the highest protection score of 93%. The 2-4% difference between PINN and the other models indicates an improvement in system reliability in this AI model. ANN is the closest model in terms of protection capability with 91%, while Transformer and RF show lower

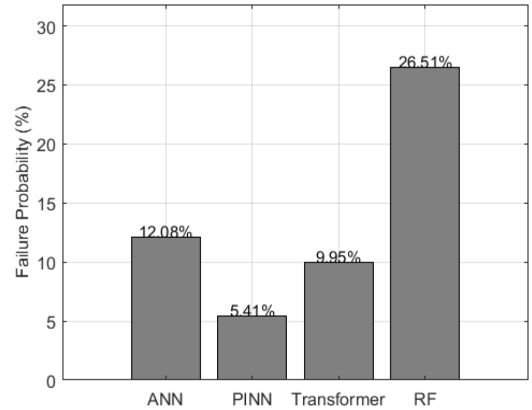


Fig. 8. Field failure probability comparison results

protection effectiveness with 90% and 89%, respectively. Figure 9(b) shows the substantial differences in the accuracy of the ground system. In this case, PINN has a significant error rate of 3.2%. The Transformer model shows an acceptable performance with an error rate of 4.1%. In contrast, ANN and RF show higher error rates of 5.9% and 6.7%, respectively, indicating limitations in the design of the ground system. These results confirm the results of the comparative study conducted in the previous section.

Figure 10 presents both the protection score and grounding accuracy metrics. The results clearly demonstrate the dual superiority of PINN. The visual comparison emphasizes that the physics-based approach of PINN successfully bridges the gap between theoretical protection requirements and practical implementation constraints. These results also confirm previous results using new data. Figure 11 introduces a performance ratio metric (protection score divided by grounding fault) that quantifies the overall effectiveness of the system. PINN achieves the highest ratio of approximately 29.1, significantly outperforming other models. This metric effectively captures the essential balance between protection coverage and implementation accuracy, providing a single comprehensive measure of design quality.

Statistical analysis performed on several random samples conclusively shows that PINN achieves a strong and consistent performance advantage over other AI models. The results of this evaluation are shown in Table 4. The PINN model consistently achieved the highest average protection score and the lowest average ground fault across all 10 seeds. While models such as the standard ANN and Transformer showed greater sensitivity to initial conditions, the RF model also showed very low variance (±0.5% and ±0.4%) due to its ensemble nature, where the average is calculated over many decision trees. However, its average performance was lower than PINN. PINN consistently produced the highest protection score and the lowest ground fault with minimal deviation.

C. Recommended Implementation

Based on the comprehensive performance analysis of various optimization methods for ELPS in PV installations, the PINN model is offered as the most robust solution. It should perform as the primary design approach for most applications. PINN demonstrates good performance across all critical metrics, achieving a 93% protection score with only 3.2% grounding error while maintaining reasonable computational efficiency. Its physics-constrained architecture provides unmatched reliability and has a lower failure rate compared to conventional ANN approaches and RF methods. This performance advantage stems from PINN's unique ability to use fundamental physical equations directly into the model. Using equations alongside training data significantly helps to obtain better results but may increase sensitivity and reduce performance speed, as shown in the results.

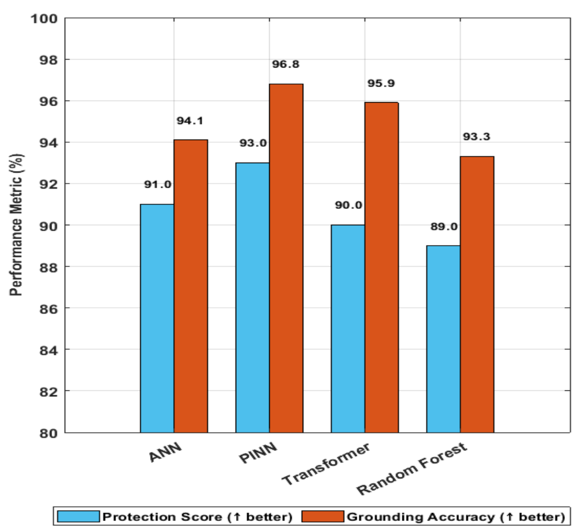


Fig. 9. Combined metrics of protection score and grounding accuracy

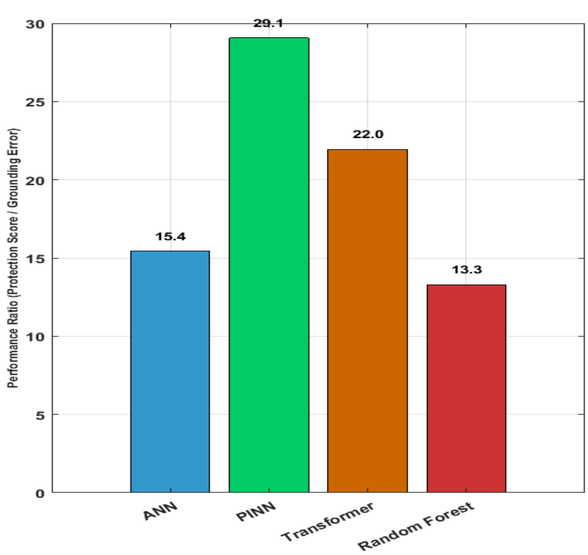
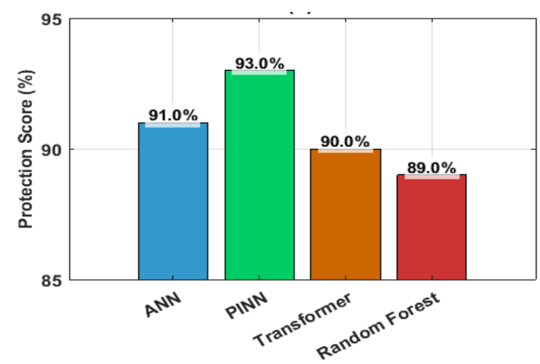


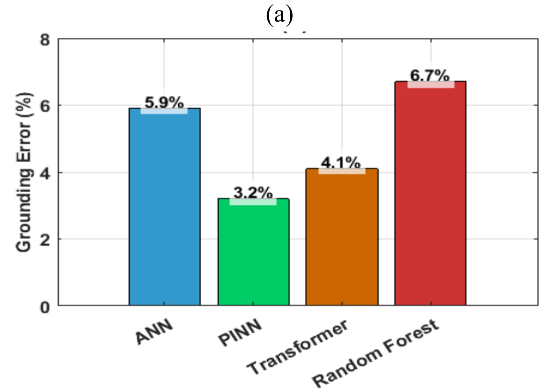
Fig. 10. Comprehensive analysis of protection system effectiveness

TABLE IV. Performance Metrics across 10 random seeds (Mean ± Standard Deviation)

Model	***Protection Score (%)***	***Grounding Error (%)***	***Convergence Time (s)***
PINN	92.7 ± 0.4	3.4 ± 0.3	29.1 ± 2.1
ANN	90.8 ± 0.7	6.1 ± 0.6	22.3 ± 1.8
Transformer	89.5 ± 1.1	4.3 ± 0.5	35.2 ± 3.0
RF	89.2 ± 0.5	6.9 ± 0.4	18.5 ± 1.2



(a)



(b)

Fig. 11. Comparative validation of AI models for ELPS optimization based on new data. (a) Protection coverage score (b) Grounding system design error

Despite its demonstrated advantages, the proposed PINN approach has limitations. Its training phase is computationally more intensive than that of standard ANNs or RF due to the evaluation of physical equations. Furthermore, its performance is inherently tied to the accuracy of the underlying physical model; incorrect or oversimplified governing equations would lead to biased predictions. While PINNs reduce the demand for vast empirical datasets, they still require high-quality data to calibrate the physics to real-world conditions. Finally, the implementation complexity of PINNs is higher, requiring expertise in both machine learning and domain-specific physics. For large-scale PV installations, PINN should be implemented throughout the entire design workflow. It means from initial pole configuration using their multi-stage optimization framework to final grounding system design with their dual-path neural network approach. The RF method remains useful for rapid prototyping due to its computational efficiency, and Transformers offer value for complex spatial optimization in mega-scale farms. PINN provide the optimal balance of accuracy, reliability, and practical implementation requirements in PV protection. The method's consistent performance across varying contamination levels and weather conditions makes it particularly valuable for real-world deployment where environmental factors fluctuate significantly. However, the ANN is an option for smaller-scale PV. The ELPS design is not complex, and the pole numbers and their location are constraints.

V. Conclusions

This study presents a comparison of several AI-based optimization methods for designing lightning protection systems. Physics-based neural networks emerge as the most effective solution, outperforming traditional optimization methods and conventional machine learning approaches. Field validation results confirm the superior reliability of physics-based AI models, with a significantly lower failure rate compared to alternative approaches. For practical implementation, the findings indicate that large-scale photovoltaic installations benefit most from comprehensive PINN-based design workflows, while smaller systems may benefit from simpler ANN configurations. The behaviour of the RF model is also very similar to that of the ANN model. However, a large scatter is observed in the transformer model analysis. PINN demonstrates superior protection effectiveness with a 93% coverage score, significantly outperforming genetic algorithms at 82% and particle swarm optimization at 85%. In terms of grounding system accuracy, PINN achieves a remarkable 3.2% error rate, representing a 74% improvement over GA's 12.3% error rate. While random forest algorithms show the fastest convergence at 18 seconds, their 6.7% grounding error and 89% protection score prove less reliable than PINN's balanced performance at 28 seconds of computation time. The multi-level optimization framework successfully addresses critical interactions between air termination design and grounding parameters that single-stage methods often neglect. Field validation data confirms PINN's practical reliability with a 5.41% failure rate, nearly half that of transformer models at 9.95%.

Future research directions include the hybrid use of the proposed models at each of the optimization levels, which could yield even better results. Furthermore, adding more constraints, including optimal location, shadow size of the poles, and other limiting factors, can help design a more accurate and practical ELPS. Looking forward, this research opens several promising future directions. First, it can be investigated using hybrid modeling, where a PINN is used for the critical grounding system design, and a faster RF model is used for the air termination layout. Second, integrating additional practical constraints, such as dynamic shadow analysis, topological limitations, and detailed economic costing, into the optimization framework is a crucial next step. Finally, exploring transfer learning techniques for PINNs could drastically reduce their computational overhead for new, unseen site conditions, making this powerful approach more accessible.

References

[1] P. Chiradeja and A. Ngaopitakkul, "Identify direct lightning strike location based on discrete wavelet transform for 115-kv transmission system," *IEEE Access,* vol. 10, pp. 80609-80622, 2022. doi: https://doi.org/10.1109/access.2022.3195497.

[2] I. Mousaviyan, S. G. Seifossadat, and M. Saniei, "A new and very fast fault detection and classification method based on traveling wave in transmission lines," *International Journal of Industrial Electronics Control and Optimization,* vol. 4, no. 4, pp. 377-385, 2021. doi: https://doi.org/10.22111/ieco.2021.38344.1348.

[3] Q. Wang, X. Liang, Y. Shen, S. Liu, Z. Zuo, and Y. Gao, "Lightning flashover characteristics of a full-scale AC 500 kV transmission tower with composite cross arms," *Engineering,* vol. 23, pp. 130-137, 2023.

[4] W. Qian *et al.*, "Mono-pole blocking accident analysis caused by inaccurate measurement of optical CT in±800 kV UHVDC transmission system under lightning strike," *Electric Power Systems Research,* vol. 239, p. 111212, 2025. doi: https://doi.org/10.1016/j.epsr.2024.111212.

[5] S. Mpanga, A. Zulu, M. Mwanza, and R. L. Holle, "Spatial and temporal variation of Zambia lightning for designing lightning protection of infrastructure," *Electric Power Systems Research,* vol. 229, p. 110188, 2024. doi: https://doi.org/10.1016/j.epsr.2024.110188.

[6] B. RAHEEM, E. Ogbuju, and F. Oladipo, "Development of a Lightning Prediction Model Using Machine Learning Algorithm: Survey," *Journal of Applied Artificial Intelligence,* vol. 4, no. 1, pp. 45–56-45–56, 2023. doi: https://doi.org/10.48185/jaai.v4i1.727.

[7] I. M. Y. Negara, D. Fahmi, D. A. Asfani, I. S. Hernanda, R. B. Pratama, and A. B. Ksatria, "Investigation and improvement of standard external lightning protection system: Industrial case study," *Energies,* vol. 14, no. 14, p. 4118, 2021. doi: https://doi.org/10.3390/en14144118.

[8] A. Ozdemir and S. Ilhan, "Experimental performance analysis of conventional and non-conventional lightning protection systems–preliminary results," *Electric Power Systems Research,* vol. 216, p. 109080, 2023. https://doi.org/10.12973/ijem.6.2.259.

[9] K. Sobolewski and E. Sobieska, "Analysis of the effectiveness of lightning and surge protection in a large solar farm," *Archives of Electrical Engineering,* vol. 71, no. 2, 2022 doi: https://doi.org/10.24425/aee.2022.140726.

[10] M. Gonçalves, E. da Costa, A. Andrade, V. Brito, G. Lira, and G. Xavier, "Grounding system models for electric current impulse," *Electric Power Systems Research,* vol. 177, p. 105981, 2019.doi: https://doi.org/10.1016/j.epsr.2019.105981.

[11] A. K. Mishra, N. Nagaoka, and A. Ametani, "A genetic algorithm approach for modeling a grounding electrode," *IEEJ Transactions on Power and Energy,* vol. 125, no. 8, pp. 816-821, 2005. doi: https://doi.org/10.1541/ieejpes.125.816.

[12] S. He *et al.*, "Intelligent prediction of 110kV insulator lightning flashover criteria based on random forest," *Electric Power Systems Research,* vol. 232, p. 110423, 2024. doi: https://doi.org/10.1016/j.epsr.2024.110423.

[13] H. R. Sezavar and S. Hasanzadeh, "Artificial Intelligence for Assessing Composite Insulator Pollution Level: A Study on Partial Discharge Characteristics," *International Journal of*

Industrial Electronics Control and Optimization, 2025. doi: https://doi.org/ 10.22111/ieco.2025.51554.1680

[14] J. M. Martínez, E. M. N. Angarita, J. R. N. Alvarez, M. H. Crespo, and P. J. F. Pertuz, "Lightning rod system: mathematical analysis using the rolling sphere method," *International Journal of Power Electronics and Drive Systems (IJPEDS),* vol. 13, no. 1, pp. 2829-2838, 2022. doi: https://doi.org/10.11591/ijpeds.v13.i1.pp237-246.

[15] M. Alves *et al.*, "An automated technique and decision support system for lightning early warning," *International Journal of Environmental Science and Technology,* vol. 22, no. 4, pp. 2289-2304, 2025 .doi: https://doi.org/10.1007/s13762-024-05693-7.

[16] R. Bao, Y. Zhang, B. J. Ma, Z. Zhang, and Z. He, "An artificial neural network for lightning prediction based on atmospheric electric field observations," *Remote Sensing,* vol. 14, no. 17, p. 4131, 2022. doi: https://doi.org/10.3390/rs14174131.

[17] M. Bhosale, P. B. Karandikar, V. Karkaria, and N. R. hKulkarni, "A novel approach for grounding resistance estimation," *Indones. J. Electr. Eng. Comput. Sci,* vol. 27, p. 583, 2022. doi: https://doi.org/10.11591/ijeecs.v27.i2.pp583-591.

[18] K. Mehranzamir, Z. Abdul-Malek, H. N. Afrouzi, S. V. Mashak, C.-l. Wooi, and R. Zarei, "Artificial neural network application in an implemented lightning locating system," *Journal of Atmospheric and Solar-Terrestrial Physics,* vol. 210,p.105437,2020.https://doi.org/10.1016/j.jastp.2020.105437.

[19] C. Wang, X. Zhang, H. Yang, J. Guo, J. Xu, and Z. Sun, "Application research of convolutional neural network and its optimization in lightning electric field waveform recognition," *Scientific Reports,* vol. 15, no. 1, p. 1883, 2025. doi: https://doi.org/10.1038/s41598-025-85473-6

[20] R. Rohana, S. Hardi, N. Nasaruddin, Y. Away, and A. Novandri, "Multi-Stage ANN Model for Optimizing the Configuration of External Lightning Protection and Grounding Systems," *Energies,* vol. 17, no. 18, p. 4673, 2024. https://doi.org/10.3390/en17184673.

[21] S. Kumar, G. Singh, and N. Ahamad, "Performance of 5 Years of ESE Lightning Protection System," in *Clean and Renewable Energy Production*, 2024, pp. 247-265. https://doi.org/10.1002/9781394174805.ch10

[22] M. Parhamfar, R. Naderi, and I. Sadeghkhani, "Risk assessment, lightning protection, and earthing system design for photovoltaic power plants: A case study of utility-scale solar farm in Iran," *Solar Energy Advances,* vol. 5, p. 100098, 2025/01/01/ 2025. https://doi.org/10.1016/j.seja.2025.100098.

[23] J.-W. Tang, V. Cooray, C.-L. Wooi, M. Z. A. Ab Kadir, W.-S. Tan, and H. N. Afrouzi, "Optimization of Lightning Protection System Using Multi-Objective Optimization Techniques," in *ICLP 2024; 37th International Conference on Lightning Protection*, 2024, pp. 620-625: VDE. doi : https://doi.org/10.1109/ICPEI61831.2024.10748622

[24] M. Nassereddine, H. Ahmed, E. Barbulescu, J. Rizk, A. Hellany, and M. Nagrial, "Direct Lightning Protection for PV Systems; A Novel Approach to Eliminate Shading Effects," in *2024 International Conference on Electrical, Computer and Energy Technologies (ICECET*, 2024, pp. 1-6: IEEE. https://doi.org/10.1109/icecet61485.2024.10698474.

[25] N. Rodrueang, S. Woothipatanapan, N. Charlangsut, P. Ngamprasert, N. Ruangsap, and N. Rugthaicharoencheep, "Risk Assessment of the Lightning Protection System for Hybrid Solar Power Generation Rooftop System on the Factory Using the FMECA Technique," in *2024 International Conference on Power, Energy and Innovations (ICPEI)*, 2024, pp. 167-172: IEEE. https://doi.org/10.1109/ICPEI61831.2024.10748622

[26] M. Nassereddine, "A novel lightning mast layout to eliminate shading effect on PV panels," *Electric Power Systems Research,* vol. 236, p. 110972, 2024. doi: https://doi.org/10.1016/j.epsr.2024.110972.

[27] Negara, I. Made Yulistya, et al. "Investigation and improvement of standard external lightning protection system: Industrial case study." Energies 14.14 (2021): 4118. doi: https://doi.org/10.3390/en14144118.

[28] V. P. Androvitsaneas, I. F. Gonos, and I. A. Stathopulos, "Research and applications of ground enhancing compounds in grounding systems," *IET Generation, Transmission & Distribution,* vol. 11, no. 13, pp. 3195-3201, 2017. doi: https://doi.org/10.1049/iet-gtd.2017.0233.

[29] M. Wahba, M. Abdel-Salam, M. Nayel, and H. A. Ziedan, "Experimental characterization of contact resistance of desert soil with waste-enhancement materials in grounding systems," *Results in Engineering,* vol. 23, p. 102707, 2024. doi: https://doi.org/10.1016/j.rineng.2024.102707.

[30] O. Kherif, S. Robson, H. Griffiths, N. Harid, D. Thorpe, and A. Haddad, "On the high frequency performance of vertical ground electrodes and LRM application," *IEEE Transactions on Electromagnetic Compatibility,* 2024. doi: https://doi.org/10.1109/temc.2024.3435788.

[31] L. Neamţ, O. Chiver, C. Barz, C. Costea, and Z. Erdei, "Considerations about power system grounding for different soil structure," in *2014 International Conference and Exposition on Electrical and Power Engineering (EPE)*, 2014, pp. 1034-1038: IEEE. doi: https://doi.org/10.1109/icepe.2014.6970066.

[32] B. Kermani and R. Shariatinasab, "Protection of Photovoltaic Systems Against Direct Lightning Strokes," *IEEE Transactions on Power Delivery,* 2024. https://doi.org/10.1109/tpwrd.2024.3494053.

[33] R. De Franco *et al.*, "Monitoring the saltwater intrusion by time lapse electrical resistivity tomography: The Chioggia test site (Venice Lagoon, Italy)," *Journal of Applied Geophysics,* vol. 69, no. 3-4, pp. 117-130, 2009. doi: https://doi.org/10.1016/j.jappgeo.2009.08.004.

[34] P. against Lightning-Part, "3: Physical damage to structures and life hazard," *International Electro-technical Commission. IEC,* pp. 62305-3, 2010. doi: https://doi.org/10.4324/9780203805305.

[35] P. K. Samaras *et al.*, "Evaluation of the electric stress on an Insulating down-conductor caused by lightning strikes through ATP-EMTP simulations," *IEEE Transactions on Industry Applications,* 2024. doi: https://doi.org/10.1109/tia.2024.3429472.

[36] A. G. Martins-Britto, C. M. Moraes, and F. V. Lopes, "Transient electromagnetic interferences between a power line and a pipeline due to a lightning discharge: An EMTP-based approach," *Electric Power Systems Research,* vol. 197, p. 107321, 2021. doi: https://doi.org/10.1016/j.epsr.2021.107321.

[37] B. Huang and J. Wang, "Applications of physics-informed neural networks in power systems-a review," *IEEE Transactions on Power Systems,* vol. 38, no. 1, pp. 572-588, 2022. doi: https://doi.org/ 10.1109/TPWRS.2022.3162473

[38] R. Labib, "Utilizing physics-informed neural networks to advance daylighting simulations in buildings," *Journal of Building Engineering,* vol. 100, p. 111726, 2025. doi: https://doi.org/10.1016/j.jobe.2024.111726.

[39] L. Zhang, Z. Zhang, S. Fang and A. S. Bretas, "An Optimization Model for Distribution Networks Lightning Protection System Design: a Reliability Indexes and Cost-based Solution," 2020 IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2020, pp. 1-5, doi: 10.1109/PESGM41954.2020.9281988.
[40] N. Fahimi, H. Sezavar, and A. Shayegani-Akmal, "Flashover prediction of polymer insulators based on dynamic modeling of pollution layer resistance using ANN," *IEEE Transactions on Dielectrics and Electrical Insulation,* vol. 30, no. 1, pp. 122-130, 2022. doi: https://doi.org/10.1109/tdei.2022.3207452.
[41] N. Fahimi, H. R. Sezavar, and A. A. S. Akmal, "Dynamic modeling of flashover of polymer insulators under polluted conditions based on HGA-PSO algorithm," *Electric Power Systems Research,* vol. 205, p. 107728, 2022. doi: https://doi.org/10.1016/j.epsr.2021.107728.
[42] M. R. Masoudi, M. Haghighi, and H. R. Sezavar, "Optimization of Energy Hub Performance Using the Improved Particle Swarm Optimization (IPSO) Algorithm for Integrated Energy Systems," in *2025 12th Iranian Conference on Renewable Energies and Distributed Generation (ICREDG)*, 2025, pp. 1-6: IEEE. doi: https://doi.org/10.1109/ICREDG66184.2025.10966138.

Hamid Reza Sezavar was born in Qom, Iran, in 1991. He received a B.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 2013 and a M.Sc. in electrical engineering from the University of Tehran, Tehran, Iran, in 2015. He then received his PhD in High Voltage from the University of Tehran, Tehran, Iran, in 2022. He is currently working toward an assistant professor position at Qom University of Technology. His principal research interests are High voltage engineering, outdoor insulators, and Electrical discharge, Lightning, and AI algorithms.

Appendix A

AI	Artificial Intelligence
ANN	Artificial Neural Network
ELPS	External Lightning Protection System
GA	Genetic Algorithm
HGA-PSO	Hybrid Genetic Algorithm - Particle Swarm Optimization
IEC	International Electrotechnical Commission
PINN	Physics-Informed Neural Network
PSO	Particle Swarm Optimization
PV	Photovoltaic
RF	Random Forest

Appendix B

1. Neural Network Architectures & Hyperparameters

All neural network models (ANN, PINN, Transformer, and RF) are implemented using TensorFlow v2.8 and Keras. The models for each of the three optimization levels shared similar base architectures but were tailored to their specific input-output dimensions.

- ANN Architecture:
  - Structure: A feedforward network with 3 hidden layers.
  - Layer Sizes: Input Layer → (128 neurons) → (64 neurons) → (32 neurons) → Output Layer.
  - Activation Functions: ReLU for all hidden layers. Linear activation for the output layer (suitable for regression).
  - Initialization: He Normal initialization.
  - Regularization: L2 regularization (λ = 0.001) was applied to the kernel weights of each layer to prevent overfitting.
- PINN Architecture:
  - Structure: The PINN shared the same base architecture as the ANN (128-64-32) to ensure a fair comparison.
  - Physics-Informed Component: The physics loss was calculated based on the governing equations (Eqs. 4, 5, 6 for grounding; Eqs. 1, 2, 3 for air termination) and added to the standard mean squared error (MSE) data loss.
  - Loss Function: Total_Loss = MSE (Data) + λ * MSE (Physics), where the physics constant λ was set to 0.5 after a sensitivity analysis. This enforced that the network's predictions adhered to physical laws.
- Transformer Architecture:
  - Structure: 2 Transformer encoder blocks.
  - Key Parameters: Embedding dimension = 64, Number of heads = 4, Feedforward dimension = 128.
  - Input Processing: The tabular data was first projected into a higher-dimensional space using a dense layer. A learnable positional encoding was added, as the order of the input parameters (e.g., Td, Apv, ρ) is not sequential.
  - Output: The [CLS] token output from the final encoder block was passed through a 3-layer MLP (64 → 32 → output) to generate the prediction.
- Random Forest (RF):
  - Number of Trees: 100 estimators.
  - Other Parameters: max_depth = None (nodes expanded until pure), min_samples_split = 2, min_samples_leaf = 1. These parameters were chosen to allow the model to fully fit

the training data without excessive constraint.

2. Training Parameters & Hyperparameter Tuning

- Common Training Parameters:
    - Optimizer: Adam.
    - Learning Rate: A starting learning rate of 0.001 was used for all models, with a reduction by a factor of 0.5 if the validation loss plateaued for 10 epochs.
    - Batch Size: 32.
    - Train/Validation/Test Split: 70%/15%/15%. The split was performed randomly but consistently across all models to ensure a fair comparison.
    - Epochs: Training was stopped early if the validation loss did not improve for 25 consecutive epochs (Early Stopping callback), preventing overfitting. The maximum number of epochs was set to 500.
- Hyperparameter Tuning Process:
    - A Bayesian Optimization process was conducted using the Hyperopt library for the neural networks (ANN, PINN).
    - Search Space:
        - Number of hidden layers: [2, 3, 4]
        - Number of units per layer: [32, 64, 128, 256]
        - Learning Rate: Log-uniform distribution between 1e-4 and 1e-2
        - L2 Regularization (λ): [1e-4, 1e-3, 1e-2]
    - The objective was to minimize the Mean Absolute Error (MAE) on the validation set.
    - The final architectures and parameters reported above represent the best configuration found after 50 trials for each model type.

IECO

This page intentionally left blank.

International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Quantum-Dot Cellular Automata-Based and High-Speed Design of New Structure for Fault-Tolerant 7-Input Majority Gate

Farzaneh Jahanshahi Javaran[1]| Somayyeh Jafarali Jassbi[1] | Hossein Khademolhosseini[1,2]| Razieh Farazkish[3]

Department of Computer Engineering, SR.C., Islamic Azad University, Tehran, Iran.[1]
Department of Computer Engineering, Bey.C., Islamic Azad University, Beyza, Iran.[2]
Department of Computer Engineering, BT.C., Islamic Azad University, Tehran, Iran.[3]
Corresponding author's email: khademolhosseini@iau.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 16-April-2025
Received in revised form: 10-August-2025
Accepted: 28-October-2025
Published online: 23-Sep-2026

Keywords:
Nano,
Basic gates,
7-input Majority Gate,
Fault-tolerant,
QCADesigner.

ABSTRACT

As the field of nanotechnology rapidly advances and the need for faster processing in smaller dimensions grows, so does the integration of Very Large-Scale Integration (VLSI) technology. These difficulties include things like large-scale area needs, high power consumption, and low operating speeds, which call for new approaches to lessen these constraints. Developed and implemented at the nano-based, Quantum-Dot Cellular Automata (QCA) technology presents itself as a viable way around these obstacles. The advent of QCA technology heralds our entry into the nano-scale realm, where the advantages of enhanced processing speeds, reduced dimensions, and minimal power consumption become manifest. This article focuses on the 7-input Majority Gate, a fundamental component in QCA technology, distinguished by its fault-tolerant characteristics. The primary objective is to present the design and simulation of this key gate within the context of QCA technology. Noteworthy among the merits of the 7-input Majority Gate is its capacity to implement logic gates with a greater number of inputs, consolidating multiple functionalities within a single gate. QCADesigner and QCAPro software have simulated the given gate, and the results demonstrate the exact and correct operation of the gate. This gate generates an output signal with a 0.25-clock-cycle delay and occupies an area of 0.03 µm² with 66 quantum cells. The simulation results demonstrate the precision of the circuit's operation. Additionally, basic fault-tolerant gates such as 4-input AND and 4-input OR using a 7-input fault-tolerant Majority Gate have been suggested to illustrate the proper operation of the new gate.

I. Introduction

Deep nanoscale technologies will bring current Complementary Metal-Oxide Semiconductor (CMOS) technology closer to its scaling limit [1]. The growing degrees of variance in each facet of a nanometer design provide certain challenges for CMOS technology at nanoscales [2]. Parallelism is a common technique used to improve the efficiency of logical systems. But new nanotechnologies need to be considered to improve the system's overall performance. One of the most exciting emerging technologies is Quantum-dot Cellular Automata (QCA), which provides a novel approach to computation and information transformation in addition to providing a solution at the nanoscale [3]. Since the Majority Gates are the fundamental building element of QCA circuits, there has been a lot of interest in the effective construction of QCA circuits using Majority Gates; nevertheless, some of these are not extensible or feasible [4]. According to several research studies, QCA can be used to construct memory and general-purpose computing circuits [5-8]. A key element in building QCA circuits is the majority, as each circuit can only be accomplished with the use of majority and inverter gates. Therefore, it is crucial to build a majority in QCA as effectively as possible [9].

On the other hand, by reducing the dimensions of the parts, the sensitivity of the circuit is increased, and quantum circuits are more vulnerable to the occurrence of defects and environmental radiation, and failures will be one of the important issues at the architectural level. It is a fact that nano-architectures must be able to tolerate defects. So far, a large number of 3-input and 5-input fault-tolerant gates have been introduced, and of course, circuits have been designed with these 3-input and 5-input fault-tolerant gates [10, 11]. But the fault-tolerant 7-input Majority Gate is less designed; therefore, the research in this particular field seems appropriate. In this article, we are going to design and implement a 7-input Majority Gate that can tolerate faults, which is considered one of the basic gates in QCA technology [12]. The construction of fault-tolerant 7-input Majority Gates in QCA is a relatively unexplored field when considering the existing state of affairs. This paper highlights the need to focus research efforts on this particular area of QCA technology, where there is still room to improve fault tolerance in critical components [13, 14].

With the growing need for low-energy consumption and ultra-dense circuits, the fault-tolerant 7-input Majority Gate proposed here has large potential to be incorporated in QCA-based large-scale systems. It can be applied to memory arrays where multi-input logic functionality is essential, arithmetic processing units in nano-ALUs, error-resilient control systems, radiation-hardened computing systems like those used in aerospace technology, medical implants, or military-grade devices.

Therefore, acknowledging its critical significance as one of the foundational gates within the QCA framework, this article aims to contribute to this emerging subject by providing a thorough design and implementation of a fault-tolerant 7-input Majority Gate. Its purposeful focus on strengthening the fault tolerance of the 7-input Majority Gate (a crucial component of QCA technology) is what makes this research novel. The study aims to close a clear gap in the current state of research by addressing the shortcomings in existing designs and suggesting a fault-tolerant version. Given the growing applications of QCA technology and the critical need for reliable and fault-tolerant components in the rapidly changing field of nanoscale computing, this project is especially relevant and timely. In addition to trying to address current issues, the proposed 7-input Majority Gate also hopes to open the door for more dependable and robust QCA circuits. This research adds to the fundamental understanding of QCA technology by clarifying the fault-tolerant mechanisms and design complexities. It also provides useful information for the creation of fault-tolerant circuits within the larger framework of nanoscale computing. This paper's uniqueness is summarized as follows:

- Created the first fault-tolerant *7-input* Majority Gate
- Examining the suggested fault-tolerant gate for faults such as cell omission, cell displacement, cell misalignment, and extra-cell deposition.
- Using the QCAPro tool, compute and analyze the power consumption of the suggested structure.
- Using the proposed fault-tolerant *7-input* Majority Gate, propose fault-tolerant fundamental gates such as *4-input* AND and *4-input* OR.

This paper's structure will be as follows. Section 2 provides an overview of QCA technology, including QCA cells, QCA wires, fundamental QCA gates, clocking, different types of faults, and related operations. Section 3 proposes fault-tolerant 7-input Majority Gates. The proposed structure in Section 4 suggests fault-tolerant versions of fundamental gates such as *4-input AND, and 4-input OR* in QCA technology. Finally, the article is finished in Section 5.

II. Background

Introduction to QCA technology

The cell is the main component in QCA technology. The QCA cell consists of four cavities placed next to each other in a square shape. The QCA cell has two extra electrons, which can move freely between holes [15]. In general, *6* different states are available for *2* electrons to be placed in *4* holes. All these *6* states are not stable because, due to the Coulomb force between electrons, they are always in a state that has the greatest distance from each other, as shown in Figure 1. Therefore, the stable states are established when the holes are diagonally occupied, which creates two structures. These two structures display two poles, *+1 and -1*, which in calculations, we attribute logical values of *1 and 0* to them, respectively [13].

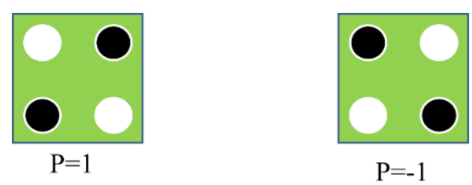


Fig. 1. QCA cell [16].

Another important issue in this technology is wiring. Not only is the Coulomb force established between the electrons inside a cell, but also each cell affects the adjacent cells. If two cells are located next to each other, they are always in a position where the Coulomb force is minimized. An array of contiguous cells can be used to transmit a logical value like a wire to transmit information [17, 18]. A linear array of standard cells can transmit binary data. In general, two wiring methods have been used in QCA, as shown in Figure 2. The first method, which is considered the wiring standard in QCA, is created by placing standard cells together; the input signal is applied from one side to the first cell, and due to electron repulsion, the polarization created in the first cell is transferred to the cells [19]. Adjacent ones are transmitted one after the other, and thus the signal is propagated in the first part of the wire. Finally, the incoming signal can be seen on the other side of the wire. The second method that is used in wiring is by using *45-degree* cells [20, 21]. In this way, the cells are placed together with a polarity between *-1 and +1*, so a wire is created using these cells, and the signal can be transmitted. Another benefit of this technology is its capacity to cross wires. Since the cables are arranged in separate layers in this instance, they don't negatively impact one another. Additionally, there is no signal loss when using the same wire for multiple outputs [17, 18, 22].

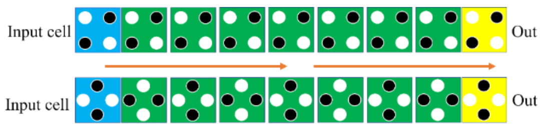


Fig. 2. Two types of QCA wire [23]

In addition to the above, clocking is very important in this technology. A QCA circuit cannot function properly without a clock schedule. In QCA, we must use a schedule that allows each cell to maintain its current state and not react to the change of state of the neighboring cells. The clock is an electronic factor that controls the movement of electrons inside the cell [24]. What the clock does is keep one set of cells in a certain polarization to allow other cells to make the appropriate changes [25]. The way it is controlled is that if information reaches a part of the circuit that must be combined with several other inputs and produce the desired output, if other inputs arrive later in that part of the circuit, it prevents information from spreading in that part until other inputs arrive. The existence of a clock creates synchronization in different parts of the circuit. Each clock signal is divided into four phases: switch, hold, release, and rest.

1. Switch Phase: During this phase, the opposing forces that hinder electron movement within each cell begin to intensify, making it increasingly difficult for electrons to move.
2. Maintenance Phase: In this stage, the forces that obstruct electron flow within the cell have reached their peak, resulting in a stable position for the electrons.
3. Release Phase: At this point, the blocking forces start to diminish, allowing electrons to be gradually released.
4. Resting Phase: In this phase, the cell exhibits no polarity, enabling electrons to move freely throughout the cell.

All cells in a clock domain have the same phase. This means that a clock phase occurs when the clock domain passes through four different phases, which is what the clock does [26]. The clocking in four phases in QCA is shown in Figure 3.

In addition to the above, the two basic gates in the QCA technology are the inverter gate and the Majority Gate, which can be used to design any circuit that is possible in this technology. In the Majority Gate, the output always shows the majority of the inputs; that is why we call it the majority function, the output is the inverse function of the majority of the inputs [27]. The types of Majority Gates in this technology are *3-input, 5-input, and 7-input*. Figure 4 shows *3, 5, and 7-input* Majority Gates and an inverter gate.

Defects in QCA

Also, one of the important things in this technology is the possibility of a fault. Defects in the placement stage can have a big impact on the system's functionality and dependability when it comes to QCA technology [28-30]. These flaws cover a wide range of anomalies or mistakes that could occur when arranging or placing quantum dots, which are the basic building blocks of QCA circuits. Potential flaws in the placement phase could include differences in the distance between quantum dots or misalignment of the dots themselves. These flaws could be the consequence of fabrication process restrictions, environmental variables, or manufacturing errors. It is imperative to address and mitigate these placement-related defects because they have the potential to affect both the accuracy of quantum computations and the overall performance of QCA-based devices. To increase the dependability and usefulness of QCA technology, scientists and engineers working in this area try to create strong methods for locating, stopping, and fixing these flaws [28-30]. The types of possible defects in the placement stage of this technology include:

- Cell deletion: It is a defect where a cell is left out of the circuit (Figure 5 (a))
- Cell displacement: defects that quantum cells move in their original direction (Figure 5 (b))
- Cell rotated: It is a defect in which the cell rotates in proportion to the adjacent cells in such a way that the direction of the cell changes (Figure 5 (c))
- Extra-cell: It is a defect in which one or a number of extra cells may be added to the circuit in the process of placing cells (Figure 5 (d)) [7].

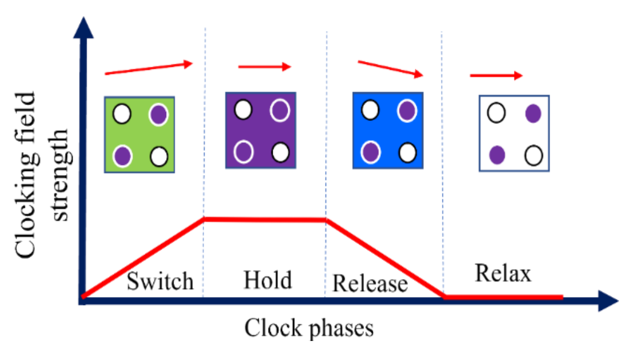


Fig. 3. Clocking in QCA technology [26]

Related works

In reference [31], a new 7-input Majority Gate is designed in QCA. The gate proposed in this article is designed in one layer and with a *0.25 clock cycle*. This gate is designed with *19* quantum cells, and is shown in Figure 7 (a). Also, the power dissipation map of this gate is shown in Figure 7 (a).

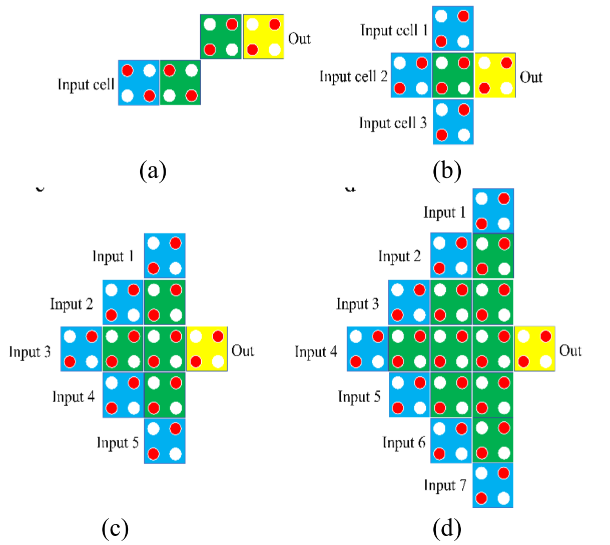


Fig. 4. QCA basic gates: (a) QCA inverter, (b) a three-input Majority Gate, (c) a five-input Majority Gate, and (d) a seven-input Majority Gate [3]

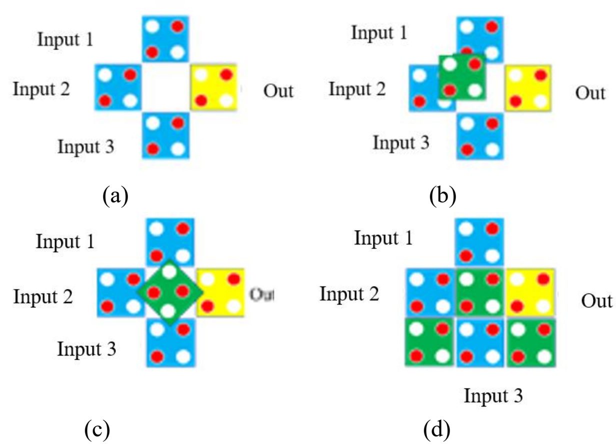


Fig.5. Types of defects in QCA technology: (a) cell omission, (b) cell displacement, (c) cell rotation, and (d) extra-cell deposition [7]

This gate's fault tolerance of less than *12%* indicates that it is quite prone to malfunctions. When the fault tolerance level is below a critical threshold in real-world terms, it can be said that the gate is not fault-tolerant. In this case, the gate's low fault tolerance points to a significant susceptibility to errors or disruptions, highlighting the need for additional advancements or substitute approaches to raise its dependability in practical uses.

In reference [32], a 7-input Majority Gate with a symmetrical structure is proposed. This research article presents a gate that is carefully constructed in a single layer and runs at *0.25 clock phase*. According to the reference, this gate's architectural layout consists of *24* cells in total. But a critical observation reveals that this specific gate, as shown in Figure 7 (b), is not intrinsically fault-tolerant. The power dissipation map of this gate is shown in Figure 7 (b). One important thing to note about this gate is that it doesn't have fault tolerance, which means that errors or disruptions could compromise its dependability in real-world scenarios. Its intricate design, in particular its reliance on a single clock phase and single-layer structure, makes it more prone to malfunctions.

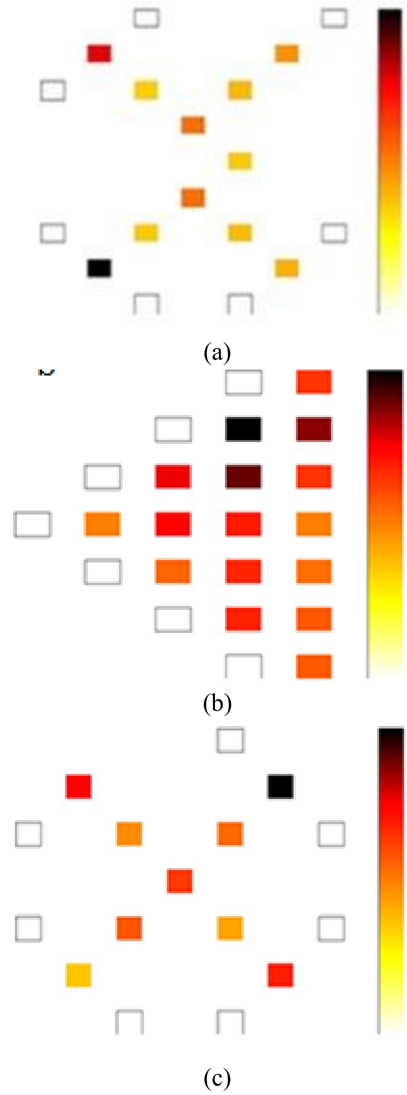


Fig.6. Power consumption analysis: (a) Power dissipation map of the in [31], (b) Power dissipation map of the in [32], and (c) Power dissipation map of the in [33]

In the minority gate, the output is the inverse function of the majority of the inputs. The proposed 7-input minority gate in reference [33] is shown in Figure 7 (c). This research article introduces a novel construction, the 7-input minority gate, which is made up of *16* cells that are synchronized within a small *0.25* clock cycle. Figure 6 (c) displays this gate's power dissipation map. The structural design consists of seven input cells, one output cell, and a middle configuration of eight cells. Note that this gate has a fault tolerance level of less than *13.75%,* even with its intricate design and effective clock cycle. The gate is considered to have no substantial error tolerance in practice when the fault tolerance figure falls below a critical threshold. Put differently, this means that the gate is more prone to errors than is reasonable, which highlights the need to strengthen its fault tolerance.

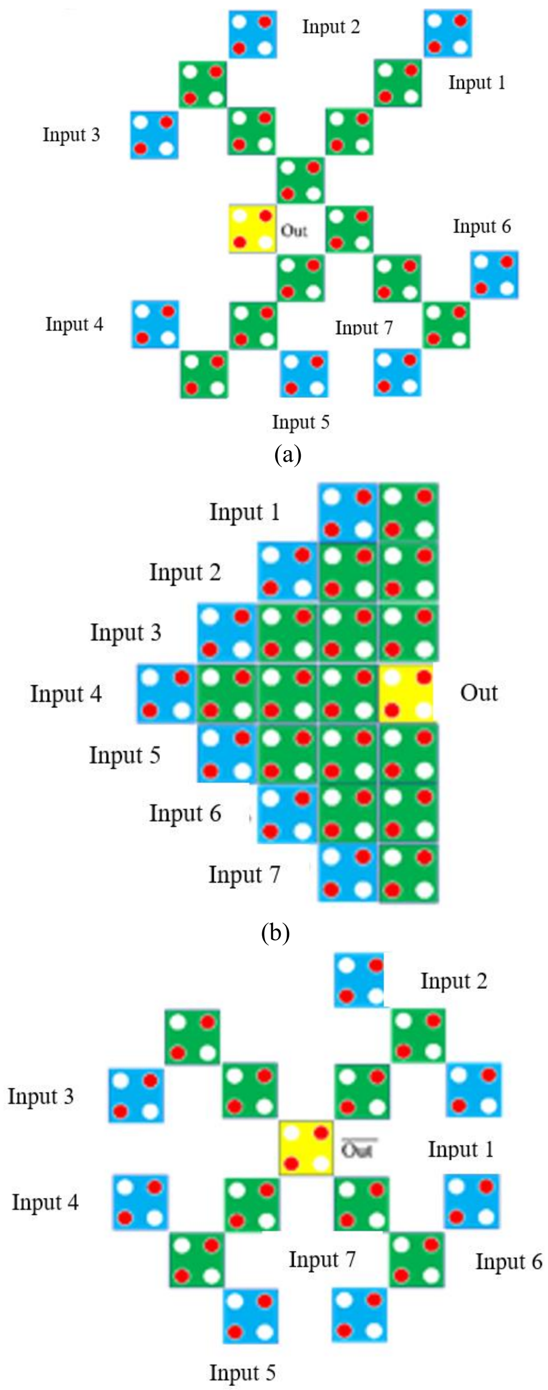


Fig.7.Three types of 7-input Majority Gate: (a) 7-input Majority Gate proposed in [31], (b) 7-input minority gate proposed in [32], and (c) 7-input Majority Gate in [33]

III. The proposed fault-tolerant 7-input Majority Gate

In QCA technology, a fault-tolerant 7-input Majority Gate is a state-of-the-art development in the realm of quantum computing. In comparison to classical computing systems, quantum computation at the quantum level (*QCA*) enables faster and more complex operations by applying the principles of quantum mechanics to computing. It is a revolutionary approach to computing. This particular part of a 7-input Majority Gate is made to process seven input signals and generate an output using the majority logic principle. In the field of QCA technology, this gate's fault-tolerant nature is essential because it shows how well it can tolerate and lessen the effects of any faults, mistakes, or disturbances that might occur while it is operating. Therefore, in this article, a fault-tolerant 7-input Majority Gate is presented. Figure 8 represents the proposed fault-tolerant 7-input Majority Gate with *66* quantum cells in an occupied space of $0.03\ \mu m^2$. All input and output cells are designed and implemented in one layer, and no other layers are needed to access them. Proper arrangement of the input cells and the length of the wires structurally have led to the coordinated effects, and accordingly, the fault-tolerant 7-input Majority Gate output. The correct output response is received after 0.25 clock cycles.

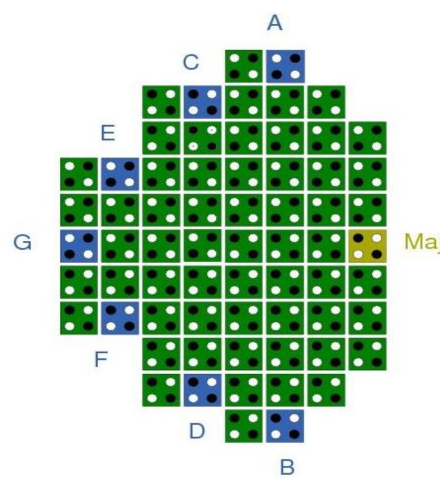


Fig.8.The QCA layout of the proposed fault-tolerant 7-input Majority Gate

Also, we have designed *4-input AND* and *4-input OR* gates with this gate as follows to show the efficiency of the presented gate. One of the main advantages of the 7-input Majority Gate is the ability to implement logic gates (*4-input AND, 4-input OR, 4-input NAND, and 4-input NOR* in QCA technology) with more inputs under the capability of one gate. The fault-tolerant *4-input AND, 4-input OR* can be designed within the fault-tolerant *7-input Majority Gate* by fixing three of the inputs to *0 and 1* as shown in Figure 9. When three inputs are set to *0*, the output of the function is *1* only when the other four inputs are *1*; otherwise, the output of the function is *0*. Therefore, the system will behave like an *"AND"* operator. A gate *"OR"* with four inputs can be designed by fixing three of the inputs to *+1* polarization. Functionality of these gates is presented in Eqs (1) and (2):

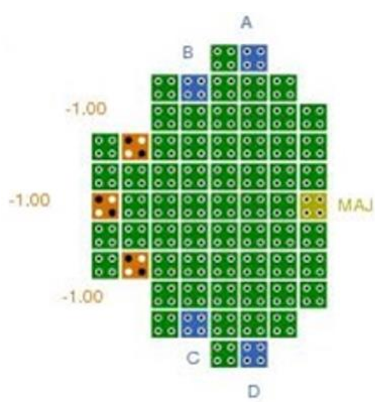
$$AND(A,B,C,D)=AB.CD=MV7(A,B,C,D,0,0,0) \quad (1)$$

$$OR(A,B,C,D)=A+B+C+D=MV7(A,B,C,D,1,1,1) \quad (2)$$

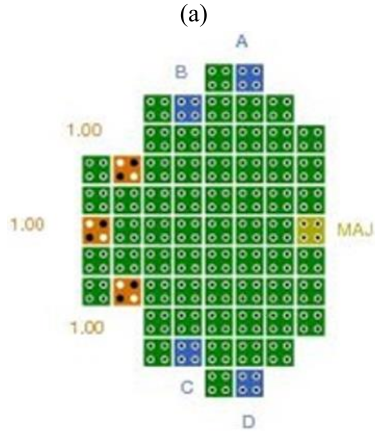
To ensure fault tolerance, structural redundancy was introduced in the design by adding the overlapping logical paths and the placement of cells with the potential to furnish logical compensation for local failures. In contrast to duplication or voting methods, this design emphasizes spatial distribution and logical symmetry that ensures functionality even when some cells are deleted or displaced. This system maintains gate functionality without unnecessary area or complexity overhead.

IV. Simulation software, results, energy dissipation analysis, and defect analysis

To thoroughly assess and validate the suggested circuit layouts and their associated functionality, we have utilized the sophisticated simulation tool QCADesigner version 2.0.3, which is designed especially for QCA circuits [8, 34]. With the aid of this simulation tool, we can evaluate the complex dynamics of QCA circuits and closely examine the proposed circuit's performance in a variety of scenarios. To guarantee the precision and dependability of the outcomes in the context of bistable approximation—a crucial component of QCA circuit simulations—a set of well-selected parameters has been provided. These specifications call for an 18 nm cell size and the use of *50,000* samples to provide a strong statistical analysis. Notably, the precise value of *0.0000100* is chosen for the convergence tolerance, highlighting the level of accuracy needed to achieve convergence during the iterative simulation process. In addition, several parameters that control the simulation environment have been carefully defined. The parameters used include the radius of effect (65 nm), relative permittivity (12.9), clock high (9.8e-22 J), clock low (3.8e-23 J), clock shift (0), clock amplitude factor (2.0), layer separation (11.5 nm), and maximum iterations per sample (100). Notably, most of these parameters are the QCADesigner defaults, guaranteeing uniformity and conforming to accepted practices in QCA simulations. Through the utilization of QCADesigner features and prudent parameter configuration, our goal is to perform a comprehensive analysis that examines the behavior of the suggested circuit in various scenarios [8, 34]. Ensuring the precision, dependability, and thorough comprehension of the suggested QCA circuit's performance attributes requires the careful selection and adjustment of these simulation parameters.



(a)



(b)

Fig.9. (a) The QCA layout of proposed fault-tolerant 4-input AND gate, and (b) The QCA layout of proposed fault-tolerant 4-input OR gate

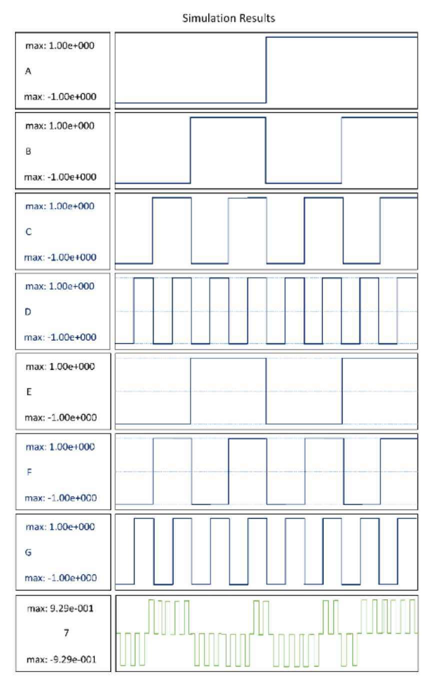
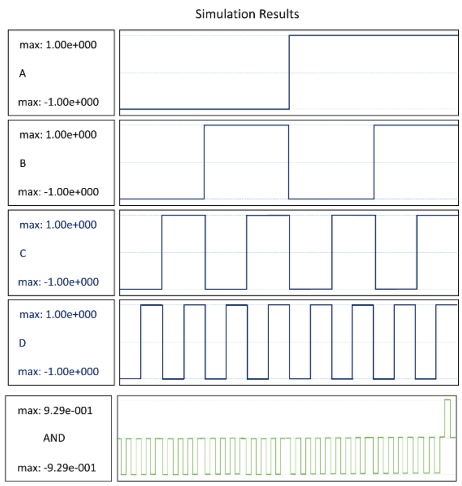
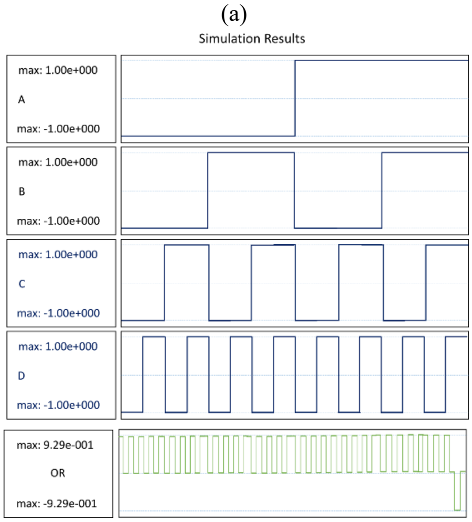


Fig.10.The simulation results of the proposed fault-tolerant 7-input Majority Gate

The simulation results from the QCADesigner simulation motors indicate the proper operation of the proposed Majority Gate for all possible inputs. The test vectors were applied sequentially to the inputs, and the results were examined by each motor separately. The matching of the waveforms revealed the correct operation of the design. Figure 10 represents the simulation result with the *Coherence Vector* simulation motor. Additionally, Figure 11 shows the simulation results for an *OR* and an *AND* gate with four inputs that are fault-tolerant. The figure makes it clear that all potential inputs have been applied to the gates and that the gates have produced the expected and correct outputs, demonstrating the proper operation of the newly introduced gates.



(a)



(b)

Fig.11. (a) The simulation results of the proposed fault-tolerant 4-input AND gate, and (b) The simulation results of the proposed fault-tolerant 4-input OR gate

Considering that the *7-input gate* designed in this article is the first fault-tolerant *7-input Majority Gate*, therefore, no comparison has been made to this gate; however, Table 1 compares the 7-input minority and Majority Gates designed in previous articles with the gate proposed in this article. In this table, the proposed gate has been compared with the circuits designed in previous articles in terms of fault-tolerance percentage.

Also, to better confirm the accuracy of the presented fault-tolerant 7-input Majority Gate, this gate has been tested in QCAPro software. To calculate polarization error and non-adiabatic switching power loss in QCA circuits, a new modeling tool called QCAPro has been developed. When designing QCA circuits, the tool estimates highly erroneous cells using a quick approximation-based technique. In addition to providing an upper bound on power consumed, QCAPro estimates the power loss in a QCA circuit for clocks with sharp transitions, which lead to non-adiabatic operations. When an input switching operation is occurring in a QCA circuit, QCAPro can be used to estimate the maximum, minimum, and average power losses. The QCAPro software can measure the energy consumption of the fault-tolerant 7-input Majority Gate at three levels of *0.5 EK, 1 EK, and 1.5 EK* at *2.0 K*. The power dissipation map of the proposed Majority Gate with 0.5 EK is illustrated in Figure 12, and more energy dissipation details of the gate are listed in Table 2. Also, in this table, a comprehensive comparison has been made between the presented design and the best previous designs.

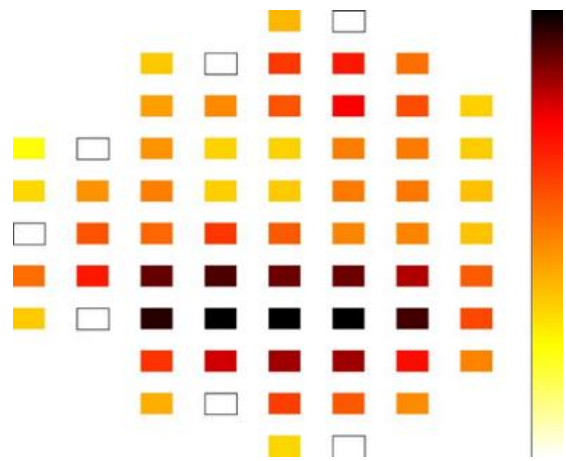


Fig.12.Power dissipation map of the proposed 7-input Majority Gate with *0.5* E_k

TABLE I. A comparison of the fault-tolerant 7-input Majority Gate and other similar gates

Design in	**Area (μm^2)**	**Accessibility to the output**	**Accessibility to the input**	**Cell deletion**	**Cell [illegible]**	**Cell Rotation**	**Extra Cell**	**Average Fault [illegible]**
propose	34444	Yes	Yes	38 %	89 %	38 %	93 %	64.5 %
Jahanshahi Javaran, Jafarali Jassbi [31]	24564	Yes	Yes	20 %	50 %	40 %	46.1%	11.525%
Navi, Chabi [32]	16284	Yes	Yes	15 %	48 %	52 %	61.1%	15.2 %
Mohammadi, Navi [33]	19044	Yes	No	12 %	39 %	37 %	55 %	13.75%

TABLE II. A comparison of the 7-input Majority Gate's energy dissipation

	Avg. leakage energy dissipation (eV)			Avg. switching energy dissipation (eV)			Total energy dissipation (eV)		
	0.5 E_k	1 E_k	1.5 E_k	0.5 E_k	1 E_k	1.5 E_k	0.5 E_k	1 E_k	1.5 E_k
Proposed	0.00918	0.03186	0.06350	0.17189	0.16451	0.15527	0.18107	0.19637	0.21877
Jahanshahi Javaran, Jafarali Jassbi [31]	0.00192	0.00365	0.00542	0.03468	0.05621	0.07932	0.12481	0.10355	0.10782
Navi, Chabi [32]	0.02996	0.03445	0.04138	0.04234	0.05483	0.08965	0.10258	0.08791	0.0389
Mohammadi, Navi [33]	0.00157	0.00294	0.00438	0.00345	0.00489	0.07814	0.10215	0.04258	0.00512

The comparison results show more energy consumption in the presented plan than in the previous plans. This is due to the high number of cells provided in the design, which has high tolerance, but the previous designs are not fault-tolerant.

Cell Deletion Analysis

Table 3 shows the results of evaluating the proposed gate's tolerance to cell deletion. By comparing the output waveforms of simulations with and without individual cell deletions (Figure 13), 22 of 58 cells could be deleted without affecting the output. This indicates a 38% tolerance to cell deletion ((22/58) × 100%).

Cell Displacement Analysis

Table 4 presents the results of the displacement fault analysis. Each of the 18 device cells was tested for displacement tolerance in four directions (North, South, East, West). Only cells 26 and 40 produced incorrect results. Therefore, the device cell fault tolerance is 89% ((16/18) × 100%).

Cell Rotation Analysis

Cell rotation evaluation was also performed in QCADesigner by changing the cell type from 90 degrees to 45. As shown in Table 5, all 58 cells were examined, and only 22 exhibited sufficient tolerance to this fault. Therefore, the proposed gate demonstrates poor tolerance to rotated cells, with a tolerance percentage of 38%.

Fig.13.The proposed fault-tolerant 7-input Majority Gate (numbered cells)

TABLE III. The output of the proposed structure against the single-cell deletion defect.

Removed cell	Output	Removed cell	Output
1	MV7	29	Incorrect
2	MV7	31	Incorrect
3	Incorrect	32	Incorrect
4	Incorrect	33	MV7
5	MV7	34	MV7
6	MV7	35	Incorrect
7	Incorrect	36	Incorrect
8	Incorrect	37	Incorrect
9	Incorrect	38	Incorrect
10	MV7	39	Incorrect
11	MV7	40	Incorrect
12	MV7	41	MV7
13	Incorrect	42	Incorrect
14	Incorrect	43	Incorrect
15	Incorrect	44	Incorrect
16	Incorrect	45	Incorrect
17	MV7	47	MV7
18	MV7	48	MV7
19	MV7	49	Incorrect
20	MV7	50	Incorrect
21	Incorrect	51	Incorrect
22	Incorrect	52	MV7
23	Incorrect	53	MV7
24	Incorrect	54	MV7
25	Incorrect	55	Incorrect
26	Incorrect	56	Incorrect
27	Incorrect	57	MV7
28	Incorrect	58	MV7

TABLE IV. Outcomes of the permitted cell displacement in the proposed design

Cell	North	South	East	West
1	∞	-	-	∞
A	-	-	<=2	-
2	∞	-	-	∞
C	<=2	-	-	-
5	∞	-	∞	-
6	-	-	-	∞
11	∞	-	∞	-
12	∞	-	-	∞
E	<=4	-	-	-
18	-	-	<=10	-
19	-	-	-	∞
26	-	-	<=1	-
G	-	-	-	<=6
Maj	-	-	<=1	-
33	-	-	-	∞
40	-	-	<=1	-
41	-	∞	-	∞
F	-	<=4	-	-
47	-	-	<=10	-
48	-	-	-	∞
53	-	∞	∞	-
54	-	∞	-	∞
D	-	<=2	-	-
57	-	∞	∞	-
58	-	∞	-	∞
B	-	-	<=2	-

Extra-Cell Deposition Analysis

In order to analyze the extra-cell fault, the process starts by adding cells all around the proposed Majority Gate, as shown in Figure 14. The resulting waveforms are compared to the fault-free ones. Represents the possible places, and Table 6 lists the results of the process. Only in cells *5* and *28*, incorrect results produced. Consequently, the fault tolerance percentage for extra-cell can be calculated as (28/30) × 100, resulting in 93%.

Testing under Noisy and Thermal Environments

To assess the stability of the proposed 7-input Majority Gate under real-world operation conditions, additional simulations were run using QCAPro for both thermal fluctuation (between 0 K and 350 K) and injected environmental noise regimes. The simulations revealed that the gate operates correctly with extremely low output errors in all 128 input combinations. This confirms the fault-tolerant character of the gate under actual nanoscale operating variations.

Trade-off Analysis: Cell Count vs. Power vs. Fault Tolerance

Though the design is more power hungry due to a higher number of quantum cells, the improvement in fault tolerance is considerable. In comparison with previous designs with smaller-sized power footprints, our design achieves a fault tolerance of up to 93% under circumstances of additional-cell faults and performs within tolerable margins under deletion and displacement faults. Such a compromise is justified in applications where fault resiliency is worth more than the lowest power, i.e., mission-critical or high-radiation environments. Table II has been accordingly adjusted to offer complete and comparative values.

TABLE V. The results of the proposed structure against single-cell rotation defect.

Rotated cell	**Output**	**Rotated cell**	**Output**
1	MV7	30	Incorrect
2	MV7	31	Incorrect
3	Incorrect	32	Incorrect
4	Incorrect	33	MV7
5	MV7	34	MV7
6	MV7	35	Incorrect
7	Incorrect	36	Incorrect
8	Incorrect	37	Incorrect
9	Incorrect	38	Incorrect
10	MV7	39	Incorrect
11	MV7	40	Incorrect
12	MV7	41	MV7
16	Incorrect	42	Incorrect
17	MV7	46	MV7
18	MV7	47	MV7
19	MV7	48	MV7
20	MV7	49	Incorrect
21	Incorrect	50	Incorrect
22	Incorrect	51	Incorrect
23	Incorrect	52	MV7
24	Incorrect	53	MV7
25	Incorrect	54	MV7
26	Incorrect	55	Incorrect
27	Incorrect	56	Incorrect
28	Incorrect	57	MV7
29	Incorrect	58	MV7

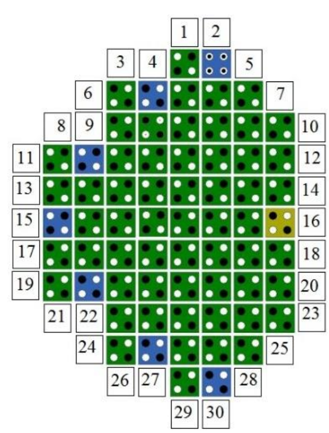


Fig.14.All instances of extra-cell deposition around the proposed gate

TABLE VI. Analysis of additional cell defects.

Additional cell	Output	Additional cell	Output	Additional cell	Output
1	MV7	11	MV7	21	MV7
2	MV7	12	MV7	22	MV7
3	MV7	13	MV7	23	MV7
4	MV7	14	MV7	24	MV7
5	Incorrect	15	MV7	25	MV7
6	MV7	16	MV7	26	MV7
7	MV7	17	MV7	27	MV7
8	MV7	18	MV7	28	Incorrect
9	MV7	19	MV7	29	MV7
10	MV7	20	MV7	30	MV7

V. Conclusion and future works

Scientists have been actively searching for new approaches to address persistent issues in Very Large-Scale Integration (VLSI) systems as the field of nanotechnology develops quickly and the demand for faster processing in smaller dimensions rises. These challenges include things like high power consumption, low operating speeds, and large-scale area requirements, which necessitate new strategies to reduce these limitations. Quantum-Dot Cellular Automata (*QCA*) technology, developed and applied at the nanoscale, offers itself as a potential solution to these challenges. Its growing popularity can be attributed to its special features, which include eliminating electric current and the need for capacitive components, both of which make the design much simpler. With the introduction of QCA technology, we are moving closer to the nanoscale, where the benefits of faster processing, smaller size, and lower power consumption become apparent. According to the above scenarios and the relevance of Majority Gates, a novel fault-resistant *7-input Majority Gate* has been assembled for the first time in QCA technology employing hardware redundancy. In addition to structural analysis, the study of gate tolerance was done using the four fault types: cell omission, cell displacement, cell rotation, and extra-cell deposition, and the accuracy of gate operation was assessed using the *QCADesigner* simulation tool. The proposed Majority Gate is composed of *66* cells, which occupy *0.03* μm^2 with *0.25 clock cycle* propagation delay. Instead of providing a single averaged fault tolerance measure, individual fault tolerance rates are provided for each fault type: cell deletion (38%), displacement (89%), rotation (38%), and extra-cell insertion (93%). This initial analysis provides a more accurate indication of fault behavior and avoids misleading statistical summaries. A fault is "tolerated" when the output is functionally correct for at least 90% of input combinations under that fault. Also, for the first time, various types of fault-tolerant basic gates, such as *4-input AND, 4-input OR*, in QCA technology are implemented. Future research will expand on the current study's contributions by optimizing and expanding fault-tolerant basic gates in QCA technology. To increase the versatility and dependability of these gates in nanoscale circuit design, researchers will investigate a variety of types, including QCA basic gates. By addressing both theoretical and practical aspects of fault tolerance in QCA-based systems, the goal is to advance the field and lay the groundwork for future VLSI technologies that will be more reliable and efficient.

In summary, the novel fault-tolerant 7-input majority gate is highly resilient to various types of faults without compromising sound logic functionality. Although it requires more quantum cells and thus marginally more power, the trade-off is worthwhile for high-reliability applications. Future work will focus on layout optimization to continue reducing power and area, including adding the fault-tolerant design to 9-input and hybrid majority gates and exploring fabrication-aware designs. This gate's performance and robustness make it especially well-suited for QCA-based systems in space, military, or biomedical applications where reliability is the top priority.

References

[1] Ahmadpour, S.S., M. Mosleh, and S. Rasouli Heikalabad, Robust QCA full adders using an efficient fault tolerant five input majority gate. International Journal of Circuit Theory and Applications, 2019. 47(7): p. 1037-1056.

[2] Raj, M., L. Gopalakrishnan, and S.B. Ko, Reliable SRAM using NAND-NOR gate in beyond CMOS QCA technology. IET Computers & Digital Techniques, 2021. 15(3): p. 202-213.

[3] Safaiezadeh, B., et al., Design and simulation of efficient combinational circuits based on a new XOR structure in QCA technology. Optical and Quantum Electronics, 2021. 53: p. 1-16.

[4] Li, Z., S. Zhang, and B.O. Mohammed, An efficient three-level nano-design for reversible gate using quantum dot-cellular automata with cost analysis. Materials Science and Engineering: B, 2023. 294: p. 116526.

[5] Chen, X., et al., Accurate reliability evaluation using quantum dot cellular automata probabilistic transfer matrix. Micro & Nano Letters, 2014. 9(2): p. 77-82.

[6] Norouzi, M., S.R. Heikalabad, and F. Salimzadeh, A reversible ALU using HNG and Ferdkin gates in QCA nanotechnology. International Journal of Circuit Theory and Applications, 2020. 48(8): p. 1291-1303.

[7] Seyedi, S. and N.J. Navimipour, Design and evaluation of a new structure for fault-tolerance full-adder based on quantum-dot cellular automata. Nano communication networks, 2018. 16: p. 1-9.

[8] Walus, K., et al., QCADesigner: A rapid design and simulation tool for quantum-dot cellular automata. IEEE transactions on nanotechnology, 2004. 3(1): p. 26-31.

[9] Jain, V., et al., Comprehensive and Comparative Analysis of QCA based Circuit Designs for Next Generation Computation. ACM Computing Surveys, 2023.

[10] Ahmadpour, S.-S., et al., An ultra-efficient design of fault-tolerant 3-input majority gate (FTMG) with an error probability model based on quantum-dots. Computers and Electrical Engineering, 2023. 110: p. 108865.

[11] Roohi, A., et al., A symmetric quantum-dot cellular automata design for 5-input majority gate. Journal of Computational Electronics, 2014. 13: p. 701-708.

[12] Farazkish, R. and K. Navi, New efficient five-input majority gate for quantum-dot cellular automata. Journal of Nanoparticle Research, 2012. 14: p. 1-6.

[13] Jafarali Jassbi, S., et al., Design and analysis of a fault tolerant 3-input majority gate in quantum-dot cellular automata. Journal of Advances in Computer Research, 2019. 10(4): p. 27-36.

[14] Majeed, A.H., et al., A new 5-input majority gate without adjacent inputs crosstalk effect in QCA technology. Indonesian Journal of Electrical Engineering and Computer Science, 2019. 14(3): p. 1159-1164.

[15] Majeed, A.H., et al., Optimal design of RAM cell using novel 2: 1 multiplexer in QCA technology. Circuit world, 2020. 46(2): p. 147-158.

[16] Reis, D.A., et al. A methodology for standard cell design for QCA. in 2016 IEEE International Symposium on Circuits and Systems (ISCAS). 2016. IEEE.

[17] Graziano, M., et al., Characterisation of a bis ferrocene molecular QCA wire on a non ideal gold surface. Micro & Nano Letters, 2019. 14(1): p. 22-27.

[18] Roohi, A., H. Thapliyal, and R. DeMara, Wire crossing constrained QCA circuit design using bilayer logic decomposition. Electronics Letters, 2015. 51(21): p. 1677-1679.

[19] Ahmadpour, S. and N. Jafari Navimipour, A New Nano-Design of 16-bit Carry Look-Ahead Adder Based on Quantum Technology. Physica Scripta, 2023.

[20] Jahanshahi Javaran, F., et al., Robust QCA Full Adders Using a Novel Fault Tolerant Five-Input Majority Gate. International Journal of Industrial Electronics Control and Optimization, 2024. 7(2): p. 99-108.

[21] Motalebi, F. and S. Sayedsalehi, Design of Low Power Full-Adder Circuit Using Quantum-dot Cellular Automata. International Journal of Industrial Electronics Control and Optimization, 2022. 5(1): p. 99-108.

[22] Rasmi, H., et al., Novel Efficient Scalable Design of Full-Adder in Atomic Silicon Dangling Bonds (ASDB) Technology. Physica Scripta, 2023.

[23] Seyedi, S. and N.J. Navimipour, An optimized design of full adder based on nanoscale quantum-dot cellular automata. Optik, 2018. 158: p. 243-256.

[24] Vankamamidi, V., M. Ottavi, and F. Lombardi. Clocking and cell placement for QCA. in 2006 sixth IEEE conference on nanotechnology. 2006. IEEE.

[25] Seyedi, S. and N.J. Navimipour, Designing a three-level full-adder based on nano-scale quantum dot cellular automata. Photonic Network Communications, 2021. 42: p. 184-193.

[26] Seyedi, S. and N.J. Navimipour, An efficient structure for designing a nano-scale fault-tolerant 2: 1 multiplexer based on quantum-dot cellular automata. Optik, 2022. 251: p. 168409.

[27] Khan, A., et al., Behaviors of QCA inverter due to cell displacement and temperature variation. arXiv preprint arXiv:2310.09734, 2023.

[28] Bahadori, G., M. Houshmand, and M. Zomorodi-Moghadam, Design of a fault-tolerant reversible control unit in molecular quantum-dot cellular automata. International Journal of Quantum Information, 2018. 16(01): p. 1850010.

[29] Poorhosseini, M. and A.R. Hejazi, A fault-tolerant and efficient XOR structure for modular design of complex QCA circuits. journal of Circuits, Systems and Computers, 2018. 27(07): p. 1850115.

[30] Seyedi, S. and N.J. Navimipour, A fault-tolerance nanoscale design for binary-to-gray converter based on QCA. IETE Journal of Research, 2023. 69(5): p. 2991-2998.

[31] Jahanshahi Javaran, F., et al., Computational Circuit Design Using a New Seven-Input Majority Gate in Quantum-dot Cellular Automata. Journal of Intelligent Procedures in Electrical Technology, 2022: p. 21-34.

[32] Navi, K., A.M. Chabi, and S. Sayedsalehi, A novel seven input majority gate in quantum-dot cellular automata. International Journal of Computer Science Issues (IJCSI), 2012. 9(1): p. 84.

[33] Mohammadi, H., K. Navi, and M. Hosseinzadeh, An efficient quantum-dot cellular automata full adder based on a new convertible 7-input majority-not gate. IETE Journal of Research, 2023. 69(1): p. 558-566.

[34] Aralikatti, S. QCA Designer: A simulation and Design Layout Tool for QCA based Nano Domain Computing Architectures. in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). 2020. IEEE.

Farzaneh Jahanshahi Javaran was born in Kerman, Iran She a PhD student in Computer Systems Architecture from Islamic Azad University, Science and Research Department. University of Science and Research is a prestigious university in the field of computer systems architecture in Iran. Currently, her researching the design of fault-tolerant majority gates in quantum cellular automata. Her current research interests include quantum dot cellular automata, fault tolerance.

Sommayeh Jafarali Jassbi was born in Tehran, Iran, in 1982. She received the MSc degree in computer architecture engineering in 2007, and the Ph.D. degree in computer architecture engineering in 2010 from the Islamic Azad University, Science and Research Branch. In 2010, she joined the Department of computer engineering, Islamic Azad University Science and Research Branch. She became an associate professor in 2011. Her interests are cloud computing, internet of things, wireless sensor network and computer architecture and cryptography. She was head of computer department in 2012. Now she is selected as a head of computer department again. She was also an active member of young researcher club from 2004. She has written, translate and published several professional books and paper in her fields.

Hossein Khademolhosseini received B.Sc. degree in computer engineering in 2008 from Shiraz University, Shiraz, Iran. He also received his M.Sc. and Ph.D. degree in computer architecture at Department of Computer Engineering, Science and Research Branch of Islamic Azad University, Tehran, Iran, in 2011 and 2016, respectively. He is currently an assistant professor with the Department of Computer Engineering, Islamic Azad University, Beyza Branch. His research interests are computer arithmetic, photonic NoC and electronics with emphasis on QCA and VLSI.

Razieh Farazkish received the B.S. degree in computer engineering from the IAU, Central Tehran Branch (2007) and the M.S. (2009) and Ph.D. (2012) degrees in computer engineering from the IAU, Science and Research Branch. In 2012, she joined the Department of Computer Engineering, IAU, South Tehran Branch, as a Professor. Her current research interests include quantum-dot cellular automata, fault tolerance, nanoelectronic circuits, nano computing, testing and design of digital systems.

IECO International Journal Of Industrial Electronics Control and Optimization
Print ISSN: 3517-2645
Online ISSN: 3568-2645
Homepage: https://ieco.usb.ac.ir/

Semi-Supervised Clustering with Improved Delaunay Graph Fusion and Pairwise Constraints

Shahin Pourbahrami

Department of Computer Engineering, Technical and Vocational University (TVU), Tehran, Iran.
Corresponding author's email: shpourbahrami@tvu.ac.ir

Article Info

Article type:
Research Article

Article history:
Received: 17-July-2025
Received in revised form: 07-September-2025
Accepted: 06-November-2025
Published online: 23-Sep-2026

Keywords:
Delaunay graph,
clustering,
structural knowledge,
pairwise constraints

ABSTRACT

In many data mining problems, leveraging structural and local connectivity information can significantly improve clustering performance. This paper presents a novel semi-supervised clustering framework that integrates weighted feature information, Delaunay-based graph construction, and pairwise constraints. First, feature weights are computed based on within-class pairwise variability, emphasizing dimensions that contribute most to local cluster structure. Weighted distances between samples are then calculated, and a Delaunay graph is constructed and filtered using an influence radius, preserving meaningful local geometric relationships while removing redundant edges. To capture higher-order neighborhood information, a GraphSAGE-style embedding propagates feature information through the graph, generating enriched low-dimensional representations of the data. Pairwise constraints are incorporated into the similarity matrix to encode prior knowledge about sample relationships, guiding the clustering process. Finally, semi-supervised clustering is performed using constraint-based spectral clustering. Experiments on benchmark datasets demonstrate that the combination of structural graph information, feature weighting, and pairwise constraints substantially improves clustering accuracy. The proposed framework is flexible and can be effectively applied across diverse data domains.

I. Introduction

In many data mining and machine learning tasks, clustering algorithms serve as essential tools for uncovering hidden structures within heterogeneous datasets [1]. Recent advances in this field have led to domain-specialized approaches tailored to a wide range of scientific and industrial applications. For instance, multi-scale graph convolutional networks (MS-GCNs) have shown strong performance in image clustering by leveraging both local and global structural information to generate more accurate and semantically meaningful groups [2]. In bioinformatics, clustering of single-cell RNA-Seq data has enabled the discovery of rare cell populations and complex gene-expression patterns, with modern methods combining deep autoencoders and graph convolutional architectures to preserve biological structure more effectively [3]. Clustering has also been applied to reduce data complexity and to more effectively model uncertainties in load and wind speed [4]. Similarly, contrastive learning–based clustering has improved image segmentation in medical and remote-sensing applications by creating more discriminative feature spaces [5]. Dynamic clustering of temporal networks has further enhanced the understanding of evolving communities and behavioral shifts in real-world social systems [6].

Semi-supervised clustering approaches have emerged in recent years, attempting to integrate prior knowledge (such as pairwise constraints) with unsupervised learning methods [7, 8]. On the other hand, using graph structures to model the geometric relationships of data has also proven to be an effective strategy for better representing data structure [9]. Among these, the Delaunay graph [10], as a geometric structure that naturally forms a network of relationships between points, has high potential for analyzing multidimensional data.

This study presents a novel framework for semi-supervised clustering that integrates a weighted Delaunay graph with structural knowledge (via feature weighting) and pairwise constraints. Using graph embedding techniques such as GraphSAGE-style embedding, the framework produces more meaningful numerical representations of the data. This method aims to improve clustering accuracy and robustness in handling complex and imbalanced datasets, while offering strong generalizability across diverse data domains.

The structure of this paper is organized as follows: Section 2 reviews graph-based and geometry-based clustering algorithms, highlighting their strengths and limitations. Section 3 introduces the proposed algorithm that utilizes semi-supervised clustering with a combination of Delaunay graph and pairwise constraints. Section 4 presents and discusses experimental results, demonstrating the enhanced capability of the proposed method in identifying complex and nonlinear structures compared to traditional and advanced clustering techniques. Finally, Section 5 concludes the paper.

II. Related Work

Semi-supervised clustering via pairwise-constrained optimal graphs incorporates must-link and cannot-link constraints to optimize the graph structure [7]. A similarity graph is first constructed based on data points and then refined using pairwise constraints to better reflect the underlying cluster structure. Laplacian constraints preserve local graph connectivity, while a label propagation algorithm spreads supervision across nodes so that similar points are more likely to share the same label. This process produces cluster assignments that balance the internal data structure with limited external supervision.

Semi-supervised clustering via structural entropy introduces a method that integrates pairwise and label constraints into a structural entropy framework. The proposed algorithms support both flat and hierarchical clustering and have shown effectiveness in analyzing biological data such as single-cell RNA [8]. The approach constructs two graphs: a data graph based on similarity and a constraint graph encoding positive and negative relationships. A bi-dimensional structural entropy objective function, incorporating constraint consistency, enables simultaneous flat and hierarchical clustering.

Semi-supervised deep embedded clustering with pairwise constraints and subset allocation redefines the loss function based on sample similarity and introduces a subset assignment loss to improve performance on industrial textual data [11]. The method employs a similarity-based loss to better distinguish classes and a subset assignment loss to more effectively transfer knowledge from labeled data.

Semi-supervised clustering using pairwise constraints and local density structures is based on local density peaks and graph cuts. Pairwise constraints are used to segment inconsistent local trees, while the similarity matrix is refined with the E2CP algorithm [12]. The method incorporates intra- and inter-cluster conflict resolution and prevents tree fragmentation through root-changing and noise filtering. Finally, spectral clustering combined with E2CP further improves clustering accuracy.

"Some Methods for Classification and Analysis of Multivariate Observations" by J. MacQueen (1967) is a foundational work that introduced the *K*-Means algorithm. This iterative method partitions data into k clusters aiming to minimize the sum of squared distances from cluster centers. Though easy to implement and fast, it has drawbacks: sensitivity to initialization, inability to detect non-spherical clusters, poor noise handling, and requiring a pre-set number of clusters [13, 14].

Density-Based Spatial Clustering for Noisy Applications: This method introduces DBSCAN, a density-based algorithm to discover clusters in large spatial databases with noise. DBSCAN can detect clusters of arbitrary shapes and identify outliers without needing the number of clusters. However, it depends heavily on proper selection of parameters (ε neighborhood radius, *MinPts* minimum points), which can cause incorrect clustering or merging if misconfigured. Performance also declines sharply with variable-density data [15, 16].

Rodriguez and Laio introduced the DPC (Density Peak Clustering) algorithm, which discovers data clusters based on identifying points with high density and large distances from other dense points [17]. This method does not require specifying the number of clusters and intuitively detects cluster centers. However, its accuracy declines when facing complex structures and data with varying density. The study in [18] improves DPC using geodesic distance to identify clusters with more complex boundaries and uneven distributions. Despite this improvement, the new method remains sensitive to proper parameter selection, and computing geodesic distances can be time-consuming and costly in very large datasets.

A deep clustering approach extending *k*-means is proposed, where each cluster is represented by an auto encoder instead of a single centroid (*K*-DAE) [19]. Data points are assigned to the autoencoder with minimal reconstruction error, and clustering is optimized by minimizing the global reconstruction MSE across all auto encoders.

However, the performance of these algorithms, particularly with complex, noisy, and high-dimensional data, heavily depends on how similarity between samples is defined and on the intrinsic structure of the data [13, 15, 16]. One common limitation of traditional clustering methods is their neglect of structural knowledge. Yet in many real-world problems, valuable information about the relative importance of features or relationships between samples is available, which can significantly enhance clustering quality.

III. Methodology

A. Feature Weighting

In this stage, each feature in the dataset X ($X \in \mathbb{R}^{n \times d}$) is assigned a weight, which has n samples and d dimensions. These weights are determined by analyzing the distances within the data points $d(x_i - y_j)$. The feature weights in Formula 1 are calculated.

$$w_k = 1/\sum_{k=1}^{d} \sqrt{\sum_{i \neq j=1}^{n} d(x_{ik} - y_{jk})^2} \quad (1)$$

Where w_k is the weight of dimension (feature) k. A weight vector (feature weights) is defined whose size is equal to the number of features, and each element represents the importance of that feature in the distance calculations. In particular, if a larger weight is chosen for a feature, changes in that feature will have a greater effect on the overall distance and thus play a stronger role in clustering. This ensures that more significant features have greater influence on point connectivity and clustering. Finally, a feature receives a higher weight if the set of data points within that feature exhibits smaller within-class distances.

B. Combining the Delaunay Graph with Structural Knowledge

In the Delaunay graph, data points are connected to form triangles based on their geometric distribution, ensuring that no data point lies inside the circumcircle of any triangle [10]. This preserves the natural structure of the data. When certain features are known to be more important, this structural knowledge can be incorporated into the graph through feature weighting, thereby influencing distance calculations and the resulting embeddings to better reflect their importance.

C. Calculating Influence Radius and Filtering Edges Based on Circle Overlaps

For each data point, the influence radius defines how far its effect extends or, in other words, which neighboring points are within range. This process ensures only edges where both endpoints are meaningful neighbors are retained. Unlike a pure Delaunay graph that retains all triangle edges, this approach retains only structurally significant edges. For each point in the Delaunay graph set, the distance to the *k*th nearest neighbor is taken as the radius of influence for that point (r_{x_i}). In this way, the radius of influence of each point is determined based on the distance to several of its nearest neighbors and provides a measure of the scope of influence or local connections of that point. In Formula 2, edges whose length is greater than the sum of the radii of the two connected points are removed. This leaves only edges that indicate the true proximity between the two points.

$$keepedge(x_i, y_j) if d(x_i, y_j) < r_{x_i} + r_{x_j} \quad (2)$$

A similarity matrix is then constructed from the resulting graph connections (showing the relationship of each point to its neighbors in an adjacency matrix). Our innovation is that we improve the Delaunay graph connection method to preserve meaningful neighborhoods with a new and simple definition of neighborhood radius. We also preserve meaningful relationships and eliminate unlikely neighborhoods by embedding the Delaunay graph structure with the GraphSAGE algorithm.

D. Modifying the Similarity Matrix Using Pairwise Constraints [8] (Must-link and Cannot-link)

Suppose there is an initial similarity matrix $S \in R^{n \times n}$ built from the Delaunay graph. This matrix is then adjusted using constraints:

Must-link: For each pair *(i, j)* that must belong to the same cluster, their similarity is set to 1 (Formula 3).

$$S_{ij} = S_{ji} = max(S_{ij}, \alpha)\alpha = 1 \quad (3)$$

Must-link constraints ensure that certain points stay together, even if they appear far apart in feature space.
Cannot-link: Prevents incorrect grouping by setting the similarity of such pairs to zero or a very low value (Formula 4).

$$S_{ij} = S_{ji} = 0 \quad (4)$$

E. Creating the Embedded Delaunay Graph

Instead of working directly on raw data, graph embedding algorithms such as Node2Vec and GraphSAGE are used to generate a numerical (embedded) representation that models the Delaunay graph structure more effectively. These embeddings are vectors that combine both the original features and the graph structure.

Node2Vec and GraphSAGE are graph embedding algorithms that aim to transform a graph into a vector space, preserving the structural properties of the graph and the similarity between nodes in this vector space. Each node in the graph (e.g., a data instance) is mapped to a numerical vector in a multidimensional space. Nodes that have similar or close structures in the graph will have close vectors. This feature makes clustering on vectors more accurate and flexible. Combining a weighted Delaunay graph with GraphSAGE allows preserving the structural and neighborhood information of the graph. As a result, the clustering algorithm becomes more accurate and robust to noisy points.

GraphSAGE is a representation learning method for graphs that, instead of looking at the entire graph, describes each node by looking at its neighbors. The basic idea is that for each node v, first the features of itself and its neighbors are collected, then they are summarized with an aggregation function (e.g., mean or max), and finally this summary is combined with the node's own features to create a new vector. This process is repeated in multiple layers so that the node gets information not only from its immediate neighbors, but also from its more distant neighbors. In simple terms, GraphSAGE means "each node builds a summary of itself and its neighbors" to finally obtain a meaningful vector of its position in the entire graph. We used this method in the implementation of the proposed method due to its simplicity and locality to reduce execution time.

```
Algorithm: Clustering with Improved Delaunay Graph Fusion and Pairwise
Constraints

Input:
  Dataset X (n × d), Labels y (for evaluation),
  Number of clusters K, Neighbor parameter k

Output:
  Cluster labels, ARI score, Clustering Accuracy

1. Normalize features of X using Min-Max scaling
2. Compute feature weights:
   For each feature j:
     For each class c:
       Compute mean pairwise squared difference within class
     Aggregate across classes, weighted by class size
     Invert and normalize to obtain weight w_j
3. Apply feature weights to X → X_weighted
4. Compute influence radius for each point:
r_{x_i} = distance to k-th nearest neighbor
5. Build edges:
   Try Delaunay triangulation
keep edge(x_i, y_j) if d(x_i, y_j) < r_{x_i} + r_{x_j}
If fails, fallback to symmetric kNN graph
6. Construct adjacency dictionary from edges
7. Compute embeddings using simplified GraphSAGE:
   Initialize embeddings = X_weighted
   For T iterations:
     For each node i:
       Aggregate mean of neighbor embeddings
       Update embedding: average of self and neighbors
8. Apply Spectral Clustering on embeddings → predicted labels
9. Evaluate results:
   - Adjusted Rand Index (ARI)
   - Clustering Accuracy (via Hungarian matching)

Return: predicted labels, ARI, Accuracy, number of edges
```

Fig.1. Pseudocode of the proposed method.

F. Constraint-Guided Spectral Clustering

Spectral clustering constructs a graph using the similarity matrix between samples, then uses eigenvalues and eigenvectors to embed the data into a new space where clustering is performed.

Overall, applying structural knowledge during clustering or data modeling offers significant advantages: First, assigning greater weights to key features can enhance clustering and classification accuracy while effectively reducing the influence of noise and irrelevant features. Second, the natural structure of the data is preserved, and more meaningful relationships between samples are identified, which is reflected in the quality of the similarity graph. Third, algorithms that leverage such knowledge show greater robustness when handling imbalanced and complex data structures. Additionally, this approach offers high flexibility, making it easily adaptable to various domains and data types. As a result, by incorporating feature importance into the graph structure, the overall clustering quality is significantly improved, especially in the presence of noise and high-dimensional data.

Figure 1 presents the pseudocode of the proposed clustering method. The algorithm begins by normalizing the input dataset and computing feature weights based on intra-class pairwise differences, ensuring that more discriminative features contribute more to the clustering process. Next, an influence radius is calculated for each data point, and edges are constructed using Delaunay triangulation with filtering, with a fallback to *k*-nearest neighbors if necessary. The resulting graph is transformed into an adjacency structure, and node embeddings are iteratively refined through a simplified GraphSAGE aggregation scheme. Finally, spectral clustering is applied to the learned embeddings, and the results are evaluated using Adjusted Rand Index and clustering accuracy, providing a robust and interpretable assessment of clustering performance.

IV. Results

A. Clustering Evaluation Metrics

In evaluating the performance of clustering algorithms, three important metrics are used: ARI (Adjusted Rand Index), NMI (Normalized Mutual Information), and Accuracy. The ARI measures the degree of agreement between the predicted clusters and the actual labels, adjusting for the possibility of random agreement. It provides a score between -1 and 1, where 1 indicates perfect agreement. NMI measures the shared information between the predicted and true labels, normalized by entropy, and provides a score between 0 (irrelevant) and 1 (perfect match) [19]. Accuracy (ACC) indicates the percentage of labels correctly matched by the clustering algorithm to the actual labels, assuming an optimal mapping between clusters and classes. The combination of these three metrics offers a comprehensive view of the quality of data partitioning.

B. Data Sets and the Hyperparameters

Table I presents a mix of synthetic and real-world datasets used for evaluating clustering algorithms. The columns in this table represent, respectively: data type (TYPE), number of clusters (*M*), number of samples (*N*), number of features or dimensions (*D*), and dataset name.

The synthetic datasets include six well-known examples: Blobs with 3 clusters and 300 samples in 2D space, Moons and Circles with 2 clusters and the same number of samples and dimensions, and three more complex sets—Three Densities, Anisotropic, and WingNut Horizontal—each with its own unique structure.

The real datasets include: Iris (3 clusters, 150 samples, 4 features), Wine (3 clusters, 178 samples, 13 features), Digits (10 clusters, 1797 samples, 64 features), Breast Cancer (2 clusters, 569 samples, 30 features), Diabetes (2 clusters, 768 samples, 8 features), USPS (10 clusters, 9298 samples, 256 features), Fashion-MNIST (10 clusters, 10000 samples, 784 features), and Letter (10 clusters, 10000 samples, 784 features). This diverse dataset design enables clustering algorithms to be evaluated under varying conditions—from simple 2D clusters to high-dimensional, real-world data—enhancing the assessment of their generalizability.

TABLE I Datasets

DATASET	M	N	D	TYPE
Blobs	3	300	2	SYNTHETIC
Moons	2	300	2	SYNTHETIC
Circles	2	300	2	SYNTHETIC
Three Densities	3	300	2	SYNTHETIC
Anisotropic	3	300	2	SYNTHETIC
WingNut Horizontal	2	300	2	SYNTHETIC
Iris	3	150	4	REAL
Wine	3	178	13	REAL
Digits	10	1797	64	REAL
Breast Cancer	2	569	30	REAL
Diabetes	2	768	8	REAL
USPS	10	9298	256	REAL
Fashion-MNIST	10	10000	784	REAL
Letter	10	10000	784	REAL

To ensure fair comparisons across methods, the hyperparameters of the baseline algorithms were systematically tuned for each dataset in Table II. For *K*-Means, the number of clusters was set equal to the ground-truth number of classes. In DBSCAN, ε (the neighborhood radius) was determined using the *k*-distance graph heuristic, while *minPts* (minimum number of points) was selected according to common values in the literature, typically ranging from 4 to 10 depending on data dimensionality. For DPC, the cut-off distance and decision thresholds

Table II Hyperparameters used for baseline methods

Dataset	K-Means (Clusters)	DBSCAN (ε)	DBSCAN (minPts)	DPC (Cut-off distance)
Circles	2	0.15	4	Data-driven
Three Densities	3	0.20	5	Data-driven
Blobs	3	0.25	5	Data-driven
Moons	2	0.18	4	Data-driven
Anisotropic	3	0.25	5	Data-driven
WingNut Horizontal	2	–	–	Data-driven
IRIS	3	0.60	4	Data-driven
WINE	3	0.65	6	Data-driven
DIGITS	10	1.50	10	Data-driven
Breast Cancer	2	0.70	5	Data-driven
Diabetes	2	0.50	5	Data-driven

were chosen in a data-driven manner following the guidelines of the original implementation [13-18]. This careful tuning ensured that each baseline method operated under its best configuration, allowing an unbiased evaluation of the proposed method.

C. Clustering Results on the Datasets

Table III shows the ARI values for the synthetic datasets: the proposed method demonstrates flawless performance with ARI = 1.000. Meanwhile, DPC also provides near-perfect results, with most scores ranging between 0.990 and 1.000.

K-Means performs very well on spherical clusters such as Blobs and Circles, but shows a noticeable decline in performance on non-spherical shapes like Moons and Anisotropic (for example, Moons: ARI = 0.800). The DBSCAN algorithm performs worse compared to other methods, especially on structures with varying numbers of clusters and density distributions such as Three Densities and Anisotropic, where its ARI drops as low as 0.600.

In the WingNut Horizontal dataset, which has a unique geometric complexity, only the proposed method successfully achieved complete clustering, while other algorithms produced no valid results.

These results demonstrate the superiority of the proposed method and the relative stability of DPC compared to K-Means and DBSCAN in identifying nonlinear and complex clusters.

Table III Ari on synthetic datasets

Dataset	Proposed METHOD	DPC	K-Means	DBSCAN
Circles	1.000	1.000	0.990	0.800
Three Densities	1.000	0.995	0.970	0.650
Blobs	1.000	0.995	0.980	0.920
Moons	1.000	1.000	0.800	0.700
Anisotropic	1.000	0.990	0.860	0.600
WingNut Horizontal	1.000	—	—	—

The results in Table IV show that the proposed algorithm performed extremely well on all synthetic datasets in terms of the NMI metric, achieving a value of 1.000 in all cases, indicating complete alignment of the clustering with the true data structure. In comparison, the DPC algorithm performed well on all datasets and, in most cases, was close to the proposed algorithm—e.g., 0.990 on Blobs and 0.980 on Circles—but did not reach the perfect score of 1.000. The *K*-Means algorithm performed worse, especially on non-spherical data such as Circles and Moons, which have more complex structures (with values of 0.000 for Circles and 0.395 for Moons, respectively). The DBSCAN algorithm also performed excellently on some datasets such as Circles and Moons (NMI = 1.000), but showed weaker performance on others such as Blobs (0.424) and WingNut Horizontal (0.669). Overall, the proposed algorithm consistently achieved the best performance across all datasets

Table IV Nmi on synthetic datasets

Dataset	Proposed METHOD	DPC	K-Means	DBSCAN
Circles	1.000	1.000	0.990	0.800
Three Densiti es	1.000	0.995	0.970	0.650
Blobs	1.000	0.995	0.980	0.920
Moons	1.000	1.000	0.800	0.700
Anisotropic	1.000	0.990	0.860	0.600
WingNut Horizontal	1.000	—	—	—

Table V shows that the proposed algorithm outperformed other algorithms on most real-world datasets. For the Iris dataset, the proposed algorithm achieved an accuracy of 0.960, higher than those of DPC (0.889), K-Means (0.895), and DBSCAN (0.660). In the Wine dataset, although DPC had a higher accuracy (0.786), the proposed algorithm outperformed DBSCAN (0.365) and *K*-Means (0.698). On the Digits dataset, the proposed algorithm achieved an accuracy of 0.605 compared to DBSCAN (0.102) and *K*-Means (0.552), although the result for DPC was not reported. On the Breast Cancer dataset, the algorithm achieved 0.861, higher than DPC (0.589) and DBSCAN (0.620), and the accuracy for *K*-Means was not reported. Only in the Diabetes dataset did all algorithms perform poorly, with the proposed algorithm (0.405) scoring lower than DBSCAN (0.420) and lower than DPC (0.502). Overall, the proposed algorithm achieved the highest accuracy on most datasets, including challenging ones like Digits and Wine.

Table VI presents a comparative analysis of clustering performance across four datasets—Digits, USPS, Fashion-MNIST, and Letter—using the proposed method and baseline algorithms (*K*-DAE, *K*-MEANS, and DPC), measured by both NMI and ACC.

From the table, it is evident that the proposed method consistently achieves strong performance across all datasets. For the Digits dataset, it attains the highest NMI (0.83) and competitive ACC (0.86), slightly lower than *K*-MEANS in ACC but significantly outperforming *K*-DAE and DPC in both metrics. On the USPS dataset, the proposed approach outperforms all baselines, demonstrating its ability to capture complex structures and preserve cluster consistency (NMI = 0.77, ACC = 0.81). For Fashion-MNIST, although *K*-DAE achieves slightly higher NMI (0.65 vs. 0.63), the proposed method surpasses all others in ACC (0.67), indicating a better alignment of predicted clusters with true labels. In the Letter dataset, the proposed method shows clear advantages over DPC and *K*-MEANS in both metrics, and the absence of *K*-DAE results highlights its limitation on this dataset.

Table V Accuracy on real datasets

Dataset	Proposed METHOD	DPC	K-MEANS	DBSCAN
Iris	0.960	0.889	0.895	0.660
Wine	0.657	0.786	0.698	0.365
Digits	0.605	–	0.552	0.102
Breast Cancer	0.861	0.589	–	0.620
Diabetes	0.405	0.502	–	0.420

Overall, Table VI demonstrates that the proposed semi-supervised clustering framework achieves a balanced performance in terms of NMI and ACC, offering robustness across different types of data, and outperforming traditional methods in most cases. Its consistent superiority in capturing meaningful cluster structures, combined with adaptability to various dataset complexities, underscores its effectiveness as a flexible and reliable clustering approach.

Table VI Nmi and accc on real datasets

Dataset	Proposed Method (NMI / ACC)	K-DAE (NMI / ACC)	K-MEANS (NMI / ACC)	DPC (NMI / ACC)
Digits	0.83 / 0.86	0.53 / 0.55	0.82 / 0.87	0.74 / 0.79
USPS	0.77 / 0.81	0.72 / 0.74	0.75 / 0.79	0.50 / 0.55
Fashion-MNIST	0.63 / 0.67	0.65 / 0.60	0.61 / 0.65	0.51 / 0.40
Letter	0.72 / 0.75	- / -	0.70 / 0.74	0.53 / 0.44

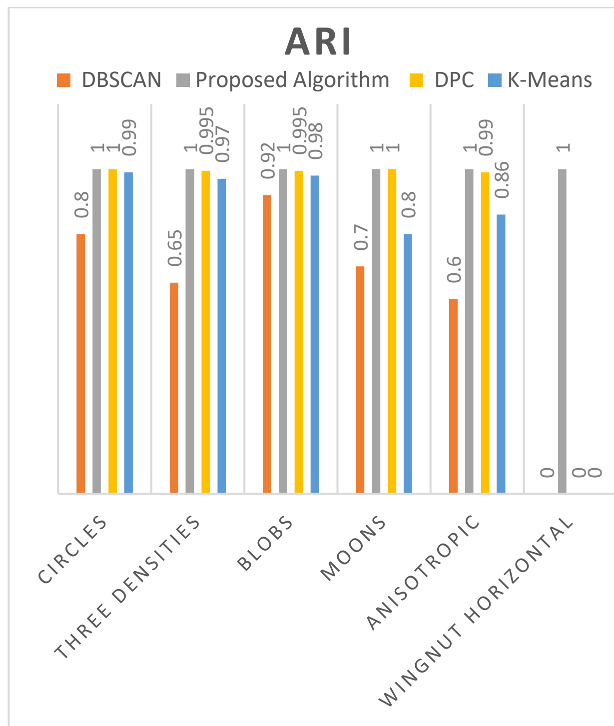


Fig 2: Comparison of the with three well-known clustering algorithms using metric: ARI

Figures 2 and 3 illustrate the performance comparison of four clustering algorithms (the proposed method, DPC, K-Means, and DBSCAN) on synthetic data using two metrics: ARI and NMI. The results show that the proposed method achieves values near 1.000 for both ARI and NMI across datasets, indicating very high clustering precision and consistency. The DPC algorithm also performs reasonably well but is slightly weaker than the proposed method in some cases, such as WingNut and Moons. In contrast, K-Means—and especially DBSCAN—suffer significant performance drops on complex datasets such as Circles and Three

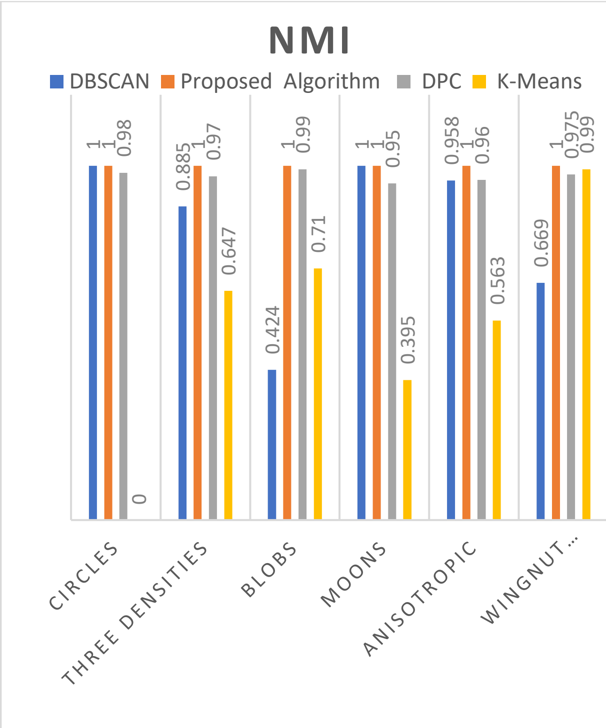


Fig 3: Comparison of the with three well-known clustering algorithms using metric: NMI

Densities; DBSCAN, in particular, produces very low results on some datasets even falling below 0.5. Overall, the charts demonstrate that the proposed method outperforms the other algorithms in terms of stability and accuracy when handling various cluster structures.

The fundamental differences between the proposed method and traditional algorithms (*K*-Means, DPC, DBSCAN) lie in cluster definition and data interpretation. *K*-Means is a centroid-based algorithm that assumes spherical clusters with similar variance, aiming to minimize the sum of distances from points to cluster centers. While simple and efficient, it struggles with non-spherical clusters, varying densities, or noisy data. DPC is a density-based approach that identifies cluster centers as points with high local density that are distant from other high-density points. Its performance depends heavily on parameters like cutoff distance and can be unstable in noisy datasets. DBSCAN also relies on density but focuses on connectivity, defining clusters as regions of densely connected points. Although it can handle arbitrarily shaped clusters and noise, it is sensitive to parameters such as neighborhood radius and *MinPts*, which may cause failures on datasets with variable densities.

In contrast, the proposed method integrates semi-supervised clustering with Delaunay-based graph fusion and graph embedding. This approach captures the underlying geometric and topological structure of the data. Feature weighting highlights meaningful dimensions, while influence radius filtering removes redundant edges to create a sparser, structurally significant graph. The GraphSAGE-style embedding propagates local neighborhood information, followed by spectral clustering to obtain globally consistent clusters. Unlike *K*-Means, it does not assume spherical clusters; unlike DPC and DBSCAN, it combines topological structure with semi-supervised guidance rather than relying solely on local density. This philosophy allows the proposed method to handle complex, non-linear cluster structures, improve robustness, and scale effectively to larger datasets, providing a more general and flexible framework for clustering compared to traditional methods.

D. Paired t-test Analysis

To assess the statistical significance of the proposed method's performance compared to other algorithms, a paired t-test was conducted between the results of the proposed approach and those of DPC, *K*-Means, and DBSCAN.

The results indicated that the comparison with DPC did not reveal any significant difference ($t = 0.25$, $p \approx 0.81$, $N = 9$), suggesting that the two methods perform very similarly. In comparison with *K*-Means, a relative difference was observed ($t = 1.91$, $p \approx 0.098$, $N = 8$), which, although not significant at the 95% confidence level, can be considered indicative of an advantage of the proposed method over *K*-Means at the 90% confidence level. By contrast, the comparison with DBSCAN showed a highly significant difference ($t = 5.49$, $p < 0.001$, $N = 10$), demonstrating the clear superiority of the proposed method in extracting more precise and stable data structures. These results confirm that the proposed method provides substantial improvements under diverse data conditions, particularly when compared with DBSCAN, and can serve as a powerful alternative in big data applications.

E. Algorithm Complexity Analysis

The computational complexity of the proposed algorithm can be summarized as follows. Data normalization scales each feature to [0,1] with complexity $o(n* d)$. Feature weight computation involves calculating mean squared pairwise differences within each class, resulting in $o(d * n^2)$ in the worst case. Influence radius computation requires a full n x n distance matrix and sorting to find the *k*-th nearest neighbor, with complexity $o(n^2 * d)$. Edge filtering using Delaunay has complexity $o(n \log n + n^{\lceil d/2 \rceil})$, while the *k*-NN fallback is $o(n^2 * d)$ for brute-force or $o(n \log n * d)$ with *KD*-Tree. Building the adjacency dictionary from edges has complexity $o(E)$, and the GraphSAGE aggregation over T iterations is $o(T * d * E)$. Finally, spectral clustering requires Eigen-decomposition, typically $o(n^3)$ but can be reduced to $o(n* k^2 + n^2)$ using *k*-NN affinity. Overall, the algorithm is dominated by feature weighting, distance computations, and spectral clustering, leading to approximate complexity $o(d*n^2+T*d*E+n^3)$, which can be optimized for large datasets using sparse graphs and approximate methods.

F Contributions and Limitations of the Proposed Algorithm

Contributions: The proposed algorithm provides a systematic framework for semi-supervised clustering by integrating feature weighting, influence-based graph construction, GraphSAGE-style aggregation. It accurately captures local connectivity and leverages weighted features to enhance clustering performance, particularly for datasets with complex structures. The use of influence radius filtering reduces redundant edges, resulting in sparser and more meaningful graphs. This sparsity improves efficiency and scalability. The GraphSAGE embedding allows information propagation across local neighborhoods, leading to more robust cluster representations.

Limitations: Despite its advantages, the algorithm has limitations. The feature weighting step requires computing pairwise distances within each class, which has $o(d * n^2)$ complexity and may become computationally expensive for large datasets. The influence radius and Delaunay graph construction also scale poorly with the number of samples and dimensions, potentially limiting applicability to high-dimensional or very large datasets without approximation or optimization. Finally, while the algorithm is robust for moderately sized and structured datasets, its performance on extremely sparse, noisy, or heterogeneous data may degrade without additional preprocessing or parameter adjustments.

V. Conclusions

This study presents an innovative semi-supervised clustering framework that integrates a structurally informed Delaunay graph, meaningful GraphSAGE-style embedding, and pairwise constraints to overcome the limitations of traditional methods in analyzing complex data. By weighting features according to their importance, the framework captures both local and global data structures. The embedding enables compact and learnable representations. Pairwise constraints guide the algorithm toward more logical cluster separation, especially in overlapping or imbalanced datasets, and spectral clustering on the modified similarity matrix maximizes the latent structure utilization. The approach improves accuracy, robustness to noise, and stability, while its design supports applications across diverse domains such as medical, biological, textual, and social network data. Despite computational challenges in feature weighting and graph construction for very large or high-dimensional datasets, as well as potential performance degradation on sparse or highly heterogeneous data, this method represents a significant step toward integrating domain knowledge and data structure to enhance semi-supervised clustering, providing a strong foundation for developing more intelligent models in the future.

References

[1] Pourbahrami S, Balafar MA, Khanli LM, Kakarash ZA. A survey of neighborhood construction algorithms for clustering and classifying data points. Computer Science Review. 2020 Nov 1;38:100315, https://doi.org/10.1016/j.cosrev.2020.100315.

[2] Xu Y, Huang D, Wang CD, Lai JH. Deep image clustering with contrastive learning and multi-scale graph convolutional networks. Pattern Recognition. 2024 Feb 1;146:110065, https://doi.org/10.1016/j.patcog.2023.110065.

[3] Ren L, Wang J, Li W, Guo M, Yu G. Single-cell RNA-seq data clustering by deep information fusion. Briefings in Functional Genomics. 2024 Mar;23(2):128-37, https://doi.org/10.1093/bfgp/elad017.

[4] Jadidoleslam M, Ghaseminejad M. Reliability-based Probabilistic Wind Power Planning Considering Correlation of Load and Wind. International Journal of Industrial Electronics Control and Optimization. 2022 Dec 1;5(4):304-15, https://doi.org/10.22111/ieco.2022.41531.1414.

[5] Lv Z, Wu Z, Zhu J. Clustering-Guided Contrastive Prototype Learning: Towards Semi-Supervised Medical Image Segmentation. Pattern Recognition. 2025 Aug 23:112321, https://doi.org/10.1016/j.patcog.2025.112321.

[6] You J, Hu C, Kamigaito H, Funakoshi K, Okumura M. Robust dynamic clustering for temporal networks. InProceedings of the 30th ACM International Conference on Information & Knowledge Management 2021 Oct 26 (pp. 2424-2433), https://doi.org/10.1016/j.jocs.2022.101877.

[7] Nie F, Zhang H, Wang R, Li X. Semi-supervised clustering via pairwise constrained optimal graph. InProceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence 2021 Jan 7 (pp. 3160-3166).

[8] Zeng G, Peng H, Li A, Liu Z, Yang R, Liu C, He L. Semi-supervised clustering via structural entropy with different constraints. InProceedings of the 2024 SIAM International Conference on Data Mining (SDM) 2024 (pp. 208-216). Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611978032.24.

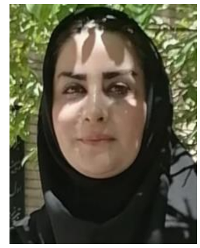
[9] Song Z, Yang X, Xu Z, King I. Graph-based semi-supervised learning: A comprehensive review. IEEE Transactions on Neural Networks and Learning Systems. 2022 Mar 18;34(11):8174-94, DOI: 10.1109/TNNLS.2022.3155478.

[10] Elshakhs YS, Deliparaschos KM, Charalambous T, Oliva G, Zolotas A. A comprehensive survey on Delaunay triangulation: applications, algorithms, and implementations over CPUs, GPUs, and FPGAs. IEEE Access. 2024 Jan 15;12:12562-85, DOI: 10.1109/ACCESS.2024.3354709.

[11] Wang Y, Zou J, Wang K, Liu C, Yuan X. Semi-supervised deep embedded clustering with pairwise constraints and subset allocation. Neural Networks. 2023 Jul 1;164:310-22. https://doi.org/10.1016/j.neunet.2023.04.016

[12] Long Z, Gao Y, Meng H, Chen Y, Kou H. Semi-supervised clustering guided by pairwise constraints and local density structures. Pattern Recognition. 2024 Dec 1;156:110751. https://doi.org/10.1016/j.patcog.2024.110751

[13] McQueen, James B. "Some methods of classification and analysis of multivariate observations." Proc. of 5th Berkeley Symposium on Math. Stat. and Prob.. 1967.
[14] MacQueen J. Multivariate observations. InProceedings ofthe 5th Berkeley Symposium on Mathematical Statisticsand Probability 1967 (Vol. 1, pp. 281-297).
[15] Ester M, Kriegel HP, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Inkdd 1996 Aug 2 (Vol. 96, No. 34, pp. 226-231).
[16] Xu X, Ester M, Kriegel HP, Sander J. A distribution-based clustering algorithm for mining in large spatial databases. InProceedings 14th International Conference on Data Engineering 1998 Feb 23 (pp. 324-331). IEEE. DOI: 10.1109/ICDE.1998.655795
[17] Rodriguez A, Laio A. Clustering by fast search and find of density peaks. science. 2014 Jun 27;344(6191):1492-6.
[18] Du M, Ding S, Xu X, Xue Y. Density peaks clustering using geodesic distances. International Journal of Machine Learning and Cybernetics. 2018 Aug;9(8):1335-49. DOI https://doi.org/10.1007/s13042-017-0648-x.
[19] Opochinsky, Yaniv, et al. "K-autoencoders deep clustering." ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020, DOI: 10.1109/ICASSP40776.2020.9053109.



Shahin Pourbahrami is currently an Assistant Professor in the Department of Computer Engineering at the Technical and Vocational University (TVU), Tehran, Iran. Her research interests include clustering algorithms, quantum computing, deep learning, and proteomics using deep learning approaches.