CONTENTS

ABSTRACTS

Behrouz Najjari, Mohammad J. Fotuhi, Mousa Vaezipour
Theoretical and Empirical Improvement of a Fast-Switching Electro-Pneumatic Valve by using Different Methods

In this paper, a non-linear model of a 2–2 way, on–off fast-switching valve is used. The model includes subsystems of electrical, magnetic, mechanical and fluid. Pulse width modulation (PWM) technique is adopted to energise the on–off solenoid valve and allow the air to flow towards the actuator. Since the non-linear behaviour of valve is of great importance, to reduce the delay in performance of switching valves, different approaches are proposed. Furthermore, hysteresis, proportional integrator (PI), optimal model predictive and fuzzy logic controller (FLC) are used and compared. Also, to improve the valve behaviour, an empirical setup based on AVR microcontroller with FLC is implemented. Empirical and simulation results indicate that all proposed control methods have superior performance. However, the fuzzy method is easy to implement in practice.

Shahida Rehman, Akhtar Hussain, Jamshaid Ul Rahman, Naveed Anjum, Taj Munir
Modified Laplace-Based Variational Iteration Method for the Mechanical Vibrations and its Applications

In this paper, we are putting forward the periodic solution of non-linear oscillators by means of variational iterative method (VIM) using Laplace transform. Here, we present a comparative study of the new technique based on Laplace transform and the previous techniques of maximum minimum approach (MMA) and amplitude frequency formulation (AFF) for the analytical results. For the non-linear oscillators, MMA, AFF and VIM by Laplace transform give the same analytical results. Comparison of analytical results of VIM by Laplace transform with numerical results by fourth-order Runge–Kutta (RK) method conforms the soundness of the method for solving non-linear oscillators as well as for the time and boundary conditions of the non-linear oscillators.

Andrzej Werner
Method for Enhanced Accuracy in Machining Free-Form Surfaces on CNC Milling Machines

The present article describes a method for enhanced accuracy in machining free-form surfaces produced on CNC milling machines. In this method, surface patch machining programs are generated based on their nominal CAD model. After the pretreatment, coordinate control measurements are carried out. The obtained results of the measurements contain information on the values and distribution of observed machining deviations. These data, after appropriate processing, are used to build a corrected CAD model of the surface produced. This model, made using reverse engineering techniques, compensates for the observed machining deviations. After regeneration of machining programs, the object processing and control measurements are repeated. As a result of the conducted procedure, the accuracy of the manufacture of the surface object is increased. This article also proposes the introduction of a simple procedure for the filtration of measurement data. Its purpose is to minimise the effect of random phenomena on the final machining error correction. The final part of the article presents the effects of the proposed method of increasing the accuracy of manufacturing on ‘raw’ and filtered measurement data. In both cases, a significant improvement in the accuracy of the machining process was achieved, with better final results obtained from the filtered measurement data. The method proposed in the article has been verified for three-axis machining with a ball-end cutter.

Mohamed Ramla, Houssem Laidoudi, Mohamed Bouzit
Behaviour of a Non-Newtonian Fluid in a Helical Tube under the Influence of Thermal

This work is an evaluative study of heat transfer in the helical-type heat exchanger. The fluid used is non-Newtonian in nature and is defined by Oswald’s model. The work was performed numerically by solving each of the Navier–Stokes equations and the energy equation using the package ANSYS-CFX. Following are the aspects that have been dealt with in this paper: the effects of thermal buoyancy, fluid nature and the tube shape on the heat transfer, and the fluid comportment. The interpretation of the obtained results was done by analyzing the isotherms and the streamlines. The mean values of the Nusselt number were also obtained in terms of the studied parameters. The results of this research enabled us to arrive at the following conclusion: the intensity of thermal buoyancy and the nature of the fluid affect the heat transfer distribution but keep the overall rate of heat transfer the same.

Ahmad Hanan, Tariq Feroze, Awais Arif, Hasan Iftikhar, Afzaal Ahmed Khan, Sarmad Javaid
Performance Evaluation of a Single Cylinder Compressed Air Engine: An Experimental Study

The quest to reduce dangerous environmental emissions has led to the research and use of alternate and renewable energy sources. One of the major contributors to the dangerous environmental emissions is the automotive industry. The world is, therefore, quickly moving towards hybrid and electric vehicles. An alternate pollution-free automotive engine is a compressed-air engine, which is powered by compressed air and is more efficient than the electric engine since it requires less charging time than a traditional battery-operated engine. Furthermore, the tanks used in compressed-air engines have a longer lifespan in comparison to the batteries used in electric vehicles. However, extensive research is required to make this engine viable for commercial use. The current study is a step forward in this direction and shows the performance analysis of a single-cylinder compressed-air engine, developed from a four-stroke, single-cylinder, 70 cc gasoline engine. The results show that compressed-air engines are economic, environmental friendly and efficient.

Agustinus Winarno, Benidiktus T. Prayoga, Ignatius A. Hendaryanto
Linear Motion Error Evaluation of Open-Loop CNC Milling Using a Laser Interferometer

The usage of computerised numerical control (CNC) machines requires accuracy verification to ensure the high accuracy of the processed products. This paper introduces an accuracy verification method of an open-loop CNC milling machine using a fringe counting of He–Ne laser interferometry to evaluate the best possible accuracy and functionality. The linear motion accuracy of open-loop CNC milling was evaluated based on the number of pulses from the controller against the actual displacement measured by the He–Ne fringe-counting method. Interval distances between two pulses are also precisely measured using the He–Ne interferometry. The linear motion error and controller error can be simultaneously evaluated in sub-micro accuracy. The linear positioning error due to the micro-stepping driver accuracy of the mini-CNC milling machine was measured with the expanded uncertainty of measurement and was estimated at 240 nm. The experimental results show that linear motion error of the open-loop CNC milling can reach up to 50 µm for 200 mm translation length.

Julius Niyongabo, Yingjie Zhang, Jérémie Ndikumagenge
Bearing Fault Detection and Diagnosis Based on Densely Connected Convolutional Networks

Rotating machines are widely used in today’s world. As these machines perform the biggest tasks in industries, faults are naturally observed on their components. For most rotating machines such as wind turbine, bearing is one of critical components. To reduce failure rate and increase working life of rotating machinery it is important to detect and diagnose early faults in this most vulnerable part. In the recent past, technologies based on computational intelligence, including machine learning (ML) and deep learning (DL), have been efficiently used for detection and diagnosis of bearing faults. However, DL algorithms are being increasingly favoured day by day because of their advantages of automatically extracting features from training data. Despite this, in DL, adding neural layers reduces the training accuracy and the vanishing gradient problem arises. DL algorithms based on convolutional neural networks (CNN) such as DenseNet have proved to be quite efficient in solving this kind of problem. In this paper, a transfer learning consisting of fine-tuning DenseNet-121 top layers is proposed to make this classifier more robust and efficient. Then, a new intelligent model inspired by DenseNet-121 is designed and used for detecting and diagnosing bearing faults. Continuous wavelet transform is applied to enhance the dataset. Experimental results obtained from analyses employing the Case Western Reserve University (CWRU) bearing dataset show that the proposed model has higher diagnostic performance, with 98% average accuracy and less complexity.

Michał Kolankowski, Robert Piotrowski
Design of Three Control Algorithms for an Averaging Tank with Variable Filling

An averaging tank with variable filling is a nonlinear multidimensional system and can thus be considered a complex control sys-tem. General control objectives of such object include ensuring stability, zero steady-state error, and achieving simultaneously shortest possible settling time and minimal overshoot. The main purpose of this research work was the modeling and synthesis of three control systems for an averaging tank. In order to achieve the intended purpose, in the first step, a mathematical model of the control system was derived. The model was adapted to the form required to design two out of three planned control systems by linearization and reduction of its dimensions, resulting in two system variants. A multivariable proportional-integral-derivative (PID) control system for the averaging tank was developed using optimization for tuning PID controllers. State feedback and output feedback with an integral action control system for the considered control system was designed using a linear-quadratic regulator (LQR) and optimization of weights. A fuzzy control system was designed using the Mamdani inference system. The developed control systems were tested using theMATLAB environment. Finally, the simulation results for each control algorithm (and their variants) were compared and their performance was assessed, as well as the effects of optimization in the case of PID and integral control (IC) systems.

Piotr Jankowski
On the Nonlocal Interaction Range for Stability Of Nanobeams with Nonlinear Distribution of Material Properties

The present study analyses the range of nonlocal parameters' interaction on the buckling behaviour of nanobeam. The intelligent nonhomogeneous nanobeam is modelled as a symmetric functionally graded (FG) core with porosity cause nonlinear distribution of material parameters. The orthotropic face-sheets are made of piezoelectric materials. These kinds of structures are widely used in nanoelectromechanical systems (NEMS). The nanostructure model satisfies the assumptions of Reddy third-order beam theory and higher-order nonlocal elasticity and strain gradient theory. This approach allows to predict appropriate mechanical response of the nanobeam regardless of thin or thick structure, in addition to including nano-sized effects as hardening and softening. The analysis provided in the present study focuses on differences in results for nanobeam stability obtained based on classical and nonlocal theories. The study includes the effect of diverse size-dependent parameters, nanobeams' length-to-thickness ratio and distributions of porosity and material properties through the core thickness as well as external electro-mechanical loading. The results show a dependence of nonlocal interaction range on geometrical and material parameters of nanobeam. The investigation undertaken in the present study provides an interpretation for this phenomenon, and thus aids in increasing awareness of nanoscale structures' mechanical behaviour.

Mohamed F. Abd Alsamieh
Numerical Study of Transient Elastohydrodynamic Lubrication Subjected to Sinusoidal Dynamic Loads for Rough Contact Surfaces

The purpose of this paper is to study the behaviour of transient elastohydrodynamic contacts subjected to forced harmonic vibrations, including the effect of surface waviness for concentrated counterformal point contact under isothermal conditions. Profiles of pressure and film thickness are studied to reveal the combined effects of sinusoidal external load and surface roughness on the lubrication problem. The time-dependent Reynolds' equation is solved using Newton–Raphson technique. The film thickness and pressure distribution are obtained at different snap shots of time by simultaneous solution of the Reynolds' equation and film thickness equation including elastic deformation and surface waviness. It is concluded that the coupling effects of the transient sinusoidal external load and wavy surface would result in increase in modulations of the pressure and film thickness profile in comparison to the case where the smooth contact surfaces are subjected to sinusoidal external load.

DOI 10.2478/ama-2022-0011 acta mechanica et automatica, vol.16 no.2 (2022)

THEORETICAL AND EMPIRICAL IMPROVEMENT OF A FAST-SWITCHING ELECTRO-PNEUMATIC VALVE BY USING DIFFERENT METHODS

Behrouz NAJJARI*, **Mohammad J. FOTUHI****, **Mousa VAEZIPOUR*****

*Department of Mechanical Engineering, University of Applied Science & Technology of IDEM, Tabriz , P.O. Box: 51335 – 4688, Iran
**Automation and Robotics Lab, Department of Mechatronics Engineering, Kocaeli University, Kabaoğlu, Baki Komsuoğlu Bulvarı No:515, Umuttepe, 41001 İzmit/Kocaeli, Turkey
***Department of Mechanical Engineering, Technical and Vocational University of Tabriz, 28QW+JRF Vila Shahr, Tabriz, East Azerbaijan Province, Iran

b.najjari@idem.ir, mohammad.fotuhi@kocaeli.edu.tr, vaezimosa@gmail.com

received 11 October 2021, revised 8 December 2021, accepted 12 December 2021

Abstract: In this paper, a non-linear model of a 2–2 way, on–off fast-switching valve is used. The model includes subsystems of electrical, magnetic, mechanical and fluid. Pulse width modulation (PWM) technique is adopted to energise the on–off solenoid valve and allow the air to flow towards the actuator. Since the non-linear behaviour of valve is of great importance, to reduce the delay in performance of switching valves, different approaches are proposed. Furthermore, hysteresis, proportional integrator (PI), optimal model predictive and fuzzy logic controller (FLC) are used and compared. Also, to improve the valve behaviour, an empirical setup based on AVR microcontroller with FLC is implemented. Empirical and simulation results indicate that all proposed control methods have superior performance. However, the fuzzy method is easy to implement in practice.

Key words: on–off valve, electro-pneumatic, PWM, empirical, AVR microcontroller

1. INTRODUCTION

Since pneumatic control systems are of low cost, easy to maintain and cheap, they are used in industrial automation systems [1,2]. However, pneumatic actuators are highly non-linear, which results from air friction and the delayed time during air compression. Hence, electro-pneumatic actuators are preferred to the pneumatic actuators. There are two major types of electro-pneumatic valves that are used for conducting air, including on–off and servo valves [3]. Servo valve offers high precision and linear behaviour, but it has a complex and also expensive structure. Fast-switching on–off valve has a cost-effective and simple structure with intrinsically non-linear behaviour [1,4]. Pulse width modulation (PWM) method is adopted to attain a linear behaviour by using fast-switching valves.

For decreasing the complexity and costs of the systems, an on–off valve, which is driven by PWM signal, is used instead of servo valve [5]. By using the PWM approach, the valve is opened and closed periodically, so the air is allowed to flow towards the actuator to control the signal [6,7]. Many studies have utilised PWM signal for control applications, but a very early application uses PWM approach to control the fluid system [8]. Furthermore, Noritsu employed PWM approach for controlling the position and velocity of pneumatic piston [9,10]. Muto [11] used differential PWM approach for controlling hydraulic actuators.

Many researchers have conducted studies on modelling and control of electro-pneumatic actuators by using on–off fast-switching valve [12-15], but these studies have scarcely focused on on–off fast solenoid valve. In [16], a dynamic non-linear model of a 3-2 on–off valve is presented. The valve performance in terms of opening and closing was improved by using the proper control approach [17]. Wang et al.[18] modified a ferromagnetic material (Al–Fe) to have a strong magnetic force to achieve fast response. Other investigations into the characteristics of fast-switching on–off valves are given in [19,20].

There are different approaches for controlling electro-pneumatic systems [21-24]. Since the non-linear behaviour of the valve leads to delay in the performance, to compensate this defect, theoretical and practical controlling of valve behaviour has major importance, but this point has been rarely studied by researchers. Also, the number of valves opening and closing affect the lifetime of the on–off valves. Therefore, reducing the valve frequency and switching time delay have a great importance. Hence, in this paper, different control methods including hysteresis, proportional integrator (PI), optimal controller based on predictive approach and fuzzy logic controller (FLC) are proposed to improve the non-linear behaviour of valve. Also, in order to verify the theoretical results, an empirical setup including alf and vegards risc (AVR) microcontroller based on FLC is designed. It should be mentioned that in many industries, Programmable logic controllers (PLCs) are employed for controlling electro-pneumatic systems. Because PLC is not a cost-effective method, in the empirical setup, an electronic circuit based on AVR microcontroller is used.

The remaining of the paper is as follows: In section 2, a non-linear model of a 2–2 way valve is presented. Then, evaluation of the solenoid valve model is shown. Section 3 is dedicated to

DOI 10.2478/ama-2022-0011

controller design for valve with PWM technique. The designs of hysteresis approach, PI controller, optimal controller based on predictive approach and FLC are also presented. In the next section, an empirical setup is presented to express the effectiveness of controllers. The conclusion is explained in the last section.

2. MODELLING OF AN ELECTRO-PNEUMATIC ON–OFF FAST-SWITCHING VALVE

As shown in Fig. 1, in this paper, for simulating the dynamic behaviour of valve equipped with designed control systems, the model of valve including four subsystems consisting of electrical, magnetic, mechanical and fluid is employed. More detailed information about the model is available in [2].

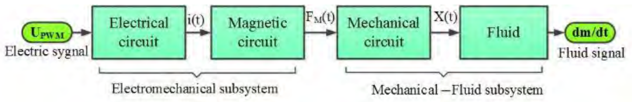


Fig. 1. Valve model components' subsystems

Now, for validation of the obtained model, the periodic input with duty cycle 50% is applied to the valve (Fig. 2). Results show the duration when the voltage input is applied; position behaviour is trivially analogous to dc input till the input is set to zero. As a result, both current and the position fall. The delay in position response originates from the practical model of the valve. Position reduction continues until the plunger goes backwards to the path beginning. Hence, it is observed that the modelled valve follows the input successfully.

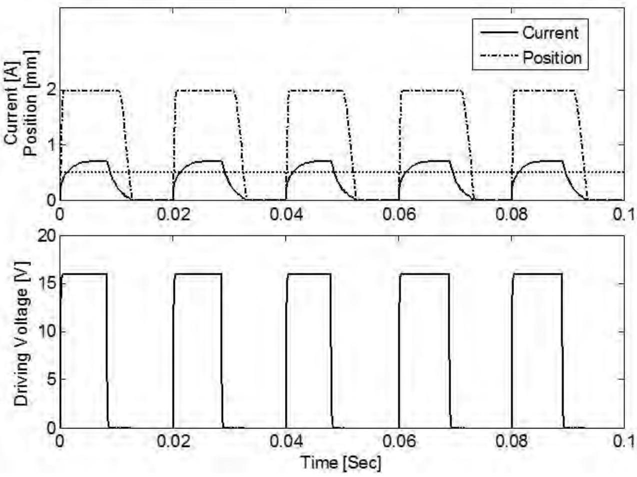


Fig. 2. Evaluation of the solenoid valve

3. DESIGN OF DIFFERENT CONTROLLERS FOR IMPROVING THE VALVE BEHAVIOUR

The valve is proposed to follow the performance as [16]; since the input voltage is applied to the coil at $t = 0$ sec, the current rises exponentially as long as the current reaches the desired amount, and then, due to the magnetic force applied to the spring force, the plunger of valve starts sliding. Since the system has a delay, the plunger of valve takes a certain time to fully open. The coil needs current, which is called holding current, to keep the valve fully open. Also, for opening the valve, the current required is called switching current. In practice, the holding current is less than the switching current [16]. In order to prevent the current from increasing after opening the valve and keeping it as holding current, several controllers are proposed and compared in the following sections. Also, for closing the valve, an electrical current should be fetched from the coil. Therefore, for reducing the time delay to close the on–off valve, using a first-order low-pass filter, a negative voltage can be applied into the coil as follows:

$$H(s) = \frac{1}{\tau s+1} \tag{1}$$

3.1. Hysteresis controller

In this section, in order to prevent the current from increasing after opening the valve, the hysteresis controller is designed in which the output is changed when the input exceeds its bound. In this controller, the signal is saturated as follows [2]:

$$\begin{cases} 1 & \text{if} \quad i_a \le i_d + \frac{H}{2} \\ 0 & \text{if} \quad i_a \ge i_d - \frac{H}{2} \\ \text{no change} & \text{if} \quad |i_a| < \frac{H}{2} \end{cases} \tag{2}$$

in which i_d and i_a are the desired and measured current, respectively, and G and H denote the output signal and the hysteresis bounds of the coil, respectively.

As seen in Fig. 3, this controller holds the current in a desired value; but due to the fluctuating nature of this method, the hysteresis controller has inevitable shortcoming and it reduces the lifetime of the valve.

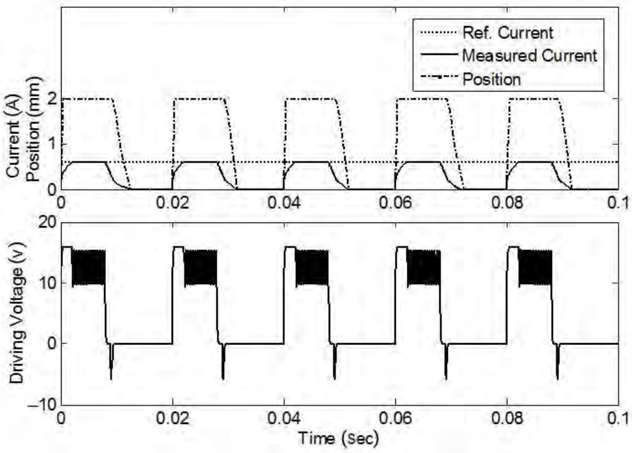


Fig. 3. Response of hysteresis controller

3.2. PI controller

As mentioned before, since the hysteresis controller has a fluctuating nature and it reduces the lifetime of valve, in this section, in order to enhance the valve behaviour during opening and closing, the PI controller, which is common in pneumatic control systems, is designed [16]. As seen in Fig. 4, this controller holds the coil current to the desired value after the valve is fully opened. In comparison with hysteresis controller, this controller holds the current of coil without frequent fluctuations, which is an important advantage of this controller.

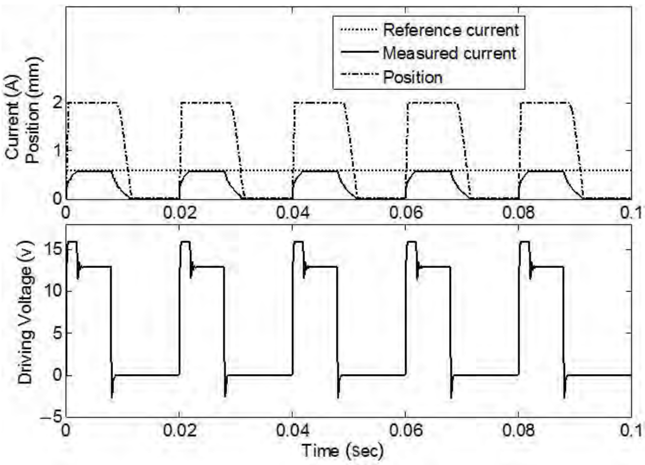


Fig. 4. Response of PI controller

3.3. Nonlinear optimal controller based on predictive approach

In this section, a model predictive-based controller is optimally designed. A 2- degree of freedom (DOF) model containing the main dynamics of solenoid is employed to calculate the voltage. The state variables of this model include the main and auxiliary ring current (I_1 and I_2). The governing equations are expressed as

$$\dot{I}_1 = c_e \left[\frac{V_1 - RI_1}{N_1} + \frac{R_2 B I_2}{D N_2}\right] \quad (3)$$

$$\dot{I}_2 = \frac{V_1 - RI_1}{N_1 B} - \frac{A c_e}{B}\left[\frac{V_1 - RI_1}{N_1} + \frac{R_2 B I_2}{D N_2}\right] \quad (4)$$

V_1 is the driving voltage of solenoid. The state space form of Eqs (3) and (4) is rewritten as

$$\dot{x}_1 = g_1 + \frac{1}{c_e N_1} V_1 \quad (5)$$

$$\dot{x}_2 = g_2 + \frac{1 - A c_e}{B N_1} V_1 \quad (6)$$

in which $X = [x_1\ x_2]T = [I_1\ I_2]T$ and V_1 are the state vector and the control signal, respectively. According to this method, a point-wise cost function that minimises the current control expenditure and the next current tracking error is defined as follows:

$$J = \frac{1}{2}\ w_I\ [I_1 - I_d]^2\ + \frac{1}{2}\ w_v V_1^2 \quad (7)$$

where, h is the prediction period and I_d is the referenced or desired response of current. Also, $w_I > 0$ and $w_v \geq 0$ are the weighting factors of current and control input, respectively. The Taylor series expansion at time t is used to expand the term $I_1(t + h)$ in Eq. (7) as follows:

$$I_1(\mathrm{t+h}) = I_1(\mathrm{t}) + h\dot{I}_1(\mathrm{t}) + \frac{h^2}{2!}\ddot{I}_1(\mathrm{t}) + \cdots + \frac{h^n}{n!}\ I_1^{(n)}(\mathrm{t}) \quad (8)$$

To prevent complexity in deriving and implementing the controller, the order of expansion in Taylor series is restricted to be equal with the relative degree of current in the non-linear system [31,32].

This selection, which is related to zero control order, removes the derivatives of the control input for predicting the control variables, which ends up in proper performance for nonlinear system with a small relative degree. Referring to Eqs (5) and (6), the relative degree of current is $\rho_1 = 1$, and therefore, the first-order Taylor series is enough to expand the current:

$$I_1(\mathrm{t+h}) = I_1(\mathrm{t}) + h\dot{I}_1(\mathrm{t}) = I_1(\mathrm{t}) + h\left[g_1 + \frac{1}{c_e N_1} V_1\right] \quad (9)$$

Substituting Eq. (9) into the cost function in Eq. (7) gives

$$J = \frac{1}{2} w_I\ \left[e_I + h\left(g_1 + \frac{1}{c_e N_1} V_1\right)\right]^2 + \frac{1}{2}\ w_v V_1^2 \quad (10)$$

where, $e_I(t) = I(t) - I_d$ is the current tracking error. The reference or desired response of current, I_d, is constant.

By minimising the cost function in Eq. (10) with respect to the current control input, the optimal control law for V_1 is achievable. The necessary optimisation condition is applied as follows:

$$\frac{\partial J}{\partial V_1} = 0\ \rightarrow\ V_1 = \frac{-c_e N_1}{h^2 + \frac{w_v c_e^2 N_1^2}{w_I\ h^2}}\ [e + h g_1] \quad (11)$$

The following theorem is given to prove the stability of the proposed integrated controller (Eq. (11)):

Theorem 1. The current dynamics in the presence of uncertainties is stable in the sense of Lyapunov under the control law in Eq. (11) with non-zero weights. Also, there exists a small predictive time h, so that the error converges to a compact set.

Proof 1. The current error dynamics in the closed-loop system is obtained by inserting the control law in Eq. (11) in the model of Eq. (5) as

$$\dot{I}_1 = g_1 - \frac{k}{h}\ [e_I + h\hat{g}_1] \quad (12)$$

where

$$K = \frac{h}{h^2 + \frac{w_v c_e^2 N_1^2}{w_I\ h^2}} \quad (13)$$

The symbol (^) indicates the nominal model used for the controller design. Eq.(12) is rewritten to derive the error dynamics as

$$\dot{e}_I + \frac{k}{h} e_I = (g_1 - \hat{g}_1) + (1 - k)\hat{g}_1 \quad (14)$$

Deviation of g_1 from $\hat{g}_1$ may be due to model uncertainties and other errors in measurement or estimation. The function $\hat{g}_1$, including the internal current, is bounded. Therefore, the positive constants $\Gamma_1 > 0$ and $\Gamma_2 > 0$ exist, so that

$$|g_1 - \hat{g}_1| \leq \Gamma_1 \qquad |\hat{g}_1| \leq \Gamma_2 \quad (15)$$

The error dynamic in Eq. (14) is rewritten by using the bounds of Eq. (15) as

$$\dot{e}_I + \frac{k}{h} e_I \leq \Gamma_1 + (1 - k)\Gamma_2 \quad (16)$$

The Lyapunov candidate can be written as follows:

$$V = \frac{1}{2} e_I^2 \quad (17)$$

Substituting Eq. (14) into the derivative of Eq. (17) gives

$$\dot{V} \leq -\frac{k}{h} e_I^2 + [\Gamma_1 + (1 - k)\Gamma_2]|e_I| \quad (18)$$

Utilising the inequality $ab \leq ma2 + b2/4m$ for any real a, b and $m > 0$ gives the following:

$$\dot{V} \leq -\frac{k}{h} e_I^2 + \frac{k}{4h} + \frac{1}{k}[\Gamma_1 + (1 - k)\Gamma_2]^2 \leq -\frac{3k}{4h} V + \frac{k}{h}[\Gamma_1 + (1 - k)\Gamma_2]^2 \quad (19)$$

DOI 10.2478/ama-2022-0011

where $m = \kappa/4h$. The inequality in Eq. (19) is solved by using the comparison lemma [31] as follows:

$$V = \frac{1}{2}e_I^2 \le \left[V(0) - \frac{2h^2}{3k^2}[\Gamma_1 + (1-k)\Gamma_2]^2\right] e_I^{-\frac{3k}{2h}t} + \frac{2h^2}{3k^2}[\Gamma_1 + (1-k)\Gamma_2]^2 \tag{20}$$

It is clear that the current tracking error converges to the compact set $|e_I| \le \frac{2}{\sqrt{3}}[\Gamma_1 + (1-k)\Gamma_2]\frac{h}{k}$. So, it is uniformly bounded for all times. For any given ε > 0, it can be chosen $0 < h < \frac{\sqrt{3}\varepsilon}{2[\Gamma_1+(1-k)\Gamma_2]}$ in the control law, Therefore, e_I converges to $|e_I| \le \varepsilon$. So that, the stability of the control law in the sense of Lyapunov is proved.

Simulation results of using optimal model predictive control are shown in Fig. 5. As seen in this figure, the coil current tracks the desired value appropriately. In comparison with hysteresis controller, this controller holds the current of coil without frequent fluctuations, which is an important advantage of this controller. Also, the proposed controller which uses the continuous dynamic model is analytically calculated in the closed form and does not need online optimisation in calculation of the control input. However, the optimal model predictive controller is fast and easy to solve and implement in real times.

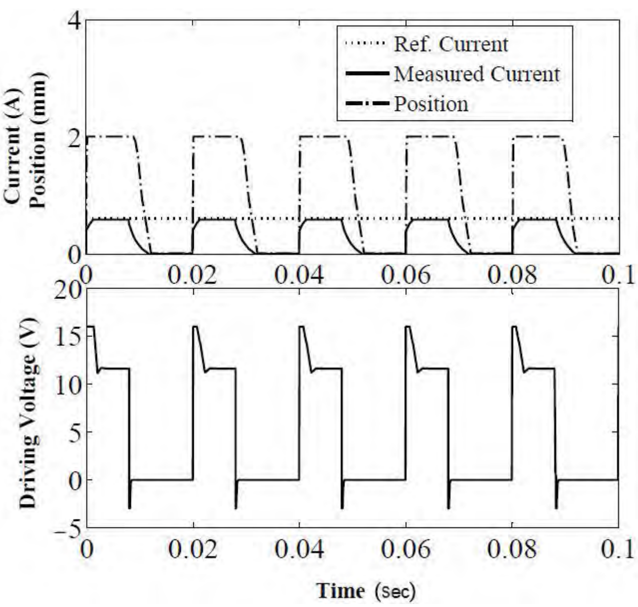


Fig. 5. Response of optimal model predictive controller

3.4. Fuzzy logic controller

In this section, the FLC is used for controlling the system current with negligible time delay. It should be mentioned that in empirical design, FLC is preferred over other controllers. In the first step of designing FLC, rules are designed for the current error and the rate of current error to get FLC response for a desired set point of bobbin current and keep it constant.

Fig. 3 shows the block diagram of FLC. In this figure, the stretched fuzzy controller provides variable voltage signal for control of bobbin current. As seen in Fig. 3, the feedforward structure is employed to compensate the measured disturbances before they have any effect on the system output. In the bobbin fuzzy control system, the current gain (Cg) is combined with an FLC to stabilise the bobbin current by tracking the input efficiently without considering external errors.

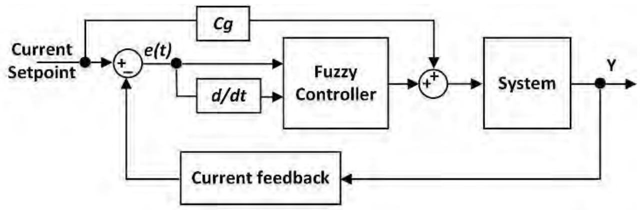


Fig. 6. Block diagram of FLC

In this paper, the FLC is designed based on Mamdani type [2]. As seen in Table 1, in the FLC, linguistic variables and rules are designed for the error and the rate of error to get an FLC response for the desired set point of the bobbin current. For instance, when the bobbin current moves down, the voltage of the bobbin increases gradually and the error decreases; on the other hand, when the bobbin current moves up, bobbin voltage decreases too.

Tab. 1. FLC rules

Error rate	Error				
	NL	**NM**	**ZE**	**PM**	**PL**
NL	VL	L	L	M	M
NM	L	L	M	S	S
ZE	L	M	S	S	VS
PM	M	M	VS	VS	VS
PL	S	S	VS	VS	VS

The linguistic variables including NL, NM, ZE, PM and PL represent negative low, negative medium, zero, positive medium and positive large, respectively.

Fig. 7 shows the general structure of the fuzzy controller consisting of fuzzification (functions that convert an explicit input to a fuzzy input) and inteference unit which is based on fuzzy rules and defuzzification (functions that convert fuzzy output to explicit output).

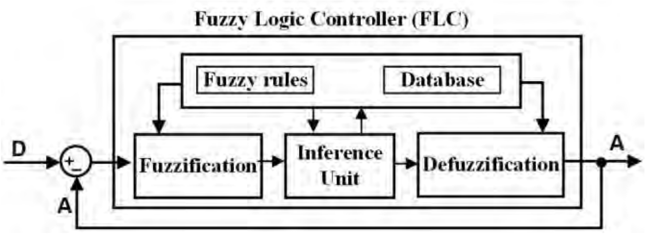


Fig. 7. General structure of FLC

FLC regulates the current by using the PWM signals in the bobbin driver. The error and the rate of error are the inputs of FLC. The error is determined as the difference between the measured and desired system current as:

$$E = A - D \tag{21}$$

where D and A are the desired and actual current, respectively.

The FLC's membership functions are shown in Fig. 8. Also, the control surface is indicated in Fig. 9.

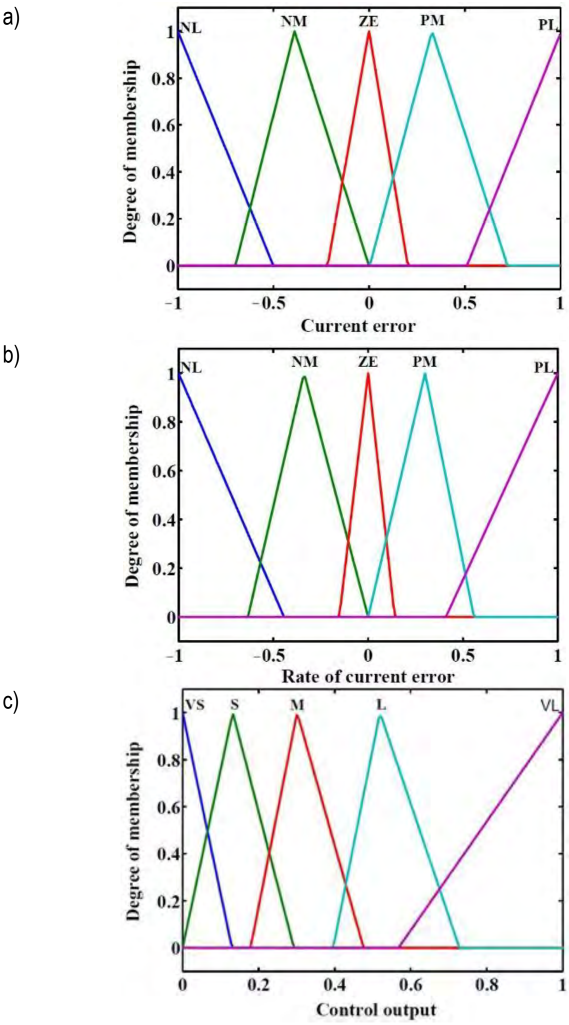


Fig. 8. Membership functions of FLC: (a) current error, (b) rate of current error, (c) control output

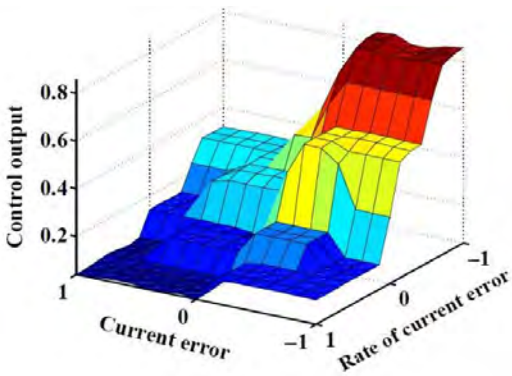


Fig. 9. Surface of FLC

Simulation results of using fuzzy control in order to track the desired current are presented in Fig. 10. As shown in this figure, the position and current verify the effectiveness of this controller. In comparison with hysteresis controller, this controller holds the current of coil without frequent fluctuations, which is an important advantage of this controller. Also, the fuzzy controller which needs no equations is easy for experimental implementation.

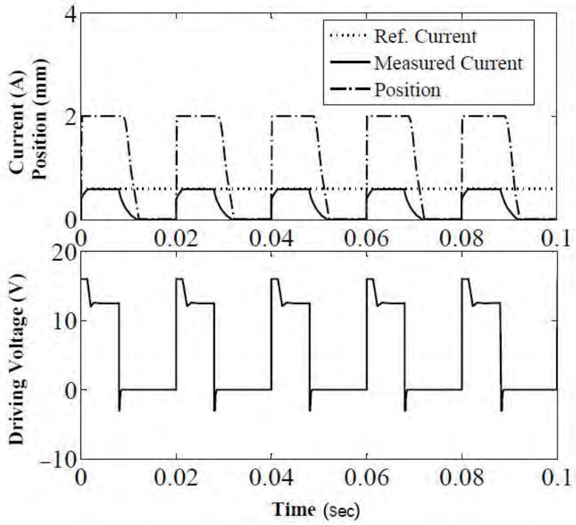


Fig. 10. Response of fuzzy controller

3.5. Comparison of designed controllers

In this section, evaluation of the controllers' performance is performed using controller effect and root mean square error (RMSE). Comparison of the dynamic performance of the four controllers is given in Table. 2. The obtained results show that the performance of FLC is better than that of the other three controllers. It should be mentioned that due to non-linearity of the system, the results of optimal controller based on predictive approach and FLC are better than the other two controllers and are approximately similar, but the fuzzy method is easy to implement in practice.

Tab. 2. Comparison of performance results for the designed controllers

Controller type	Controller effect	RMSE
Hysteresis	1.300247	0.095237
PI	1.250031	0.086718
Optimal model predictive	1.220253	0.080158
FLC	1.216844	0.079451

FLC, fuzzy logic controller; PI, proportional integrator

4. EMPIRICAL SETUP

In this section, an empirical setup consists of a solenoid, and a control circuit is implemented to enhance the performance of a solenoid valve. Fig. 11 shows an empirical setup of the system, which is controlled by using AVR microcontroller and FLC designed in the previous section. A solenoid of valve is a coil that pushes (or pulls) a plunger or spool until the current flows through it. There are several types of solenoids: push and pull types, with and without springs for pushing back the spool or plunger until the current no longer flows. Solenoids have some advantage in comparison with motors, including effecting a linear movement based on a compact mechanism without gears. On the other side, solenoids have a limited stroke.

DOI 10.2478/ama-2022-0011

In this paper, a small pull-type solenoid is utilised, which has a spring for pushing the plunger or spool back out.

The coil's resistance is 3.6 ohms, which draws 3.33 A when it is connected to a 12 V supply.

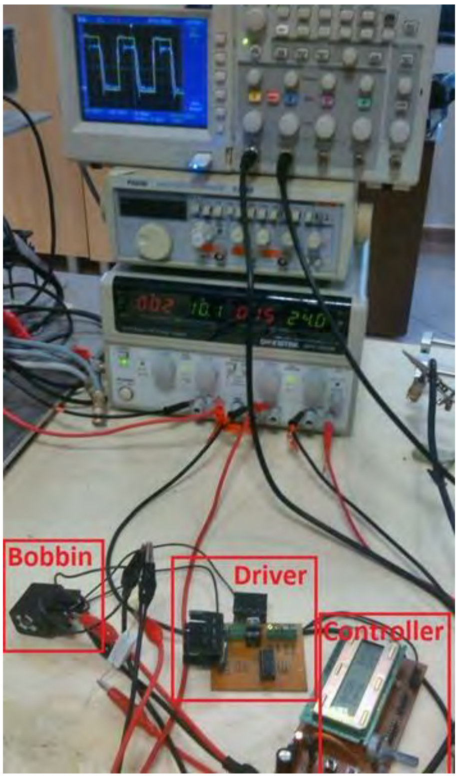


Fig. 11. The structure of empirical setup

4.1. Driving the solenoid

Fig. 12 shows the structure of the control circuit. The solenoid is connected to one of the controller ports in series with two resistors, where each resistor is equal to 1 ohm, which is for the safety of the circuit. It should be mentioned that the total resistance is more than 6 ohms, which limits the current of port to less than 1.5 A.

The solenoid is activated using the PWM at the specified duty cycle. The port voltage and the solenoid's current are measured using an oscilloscope, which is shown in Fig. 13.

The controller increases the voltage of the coil immediately. Due to the inductance of the coil, the current ramps up. This causes the voltage of coil to drop from about 15 V to less than 12 V. Then, the current continues to remain constant.

When the controller reduces the voltage, the current ramps down gradually, which causes a short negative voltage spike when the protection diode in the system starts conducting. When the voltage across it drops to under the threshold of diode, the protection diode stops conducting and the low negative voltage lingers in the port for a while longer.

The capacitor connected to the solenoid coil is charged through a 10 resistor that gives a 0.33 msec RC constant. It should be mentioned that, the used capacitor has an adequate capacitance to store sufficient charge and also low adequate internal resistance in discharge process.

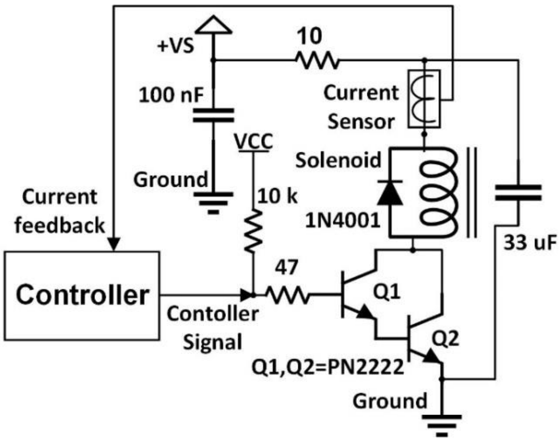


Fig. 12. The structure of control circuit

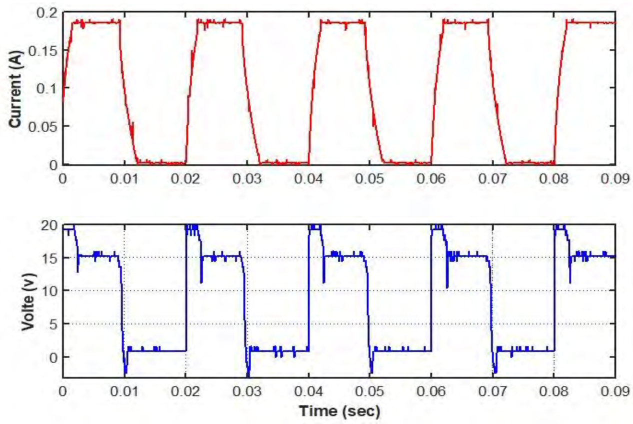


Fig. 13. Empirical performances: current and voltage

5. CONCLUSION

In this paper, the model of a fast-switching valve of 2–2 way, based on PWM, has been employed. To enhance the valve performance and decrease the delay in opening and closing, several controllers consisting of hysteresis, PI, optimal model predictive and fuzzy controllers are comparatively designed and implemented. Among the employed control approaches, PI, optimal model predictive and fuzzy controllers have faster performance, and it causes the valve to have long lifetime. Also, in comparison with other controllers, the FLC needs no equations and is easy to implement empirically.

REFERENCES

1. Boubakir A, Labiod L, Boudjema F. Direct adaptive fuzzy position controller for an electropneumatic actuator: Design and experimental evaluation. Mechanical Systems and Signal Processing. 2021;147. https://doi.org/10.1016/j.ymssp.2020.107066.
2. Najjari B, Barakati SM, Mohammadi A, Fotuhi MJ. Bostanian M. Position control of an electropneumatic system based on PWM technique and FLC, ISA Transaction. 2014;53(2):647–657. https://doi.org/10.1016/j.isatra.2013.12.023
3. Miha P, Niko H. Closed-loop volume flow control algorithm for fast switching pneumatic valves with PWM signal, Control Engineering Practice. 2018;70:114–120. https://doi.org/10.1016/j.conengprac.2017.10.008

4. Vinit S, Hitensinh V, Shk Madeenav L, Bikash RD, Anuj G. Effect of magnetic field environment on the performance of 3/2 solenoid valve, Fusion Engineering and Design. 2020;156(3):1–5. https://doi.org/10.1016/j.fusengdes.2020.111618
5. Taghizadeh M, Ghaffari A, Najafi F. Improving dynamic performances of PWM-driven servo-pneumatic systems via a novel pneumatic circuit, ISA Transaction. 2009; 48(4): 512–518. https://doi.org/10.1016/j.isatra.2009.05.001
6. Keles O, Ercan Y. Theoretical and experimental investigation of a pulse-width modulated digital hydraulic position control system, Control Engineering Practice. 2002;10(6): 645-654. https://doi.org/10.1016/S0967-0661(02)00021-7
7. Situm Z, Zilic T, Essert M. High Speed Solenoid Valves in Pneumatic Servo Applications, International Conference on Control, Automation and Systems Engineering. 2007;1–6. https://doi.org/10.1109/MED.2007.4433746
8. Stephen A, Murtaugh Jr. An introduction to time-modulated acceleration switching electro-hydraulic switching servomechanism. Journal of Basic Engineering. 1959;81(2):263–268. https://doi.org/10.1115/1.4008436
9. Noritsugu T. Development of PWM mode electro-pneumatic servo mechanism, part I: Speed control of a pneumatic cylinder. In: J Fluid Control. 54:65-80
10. Noritsugu T. Development of PWM mode electro-pneumatic servo mechanism, part II: Position control of a pneumatic cylinder, Journal of Fluid Control. 1986; 59:65-80.
11. Muto T, Kato H, Yamada H, Suematsu Y. Digital control of an HST system with load cylinder operated by differential pulse width modulation, Digital control of an HST system with load cylinder. 1993; 1993(2): 321-326. https://doi.org/10.5739/isfp.1993.321
12. Rao Z, Bone GM. Nonlinear Modeling and Control of Servo Pneumatic Actuators. IEEE Transactions on Control Systems Technology. 2008;16(3): https://doi.org/562–569. 10.1109/TCST.2007.912127
13. Messina A, Giannoccaro NI, Gentile A. Experimenting and modelling the dynamics of pneumatic actuators controlled by the pulse width modulation (PWM) technique, Mechatronics. 2005;15(7):859-881. https://doi.org/10.1016/j.mechatronics.2005.01.003
14. Leephakpreeda T. Fuzzy logic based PWM control and neural controlled-variable estimation of pneumatic artificial muscle actuators. Expert Systems with Applications. 2011;38(6):7837-7850. https://doi.org/10.1016/j.eswa.2010.12.120
15. Hodgson SM, Le Q, Tavakoli M, Pham MT. Improved tracking and switching performance of an electropneumatic positioning system. Mechatronics. 2012; 22(1):1-12. https://doi.org/10.1016/j.mechatronics.2011.10.007
16. Taghizadeh M, Ghaffari A, Najafi F. Modeling and identification of a solenoid valve for PWM control applications. Comptes Rendus Mécanique. 2009; 337(3): 131–140. https://doi.org/10.1016/j.crme.2009.03.009
17. Tao G, Chen HY, J YY, He ZB. Optimal design of the magnetic field of a high-speed response solenoid valve. Journal of Materials Processing Technology. 2002;129(3):555-558. https://doi.org/10.1016/S0924-0136(02)00633-7
18. Wang Q, Yang F, Yang Q, Chen J, Guan H. Experimental analysis of new high-speed powerful digital solenoid valves. Energy Conversion and Management. 2011;52(5):2309-2313. https://doi.org/10.1016/j.enconman.2010.12.032
19. Szente V, Vad J. Computational and Experimental Investigation on Solenoid Valve Dynamics. IEEE/ASME International Conference on Advanced Intelligent Mechatronics. Proceedings. 2001. https://doi.org/10.1109/AIM.2001.936537
20. Dulk I, Kovacshazy T. Modelling of a linear proportional electromagnetic actuator and possibilities of sensorless plunger position estimation. 12th International Carpathian Control Conference. 2011; 89–93. https://doi.org/10.1109/CarpathianCC.2011.5945822
21. Beater P. Pneumatic drives: system design, modelling and control, 4th ed. Springer. 2007
22. Murali MG, KK M. Modeling and PWM Control of Electro-Pneumatic Actuator for Missile Applications. In: IFAC-PapersOnLine. 2018; 51(1):237–242. https://doi.org/10.1016/j.ifacol.2018.05.057
23. Kuo BC, Golnaraghi MF. Automatic control systems. 3th ed. John Wiley and Sons. 2003.
24. Skogestad S, Postlethwaite I. Multivariable Feedback Control: Analysis and Design, 3th ed. John Wiley and Sons. 2005.

Behrouz Najjari: https://orcid.org/0000-0003-2132-6408

Mohammad J. Fotuhi: https://orcid.org/0000-0002-2271-8422

Mousa Vaezipour: https://orcid.org/0000-0003-0784-4137

DOI 10.2478/ama-2022-0012

MODIFIED LAPLACE BASED VARIATIONAL ITERATION METHOD FOR THE MECHANICAL VIBRATIONS AND ITS APPLICATIONS

Shahida REHMAN*, Akhtar HUSSAIN*, Jamshaid Ul RAHMAN*, Naveed ANJUM, Taj MUNIR***

*Abdus Salam School of Mathematical Sciences, Government College University,
68-B, New MuslimTown, Lahore 54600, Pakistan
**Faculty of Physical Sciences, Department of Mathematics, Government College University,
Allama Iqbal Road, Faisalabad-38000, Pakistan

shahidarehman21@sms.edu.pk, akhtarhussain21@sms.edu.pk, jamshaidrahman@gmail.com,
xsnaveed@yahoo.com, taj_math@hotmail.com

received 25 October 2021, revised 10 January 2022, accepted 12 January 2022

Abstract: In this paper, we are putting forward the periodic solution of non-linear oscillators by means of variational iterative method (VIM) using Laplace transform. Here, we present a comparative study of the new technique based on Laplace transform and the previous techniques of maximum minimum approach (MMA) and amplitude frequency formulation (AFF) for the analytical results. For the non-linear oscillators, MMA, AFF and VIM by Laplace transform give the same analytical results. Comparison of analytical results of VIM by Laplace transform with numerical results by fourth-order Runge–Kutta (RK) method conforms the soundness of the method for solving non-linear oscillators as well as for the time and boundary conditions of the non-linear oscillators.

Key words: variational iterative method, non-linear oscillator, Laplace transform

1. INTRODUCTION

The variational iterative method was suggested in 1990s to solve an oozing of with fractional deriva-tives and non-linear oscillator [2, 3, 4], and this method is used enormously as a main mathematical instrument for solving non-linear oscillators in various sciences (e.g. see [5–10]). This is a very popu-lar method in the list of methods for non-linear systems, and it includes high citation index articles dealing with the 'variational iterative method' (VIM). This method deals with non-linear oscillators like Fangzhu oscillators [11], fractal Toda oscillator [12], HIV models [13], biological models [14], fractal vibration models [15], microelectromechanical system oscillators [16, 17], fractal/fractional/ two-scale oscillators [18], interconnected spring carts [19, 20], etc. Naveed et al. [21, 22] investigated the ho-motopy perturbation method for the oscillators in nanotechnology. This paper suggests the periodic solution of the governing differential equations (non-linear oscillators) obtained by Laplace transform and VIM. The VIM retains a series of linear equations that can be solved by Laplace transform. This method identifies some obvious benefits, and its Lagrange multiplier is much trouble-free than that of variational theory [23–27]. The general recognition of the Lagrange multiplier by Laplace transform is given by Eq. (5) in [28].

Consider a non-linear oscillator in the equation form as

$$u''(t) + f(u) = 0 \tag{1}$$

with initial conditions $u(0) = A, u'(0) = 0$. Eq. (1) can be written as:

$$u'' + w^2 u + p(u) = 0 \tag{2}$$

where w is unknown frequency, $\mathrm{p(u)} = f(u) - w^2 u$. As claimed by the VIM, the correction functional which is basically a convolution for Eq. (2) is given as [29–33]:

$$u_{n+1}(t) = u_n(t) + \int_0^t \zeta(t,\xi)[u_n''(\xi) + w^2 u_n(\xi) + \tilde{p}(u_n)]d\xi \qquad n = 0,1,2,\ldots \tag{3}$$

where ζ is a general Lagrange multiplier, and it can be choicely determined by immobilised conditions of Eq. (3) with respect to using variational theory [10–13]. The subscript *n* denotes the *n*th approximation of the solution and $\tilde{p}$ is a restricted variation. This convolution gives the value of ζ by making $u_n(t)$ immobilised. This method is applicable to derive the analytical solution for the motion of non-linear unbound vibration of conservative, single degree of freedom systems.

Now we implement this method to justify the motion of two oscillators by making Laplace transform in the well-known VIM to obtain the relationship between amplitude and angular frequency. This method is equally good when compared with the older versions of VIM, Ganji and Azimi [1] maximum minimum approach (MMA), and amplitude frequency formulation (AFF) to non-linear oscillation systems. For the first problem shown in (Fig. 1), consider a block of mass m_1, which is on the horizontal surface, while another block of mass m_2 is just slipped vertical and is also attached with mass m_1. In this system, the length of support is *L*, gravitational acceleration produced due to free motion of blocks is denoted by *g* and *K* is spring constant. If we assume that $u = \frac{x}{L} \ll 1$, then the equation in the range of time and boundary conditions is given as:

$$u'' + \left(\frac{m_2}{m_1}\right)u^2u'' + \left(\frac{m_2}{m_1}\right)uu'^2 + \left(\frac{K}{m_1} + \frac{m_2\boldsymbol{g}}{m_1}\right)u + \frac{m_2\boldsymbol{g}}{2Lm_1}u^3 = 0, \quad u(0) = A, \quad u'(0) = 0 \tag{4}$$

where both *u*, the generalised displacement, and *t*, the time variable, are dimensionless.

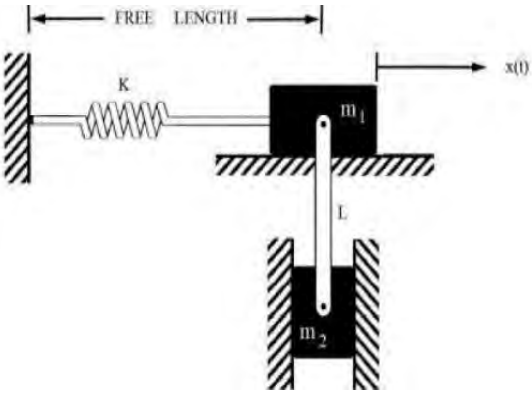


Fig. 1. Geometry of the first Problem

2. APPLICATIONS OF LAPLACE-BASED VIM TO THE FIRST PROBLEM

In order to solve the first problem given in Eq. (4), by making Laplace transform in VIM, we rewrite the problem as:

$$u'' + \left(\frac{m_2}{m_1}\right)u^2u'' + \left(\frac{m_2}{m_1}\right)uu'^2 + w_0^2u + \frac{m_2\boldsymbol{g}}{2Lm_1}u^3 = 0, \quad u(0) = A, \quad u'(0) = 0 \tag{5}$$

where $\mathrm{w}_0^2 = \left(\frac{\mathrm{K}}{\mathrm{m}_1} + \frac{\mathrm{m_2g}}{\mathrm{Lm}_1}\right)$. We can rewrite the equation in the form:

$$(1 + \alpha u^2)u'' + \alpha uu'^2 + w_0^2u + \beta u^3 = 0, \quad u(0) = A, \quad u'(0) = 0 \tag{6}$$

where $\alpha = \frac{\mathrm{m}_2}{\mathrm{m}_1}$ and $\beta = \frac{\mathrm{m_2g}}{2\mathrm{Lm}_1}$. To solve the above equation, we use VIM by Laplace transform. To approach the correctional functional, we write the above equation in general non-linear oscillator form as:

$$u'' + w^2u + p(u) = 0 \tag{7}$$

where

$$p(u) = \alpha u^2u'' + \alpha uu'^2 + (w_0^2 + w^2)u + \beta u^3 \tag{8}$$

The correctional functional is defined as

$$u_{n+1}(t) = u_n(t) + \int_0^t \zeta(t,\xi)[u_n''(\xi) + w^2u_n(\xi) + \tilde{p}(u_n)]d\xi \tag{9}$$

where the Lagrange multiplier is given by $\zeta(\xi) = -\frac{1}{w}\sin wt$. Now, by applying Laplace transform to Eq. (9), we have

$$L[u_{n+1}(t)] = L[u_n(t)]]$$

$$-L[\int_0^t \frac{1}{w}\sin w(t-\xi)[u_n''(\xi) + w^2u_n(\xi) + p(u_n)]d\xi$$

$$L[u_{n+1}(t)] = L[u_n(t)] - \frac{1}{w}L[\sin wt]L[u_n'' + \alpha u_nu_n'^2 + \alpha u_n^2u_n'' + w_0^2u_n + \beta u_n^3] \tag{10}$$

For a first-order approximate solution, put $n = 0$, we get

$$L[u_1(t)] = L[u_0(t)] - \frac{1}{w}L[\sin wt]L[u_0'' + \alpha u_0u_0'^2 + \alpha u_0^2u_0'' + w_0^2u_0 + \beta u_0^3]$$

Using the trail function $\mathrm{u}_0(\mathrm{t}) = \mathrm{A\cos wt}$, the above equation gets the form

$$L[u_1(t)] = L[A\cos wt] - \frac{1}{w}L[\sin wt]L[-Aw^2\cos wt - \alpha A^3w^2\cos^3 wt + \alpha A^3w^2\cos wt\sin^2 wt + w_0^2A\cos wt + \beta A^3\cos^3 wt]$$

$$L[u_1(t)] = L[A\cos wt] - \frac{1}{w}L[\sin wt]L[-Aw^2\cos wt - \frac{\alpha A^3w^2}{4}(3\cos wt - \cos 3wt) + \frac{\alpha A^3w^2}{4}(\cos wt + \cos 3wt) + w_0^2A\cos wt + \frac{\beta A^3}{4}(3\cos wt - \cos 3wt)]$$

After some simplification, we have the expression as

$$\mathrm{L}[\mathrm{u}_1(\mathrm{t})] = \mathrm{L}[\mathrm{A\cos wt}] - \frac{1}{w}\left[-Aw^2 - \frac{\alpha A^3w^2}{2} + w_0^2A + \frac{3\beta A^3}{4}\right]L[\sin wt]L[\cos wt] - \frac{1}{w}\left[\frac{\alpha A^3w^2}{4} - \frac{\beta A^3}{4}\right]L[\sin wt]L[\cos 3wt]$$

By inverse Laplace transform, the expression for the first-order approximate solution is

$$u_1(t) = A\cos wt - \frac{1}{w}\left[-Aw^2 - \frac{\alpha A^3w^2}{2} + w_0^2A + \frac{3\beta A^3}{4}\right]\left(\frac{1}{2}t\sin wt\right) - \frac{1}{w}\left[\frac{\alpha A^3w^2}{4} - \frac{\beta A^3}{4}\right]\left(\frac{1}{8w}\cos wt - \cos 3wt\right) \tag{11}$$

In Eq. (11), the second term is a secular term because it grows in amplitude with time, so avoiding the secular term in approximate solution requires that

$$-\frac{1}{w}\left[-Aw^2 - \frac{\alpha A^3w^2}{2} + w_0^2A + \frac{3\beta A^3}{4}\right] = 0$$

$$Aw^2 + \frac{\alpha A^3w^2}{2} = w_0^2A + \frac{3\beta A^3}{4}$$

$$w^2 = \frac{w_0^2 + \frac{3A^2}{4}\frac{m_2\boldsymbol{g}}{2Lm_1}}{1 + \frac{m_2A^2}{2m_1}}$$

This leads to the expression for angular frequency of the system:

$$w = \frac{1}{2}\sqrt{\frac{8w_0^2Lm_1 + 3m_2\boldsymbol{g}A^2}{L(m_2A^2 + 2m_1)}}$$

This expression for angular frequency of the first problem is exactly the same as obtained by the MMA in Eq. (11) and the AFF method in Eq. (22) by Ganji and Azimi [1] and He in [34]. So, the periodic solution in this case becomes the same as that of MMA and AFF, while the first-order approximate solution is given as

$$u_1(t) = A\cos wt - \frac{1}{w}\left[\frac{\alpha A^3w^2}{4} - \frac{\beta A^3}{4}\right]\left(\frac{1}{8w}\cos wt - \cos 3wt\right). \tag{12}$$

3. APPLICATION OF THE LAPLACE-BASED VIM TO THE SECOND PROBLEM

The second problem deals with the motion of simple pendulum devoted to a spinning rigid frame shown in (Fig. 2), which has the differential equation:

DOI 10.2478/ama-2022-0012

$$\theta'' + (1 - \Lambda\cos\theta)\sin\theta = 0, \quad \theta'(0) = 0, \; \theta(0) = A \tag{13}$$

where θ is generalised displacement without dimensions, t is the time variable and Λ indicates the relation $\Lambda = \frac{w^2 r}{g}$.

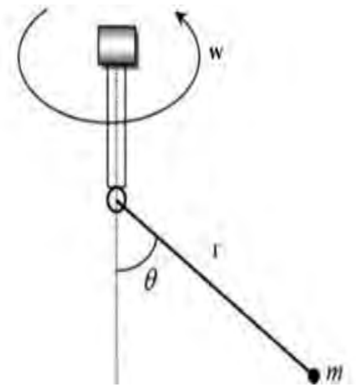


Fig. 2. Geometry of the first Problem

In order to solve this problem by using VIM with Laplace transform, we write Eq. (13) as

$$\theta'' + (1-\Lambda)\theta + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)\theta^3 + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)\theta^5 = 0, \quad \theta'(0) = 0, \; \theta(0) = A \tag{14}$$

Eq. (14) in the form of general non-linear oscillator has the form

$\theta'' + w^2\theta + p(\theta) = 0$, where $p(\theta) = -w^2\theta + (1-\Lambda)\theta + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)\theta^3 + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)\theta^5$.

The correctional functional to approximate solution is defined as

$$\theta_{n+1}(t) = \theta_n(t) + \int_0^t \zeta(t,\xi)[\theta_n''(\xi) + w^2\theta_n(\xi) + \tilde{p}(\theta_n)]d\xi$$

By using Lagrange multiplier $\zeta(\xi) = -\frac{1}{w}\sin wt$, we have the iterative formula

$$\theta_{n+1}(t) = \theta_n(t) - \frac{1}{w}\int_0^t \sin(t-\xi)[\theta'' + (1-\Lambda)\theta + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)\theta^3 + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)\theta^5]d\xi$$

Now, by applying Laplace transform on the above iterative formula, we get

$$L[\theta_{n+1}(t)] = L[\theta_n(t)] - \frac{1}{w}L[\sin wt]L[\theta'' + (1-\Lambda)\theta + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)\theta^3 + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)\theta^5]$$

For the first-order approximate solution, use $n = 0$ and trail function $u_0(t) = A\cos wt$.

$$L[\theta_1(t)] = L[A\cos wt] - \frac{1}{w}L[\sin wt]L[-Aw^2\cos wt + (1-\Lambda)A\cos wt + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)A^3\cos^3 wt + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)A^5\cos^5 wt]$$

$$L[\theta_1(t)] = L[A\cos wt] - \frac{1}{w}L[\sin wt]L[-Aw^2\cos wt + (1-\Lambda)A\cos wt + \left(\frac{-1}{6} - \frac{2\Lambda}{3}\right)\frac{A^3}{4}(\cos 3wt + 3\cos wt) + \left(\frac{1}{120} - \frac{2\Lambda}{15}\right)\frac{A^5}{16}(\cos 5wt + 5\cos 3wt + 10\cos wt)]$$

After simplification, we get

$$L[\theta_1(t)] = L[A\cos wt] - \frac{1}{w}\left(-Aw^2 + (1-\Lambda)A - \frac{A^3}{8} + \frac{A^3\Lambda}{2} + \frac{A^5}{192} - \frac{A^5\Lambda}{12}\right)L\left(\frac{t}{2}\sin wt\right) - \frac{1}{w}\left(-\frac{A^3}{24} + \frac{A^3\Lambda}{6} + \frac{A^5}{384} - \frac{A^5\Lambda}{24}\right)L\left(\frac{1}{8w}(\cos wt - \cos 3wt)\right)$$

By applying inverse Laplace, the expression reduces to

$$\theta_1(t) = A\cos wt - \frac{1}{w}\left(-Aw^2 + (1-\Lambda)A - \frac{A^3}{8} + \frac{A^3\Lambda}{2} + \frac{A^5}{192} - \frac{A^5\Lambda}{12}\right)\left(\frac{t}{2}\sin wt\right) - \frac{1}{w}\left(-\frac{A^3}{24} + \frac{A^3\Lambda}{6} + \frac{A^5}{384} - \frac{A^5\Lambda}{24}\right)\left(\frac{1}{8w}(\cos wt - \cos 3wt)\right)$$

Here, in this equation, the second term is a secular term because it grows in amplitude with time, so avoiding the secular term in approximate solution required that

$$-\frac{1}{w}\left(-Aw^2 + (1-\Lambda)A - \frac{A^3}{8} + \frac{A^3\Lambda}{2} + \frac{A^5}{192} - \frac{A^5\Lambda}{12}\right) = 0$$

$$w^2 = \left((1-\Lambda) - \frac{A^2}{8} + \frac{A^2\Lambda}{2} + \frac{A^4}{192} - \frac{A^4\Lambda}{12}\right)$$

The expression for the angular frequency is given as:

$$w = \sqrt{1 - \Lambda - \frac{A^2}{8} + \frac{A^2\Lambda}{2} + \frac{A^4}{192} - \frac{A^4\Lambda}{12}}$$

The expression for angular frequency of the second problem is exactly the same as obtained by the MMA in Eq. (30) and the AFF method in Eq. (35) by Ganji and Azimi [1]. So, the periodic solution in this case becomes the same as that of MMA and AFF, while the approximate solution is

$$\theta_1(t) = A\cos wt - \frac{1}{w}\left(-\frac{A^3}{24} + \frac{A^3\Lambda}{6} + \frac{A^5}{384} - \frac{A^5\Lambda}{24}\right) * \left(\frac{1}{8w}(\cos wt - \cos 3wt)\right) \tag{15}$$

4. RESULTS AND DISCUSSION FOR THE FIRST PROBLEM

In this section, we have compared the numerical solution of non-linear oscillator (4) obtained by fourth-order Runge–Kutta (RK) method with analytical solutions obtained by Laplace based VIM. The analytical solution by VIM using Laplace transform coincides analytically with MMA and AFF techniques. In (Fig. 3), the comparison between analytical solution by VIM, VIM with Laplace and numerical solution by fourth-order RK method shows the validity of Laplace-based VIM.

In this session, we have characterised the error analysis of the analytical solution by VIM with Laplace transform and numerical solution by fourth order RK method. In Table 1, the error terms, e_1 and e_2 are by VIM and VIM with Laplace transform, respectively. Error e_2 conforms the validity of the solution by VIM with Laplace transform.

sciendo
DOI 10.2478/ama-2022-0012 acta mechanica et automatica, vol.16 no.2 (2022)

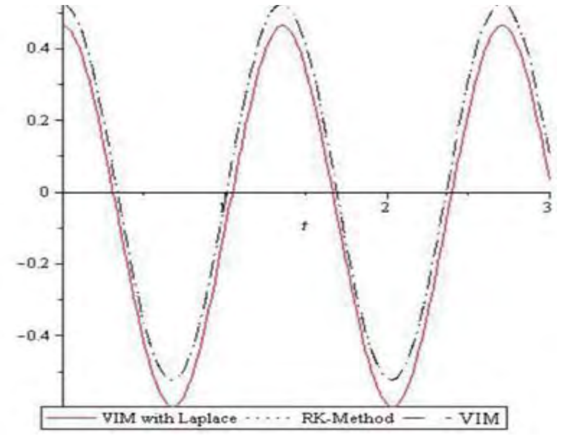


Fig. 3. Comparison among VIM, VIM with Laplace and fourth-order RK method in the first problem for $= \frac{\pi}{6}$; g = 9.81 $\mathrm{m\ s^{-2}}$, K = 100 $\mathrm{N\ m^{-2}}$, $\mathrm{m_1} = 5\ \mathrm{kg}$, $\mathrm{m_2} = 1\ \mathrm{kg}$, L= $1\ \mathrm{m}$, $\alpha = \frac{1}{5}$, $\beta = 0.981\ \mathrm{s^{-2}}$

Tab. 1. Error analysis for the first problem

Time steps	Previous results	Our results	RK method	e_1	e_2
1	0.468130637	0.402613907	0.469219893	0.001089256	0.066606
2	0.313478406	0.239576597	0.315842007	0.002363601	0.0762654
3	0.0924086651	0.021854038	0.093445360	0.001036695	0.0715913
4	−0.148239947	−0.209104415	−0.149924001	0.001684053	0.0591804
5	−0.357480565	−0.418137275	−0.359780544	0.002299979	0.0583567
6	−0.490980817	−0.561137359	−0.491696084	0.000715266	0.0694413
7	−0.520455648	−0.594098506	−0.520518333	0.042237427	0.0735802
8	−0.439660145	−0.505017274	−0.441122041	0.001461896	0.0638952
9	−0.265712668	−0.324704000	−0.267960752	0.002248084	0.0567432
10	−0.035467936	−0.100595290	−0.035762460	0.000294524	0.0648328

5. RESULTS AND DISCUSSION FOR THE SECOND PROBLEM

In this sectio, we have compared the numerical solution of non-linear oscillator (13) obtained by fourth-order RK method and the analytical solution. The analytical solution by VIM using Laplace transform coincides analytically with MMA and AFF techniques. In (Fig. 4), the comparison among analytical solution by VIM, VIM with Laplace and numerical solution by fourth-order RK method shows the validity of VIM with Laplace.

In this session, we have characterised the error analysis of the analytical solution by VIM with Laplace transform and numerical solution by fourth-order RK method. In Table 2, the error terms $\mathrm{e_1}$ and $\mathrm{e_2}$ are by VIM and VIM with Laplace transform, respectively. Error $\mathrm{e_2}$ conforms the validity of the solution by VIM with Laplace transform.

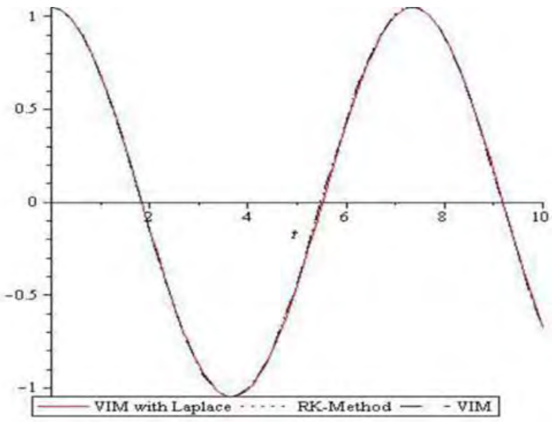


Fig. 4. Comparison among VIM, VIM with Laplace and fourth-order RK method in the second problem for $\mathrm{A} = \frac{\pi}{3}$ and $\Lambda = 0.25$

Tab. 2. Error analysis for the second problem

Time steps	Previous results	Our results	RK method	e_1	e_2
1	1.043371279	1.043420173	1.043429864	0.039474666	0.024021587
2	1.031920422	1.032112440	1.032149462	0.000229041	0.000037022
3	1.012928661	1.013347597	1.013424679	0.000496018	0.000077082
4	0.986534778	0.987247802	0.987370428	0.000835650	0.000122626
5	0.952931653	0.953984220	0.954149663	0.001218009	0.000165443
6	0.912364844	0.913777051	0.913974673	0.001609828	0.000197622
7	0.865130800	0.866895416	0.867108671	0.001977871	0.000213255
8	0.811574689	0.813657009	0.813866137	0.002291448	0.000209128
9	0.752087881	0.754427405	0.754613127	0.002525246	0.000185722
10	0.687105084	0.689618923	0.689765626	0.002660543	0.000146703

6. CONCLUSIONS

In this paper, VIM by Laplace transform is applied to non-linear oscillators to compute the analytical results. Earlier, two techniques, MMA and AFF, were used for analytical results. Our technique, VIM with Laplace, coincides analytically with MMA and AFF, but is graphically slightly different than that of the numerical solution by fourth-order RK method, MMA and AFF.

REFERENCES

1. Ganji DD, Azimi M. Application of max min approach and amplitude frequency formulation to nonlinear oscillation systems. UPB Scientific Bulletin. 2012 Jan 1;74(3):131-40.
2. Suleman M, Lu D, Yue C, Ul Rahman J, Anjum N. He–Laplace method for general nonlinear periodic solitary solution of vibration equations. Journal of Low Frequency Noise, Vibration and Active Control. 2019 Dec;38(3-4):1297-304.
3. He JH. A short remark on fractional variational iteration method. Physics Letters A. 2011 Sep 5;375(38):3362-4.

DOI 10.2478/ama-2022-0012

4. He JH. Variational iteration method–a kind of non-linear analytical technique: some examples. International journal of non-linear mechanics. 1999 Jul 1;34(4):699-708.
5. He JH. Variational principles for some nonlinear partial differential equations with variable coefficients. Chaos, Solitons & Fractals. 2004 Mar 1;19(4):847-51.
6. He JH. Variational approach to (2+ 1)-dimensional dispersive long water equations. Physics Letters A. 2005 Feb 7;335(2-3):182-4.
7. ul Rahman J, Mohyuddin MR, Anjum N, Zahoor S. Mathematical Modelling and Simulation of Mixing of Salt in 3-Interconnected Tanks. Journal of Advances in Civil Engineering. 2015;1(1):1-6.
8. Anjum N, Ain QT. Application of He's fractional derivative and fractional complex transform for time fractional Camassa-Holm equation. Thermal Science. 2020;24(5 Part A):3023-30.
9. Anjum N, He JH. Analysis of nonlinear vibration of nano/ microelectromechanical system switch induced by electromagnetic force under zero initial conditions. Alexandria Engineering Journal. 2020 Dec 1;59(6):4343-52.
10. Ain QT, Anjum N, He CH. An analysis of time-fractional heat transfer problem using two-scale approach. GEM-International Journal on Geomathematics. 2021 Dec;12(1):1-0
11. He JH, El-Dib YO. Homotopy perturbation method for Fangzhu oscillator. Journal of Mathematical Chemistry. 2020 Nov; 58(10): 2245-53.
12. He JH, El-Dib YO, Mady AA. Homotopy perturbation method for the fractal toda oscillator. Fractal and Fractional. 2021 Sep;5(3):93.
13. Suleman M, Lu D, He JH, Farooq U, Hui YS, Rahman JU. Numerical investigation of fractional HIV model using Elzaki projected differential transform method. Fractals. 2018 Oct 5;26(05):1850062.
14. Ul Rahman J, Lu D, Suleman M, He JH, Ramzan M. He–Elzaki method for spatial diffusion of biological population. Fractals. 2019 Aug 13;27(05):1950069.
15. He CH, Liu C, He JH, Gepreel KA. Low frequency property of a fractal vibration model for a concrete beam. Fractals. 2021;29(5):2150117-33.
16. Anjum N, He JH. Higher-order homotopy perturbation method for conservative nonlinear oscillators generally and microelectro-mechanical systems' oscillators particularly. International Journal of Modern Physics B. 2020 Dec 30;34(32):2050313.
17. Tian D, Ain QT, Anjum N, He CH, Cheng B. Fractal N/MEMS: from pull-in instability to pull-in stability. Fractals. 2021 Mar 10;29(02):2150030.
18. Ain QT, Anjum N, He CH. An analysis of time-fractional heat transfer problem using two-scale approach. GEM-International Journal on Geomathematics. 2021 Dec;12(1):1-0.
19. Ain QT, He JH, Anjum N, Ali M. The fractional complex transform: A novel approach to the time-fractional Schrödinger equation. Fractals. 2020 Nov 2;28(07):2050141.
20. ul Rahman J, Mohyuddin MR, Anjum N, Butt R. Modelling of Two Interconnected Spring Carts and Minimization of Energy. DJ Journal of Engineering and Applied mathematics. 2016;2(1):7-11.
21. Ali M, Anjum N, Ain QT, He JH. Homotopy perturbation method for the attachment oscillator arising in nanotechnology. Fibers and Polymers. 2021 Jun;22(6):1601-6.
22. Rahman JU, Suleman M, Anjum N. Solution of unbounded boundary layer equation using modified homotopy perturbation method. Int. J. Macro Nano Phys. 2018;3(1):11-5.
23. He JH. Some asymptotic methods for strongly nonlinear equations. International journal of Modern physics B. 2006 Apr 20;20(10):1 141-99.
24. Noor MA, Mohyud-Din ST. Variational iteration method for solving higher-order nonlinear boundary value problems using He's polynomials. International Journal of Nonlinear Sciences and Numerical Simulation. 2008 Jun 1;9(2):141-56.
25. He JH. Generalized equilibrium equations for shell derived from a generalized variational principle. Applied Mathematics Letters. 2017 Feb 1;64:94-100.
26. He JH. An alternative approach to establishment of a variational principle for the torsional problem of piezoelastic beams. Applied Mathematics Letters. 2016 Feb 1;52:1-3.
27. Wu Y, He JH. A remark on Samuelson's variational principle in economics. Applied Mathematics Letters. 2018 Oct 1;84:143-7.
28. Anjum N, He JH. Laplace transform: making the variational iteration method easier. Applied Mathematics Letters. 2019 Jun 1;92:134-8.
29. He JH. Variational iteration method—some recent results and new interpretations. Journal of computational and applied mathematics. 2007 Oct 1;207(1):3-17.
30. He JH, Wu XH. Variational iteration method: new development and applications. Computers & Mathematics with Applications. 2007 Oct 1;54(7-8):881-94.
31. He JH. Variational iteration method for autonomous ordinary differential systems. Applied mathematics and computation. 2000 Sep 11;114(2-3):115-23.
32. He JH. Variational theory for linear magneto-electro-elasticity. International Journal of Nonlinear Sciences and Numerical Simulation. 2001 Dec 1;2(4):309-16.
33. He J. Variational iteration method for delay differential equations. Communications in Nonlinear Science and Numerical Simulation. 1997 Dec 1;2(4):235-6.
34. He JH. An improved amplitude-frequency formulation for nonlinear oscillators. International Journal of Nonlinear Sciences and Numerical Simulation. 2008 Jun 1;9(2):211-2.

Shahida Rehman: https://orcid.org/0000-0001-7137-9079

Akhtar Hussain: https://orcid.org/0000-0002-1892-3212

Jamshaid Ul Rahman: https://orcid.org/0000-0001-8642-0660

Naveed Anjum: https://orcid.org/0000-0001-7810-7877

Taj Munir: https://orcid.org/0000-0001-8998-7937

sciendo
DOI 10.2478/ama-2022-0013 acta mechanica et automatica, vol.16 no.2 (2022)

METHOD FOR ENHANCED ACCURACY IN MACHINING FREE-FORM SURFACES ON CNC MILLING MACHINES

Andrzej WERNER*

*Faculty of Mechanical Engineering, Bialystok University of Technology, ul. Wiejska 45C, 15-351 Bialystok, Poland

a.werner@pb.edu.pl

received 29 October 2021, revised 13 January 2022, accepted 22 January 2022

Abstract: The present article describes a method for enhanced accuracy in machining free-form surfaces produced on CNC milling machines. In this method, surface patch machining programs are generated based on their nominal CAD model. After the pretreatment, coordinate control measurements are carried out. The obtained results of the measurements contain information on the values and distribution of observed machining deviations. These data, after appropriate processing, are used to build a corrected CAD model of the surface produced. This model, made using reverse engineering techniques, compensates for the observed machining deviations. After regeneration of machining programs, the object processing and control measurements are repeated. As a result of the conducted procedure, the accuracy of the manufacture of the surface object is increased. This article also proposes the introduction of a simple procedure for the filtration of measurement data. Its purpose is to minimise the effect of random phenomena on the final machining error correction. The final part of the article presents the effects of the proposed method of increasing the accuracy of manufacturing on 'raw' and filtered measurement data. In both cases, a significant improvement in the accuracy of the machining process was achieved, with better final results obtained from the filtered measurement data. The method proposed in the article has been verified for three-axis machining with a ball-end cutter.

Key words: free-form surface, milling, coordinate measurements, CAD model, data filtration, accuracy improvement

1. INTRODUCTION

The machining of objects containing curved geometries is currently used in the industry for the production of various types of cams, blanking dies and electrodes for electrical discharge machining (EDM). When producing these elements, it is necessary to maintain high accuracy. This fact often requires a machining error compensation procedure. Various approaches are currently in use to improve manufacturing accuracy [1, 2]. One of these approaches is to determine the geometrical errors of the CNC (computer numerical control) machine tool and use them for the correction of machining programs [3, 4]. This approach requires a series of machining tests and control measurements aimed at the determination of models describing the machine tool errors [5, 6, 7]. These models are used to correct machining programs before machining [8, 9].

A similar approach is to identify and eliminate geometric errors in a CNC machine tool. For example, the recent researches focus on development of a method to accurately identify geometric errors of five-axis CNC machines. A theoretical model for identification of geometric errors is created. In this model, both position-independent errors and position-dependent errors are considered as the error sources. Experiments on a five-axis CNC machine tool also demonstrate significant reduction in the volumetric error after error compensation [10].

There is another interesting approach in which the cutter/workpiece engagement is taken into account. It varies with the tool orientation continuously including lead and tilt angles during machining, which results in the obvious time-varying characteristic for consecutive cutting forces. Considering tool orientation, actual cutter runout and cutter motion process, an accurate calculation model for instantaneous cutter/workpiece engaging process in five-axis ball-end milling is calculated based on an improved analytical method. Then, based on the cutting force model, the tool orientation optimisation strategies with a flexible cutter and rigid workpiece for roughing and finishing milling operation are elaborated [11].

An interesting and complicated approach is to take into account the geometrical aspects of the tool and the workpiece. An example is a precise approach to the generation of optimised collision-free and gouging-free tool paths for five-axis CNC machining of freeform NURBS (Non-Uniform Rational B-Spline) surfaces using flat-end and rounded-end (bull nose) tools having cylindrical shank. A global optimisation is performed to find the tool path that maximises the approximation quality of the machining [12]. Another example of this approach is five-axis machining of freeform surfaces, where the degrees of freedom in selecting and moving the cutting tool allow one to adapt the tool motion optimally to the surface to be produced. A careful geometric analysis of curvature-adapted machining via so-called second-order line contact between tool and target surface is performed. As a result, better toolpath generation results are obtained, which results in improved machining accuracy [13].

Another approach is to analyse the errors, which are the source of the machining process itself and the accompanying phenomena [14]. These are mainly deformations of the tool and workpiece, originating from cutting forces and inertial forces [15]. Literature analysis related to this approach indicates that many methods have been developed to increase manufacturing accura-

10.2478/ama-2022-0013

cy. One of these is the design of the machining process, in which the cutting forces are controlled by adjusting parameters such as the feed or the width of the machined layer [16]. As a result, the tool does not bend beyond a certain limit. Adaptive control systems are being developed to correct the position of the tool in real time [17]. They require the machine tool to be set up with sensors to control the parameter values affecting the accuracy of machining (e.g. tool deflection, displacement of the machine's executive unit, etc.). Another way is the method of modifying the tool path based on the calculated tool deflection [18]. Deflection is determined by processing the process simulation models that have been developed. These models take into account deformation of the tool, the tool holder and the spindle when compensating the tool path due to the cutting forces.

Another universal method of increasing the accuracy of manufactured elements is the method using coordinate measurements. The correction of the machining process is based on the results of measurements made on a CNC machine tool [19] or a coordinate measuring machine (CMM) [20]. Measurements performed on the machine tool use machine tool probes. The advantage of this approach is that machining and measurements are made without changing the position of the workpiece in the same coordinate system. In the case of measurements made on a CMM, the workpiece must be removed from the processing station and transferred to the measuring station. A very important element of the measurement process in this case is reconstruction of the (coordinate system) setting of the workpiece on a CMM [21]. Measurement data obtained in two ways are usually compared to nominal CAD models of workpieces [22]. On this basis, machining deviations are determined and are then used in the process of correction of manufacturing errors.

The method of increasing the accuracy of manufacturing proposed in the article has been applied to the treatment of a free-form surface on a CNC milling machine. In this method, the creation of functions describing the distribution of machining deviations is omitted. Correction of machining programs is carried out indirectly by converting the nominal CAD model of the object into a corrected geometrical model, taking into account the deviations observed after the initial machining of the workpiece. Reconstruction of the CAD model takes place through the direct use of the results of coordinate measurements (observed deviations) carried out after the initial treatment. The expected effects of the proposed method are an increase in the accuracy of free surface preparation and simplification of the entire process of increasing the machining accuracy. An additional aspect proposed in the article is the introduction of a simple procedure for the filtration of measurement data. Its purpose is to minimise the effect of random phenomena on the final machining error correction. The method proposed in the article has been verified for three-axis machining with a ball-end cutter.

2. METHOD FOR INCREASING THE ACCURACY OF MANUFACTURING

2.1. Procedure

For increasing the accuracy of free-form surfaces produced on CNC milling machines, a method that is based on coordinate measurements performed outside the machine tool is presented in this article (Fig. 1). The results of measurements carried out after machining are data on the basis of which the values and distribution of observed machining deviations are determined. In this method, the nominal geometric model of the surface object is created first. This model accurately reproduces the shape and dimensions of the surface area being produced. On this basis, programs controlling milling centre machining are created in the CAD/CAM system. In the next step, after the machine is properly equipped, the parts are machined. After the processing, coordinate control measurements of the created object are carried out.

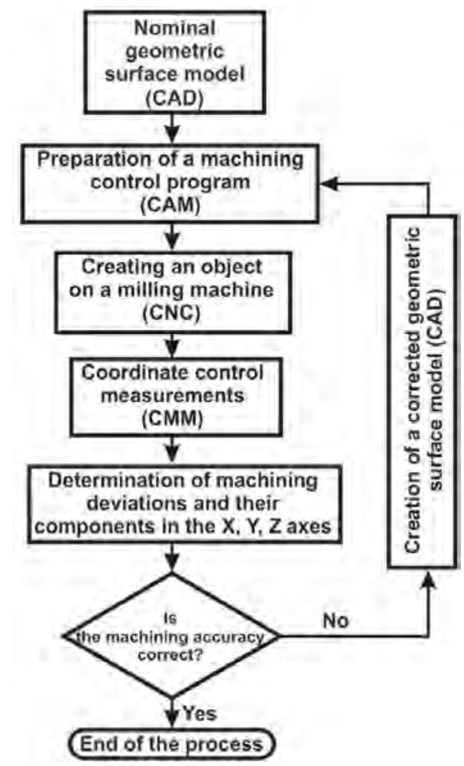


Fig. 1. Method for increasing the accuracy of the manufacturing of shaped surfaces

In the next stage of the process, based on the results of coordinate measurements, observed deviations and their components in individual axes are determined. The achieved machining accuracy can be estimated. The observed machining errors are compared with the required accuracy. If the machining accuracy meets the expectations, the manufacturing process is finished. In the case when the obtained accuracy is not satisfactory, machining errors are compensated. This compensation requires the reconstruction of the nominal geometric model of the manufactured object into a model that corrects the machining errors that occur, the re-creation of machining control programs and repetition of machining. The reconstruction of the geometrical model uses the results of coordinate measurements. These data contain information such as coordinates of nominal and observed points, deviations observed and directional cosines describing the direction of deviations observed. On the basis of this information, the components of deviations observed in the X, Y and Z axes are determined. These components are used in creating the corrected geometric model of the object. The method of determining the components of machining deviations and the reconstruction of the nominal surface model are described later in this article. The corrected surface geometric model is used for the re-creation of machining programs. After they are obtained, parts processing and coordinate control measurements are repeated. If, after the repetition of machining, the accuracy obtained is still unsatisfactory, the correction process can be repeated again. It should be noted that in control coordinate measurements, machining deviations are always determined in relation to the initial, nominal geometric surface model.

2.2. Determining machining deviations and reconstructing the geometrical model

Control coordinate measurements of a free-form surface with a rectangular contour can be performed with a bidirectional uniform distribution of measurement points. In this way, the *n*x*m* grid of observed points is obtained. For this purpose, automatic surface scanning procedures can be used, for example, *UVScan* or *Grid* (PC-DMIS system). The number of measuring points is adjusted to the degree of complexity of the shape of the object.

In order to estimate the accuracy of making the free-form surface describing the object to be manufactured, machining deviations should be determined at the measuring points. The measure of the determined deviations are the distances between the points on the surface of the CAD model (nominal surface) and the corresponding points observed as a result of control measurements on the coordinate machine. Deviations are determined in the normal direction to the work surface (Fig. 2).

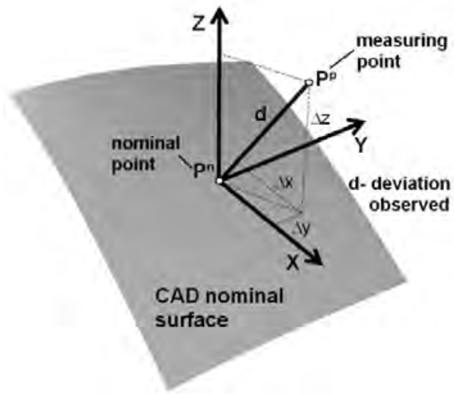


Fig. 2. Graphical representation of machining deviation

The data necessary for the determination of deviations is included in the measurement program controlling the CMM. After the measurements, information on the coordinates of the nominal points and the points observed during the measurement, observed machining deviations and directional cosines at the measuring points is available.

On the basis of this information, the correction of the surface describing the object being manufactured is made. First, the components of the observed machining deviations in the individual axes of the coordinate system are determined. The following relationships are used for calculations:

$$\Delta x_{ij} = d_{ij} * \cos\alpha_{ij}$$
$$\Delta y_{ij} = d_{ij} * \cos\beta_{ij} \qquad (1)$$
$$\Delta z_{ij} = d_{ij} * \cos\gamma_{ij}$$

where:
– Δx_{ij}, Δy_{ij}, Δz_{ij} are the components of observed machining deviations;
– d_{ij} is deviation observed at the measuring point;
– $\cos\alpha_{ij}$, $\cos\beta_{ij}$, $\cos\gamma_{ij}$ are directional cosines at the measuring points;
– i, j are coefficients describing the location of the observed point.

Determining the components of machining deviations makes it possible to calculate the corrected coordinates of points.

If the correction is carried out on 'raw' measurement data, the corrected coordinates will be determined from the following relationships:

$$x_{ij}^{\mathrm{cor}} = x_{ij}^{\mathrm{nom}} - \Delta x_{ij}$$
$$y_{ij}^{\mathrm{cor}} = y_{ij}^{\mathrm{nom}} - \Delta y_{ij} \qquad (2)$$
$$z_{ij}^{\mathrm{cor}} = z_{ij}^{\mathrm{nom}} - \Delta z_{ij}$$

where: $x_{ij}^{\mathrm{cor}}, y_{ij}^{\mathrm{cor}}, z_{ij}^{\mathrm{cor}}$ are the coordinates of the corrected surface patch and $x_{ij}^{\mathrm{nom}}, y_{ij}^{\mathrm{nom}}, z_{ij}^{\mathrm{nom}}$ are the coordinates of points on the nominal surface (CAD model).

The presented approach is the simplest, but it does not guarantee achieving the best final result. Due to the complexity of the machining and measurement process, deviations observed may contain significant effects of random phenomena. These deviations have two components: determined and random. The introduction of the measurement data filtration makes it possible to minimise the influence of random deviations on the final effect of machining deviations' correction. Equation (1) changes its form, and the components of the adjusted deviations are determined according to the following relationship:

$$\Delta x_{ij} = d_{ij}^{f} * \cos\alpha_{ij}$$
$$\Delta y_{ij} = d_{ij}^{f} * \cos\beta_{ij} \qquad (3)$$
$$\Delta z_{ij} = d_{ij}^{f} * \cos\gamma_{ij}$$

where: d_{ij}^{f} are the filtered components of observed machining deviations (determined components).

The methods of data filtration are many. The key to choosing the right method should be the simplicity of its use. The next part of the article will present an approach taken from the techniques used in image filtering.

The determined corrected coordinates are used to create the corrected surface patch. It contains information about manufacturing errors. In the construction of the patch of the corrected surface, reverse engineering techniques are used [23]. Firstly, a net of *n*x*m* corrected points is created (Fig. 3a). Next, a series of curves is interpolated onto the grid (Fig. 3b), on which, subsequently, a surface patch is applied (Fig. 3c). It is important to maintain the same *uv* parameterisation directions as in the nominal model when creating a corrected surface patch. The surface thus constructed compensates for the machining deviations that occur. It is necessary to create modified machining programs.

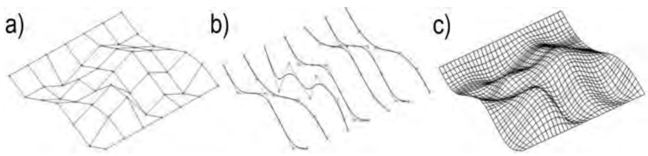


Fig. 3. Creation of the surface: (a) grid of points, (b) series of curves, (c) surface patch

10.2478/ama-2022-0013

3. EXPERIMENTAL VERIFICATION OF THE PROPOSED METHOD OF INCREASING THE ACCURACY OF MANUFACTURING

The method of correction of machining errors described in the previous part of the article was verified using the example of an object described using the NURBS surface patch (Fig. 4). The surface patch was built on a control grid built on 49 control points. The degree of B-spline functions in two directions of the *uv* surface patch parameterisation equalled 3. As can be seen in the attached drawing, the surface smoothly reproduced the shape of the control grid. The surface model was the basis for the preparation of programs controlling the machining of parts and measurement programs used to control the accuracy of manufacturing.

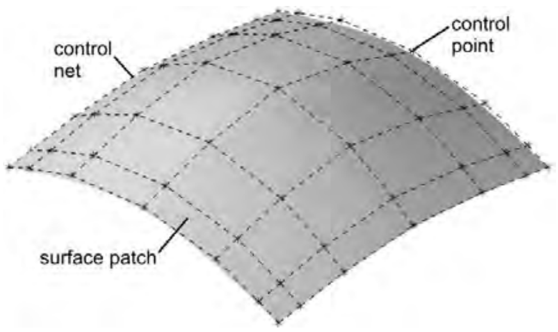


Fig. 4. A surface model describing the object being manufactured

The object was made of aluminium 2017A (Fig. 5). The milling centre control program included roughing, shaping and finishing. After the shaping treatment, a 0.3 mm allowance was left on the machined surface. To remove this allowance, a spherical cutter with a diameter of 6 mm was used to process the aluminium. Parallel tool passes with 0.2 mm spacing were programmed. Finishing was carried out at a spindle speed of 7.500 rpm and a feed of 300 mm/min. The surface produced was contained within a square of 45 mm sides.

Fig. 5. The created object

After completion of the manufacturing stage, the test object was subjected to control measurements. The control measurements were carried out on a Hexagon Metrology Global Performance measuring machine (PC-DMIS software, MPEE = 1.5 + L/333 [μm], Renishaw SP25M measuring head, 20-mm-long stylus with a spherical 2-mm-diameter tip). Detailed information on coordinate measurements of different examples of machined surfaces can be found in references [20] and [21].

Due to the fact that the shape of the manufactured object was described by the patch of the NURBS surface of the third degree, the surface produced did not show any sudden changes in shape. This allowed the use of one of the automatic surface scanning procedures available in the PC-DMIS system. This procedure, called *UVScan*, allows for an even distribution of measuring points. Finally, the control measurements were programmed for a grid of 45×45 measurement points (distance between points = 1 mm). The distribution of measuring points on the measured surface is shown in Fig. 6.

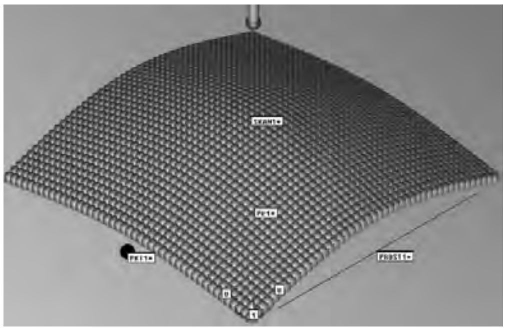


Fig. 6. Distribution of measurement points

As a result of the measurements carried out, information on 2025 deviations that were observed was obtained. The map and the distribution of spatial deviations are presented in Fig. 7. All determined deviations were within the range (−0.03, +0.045) mm.

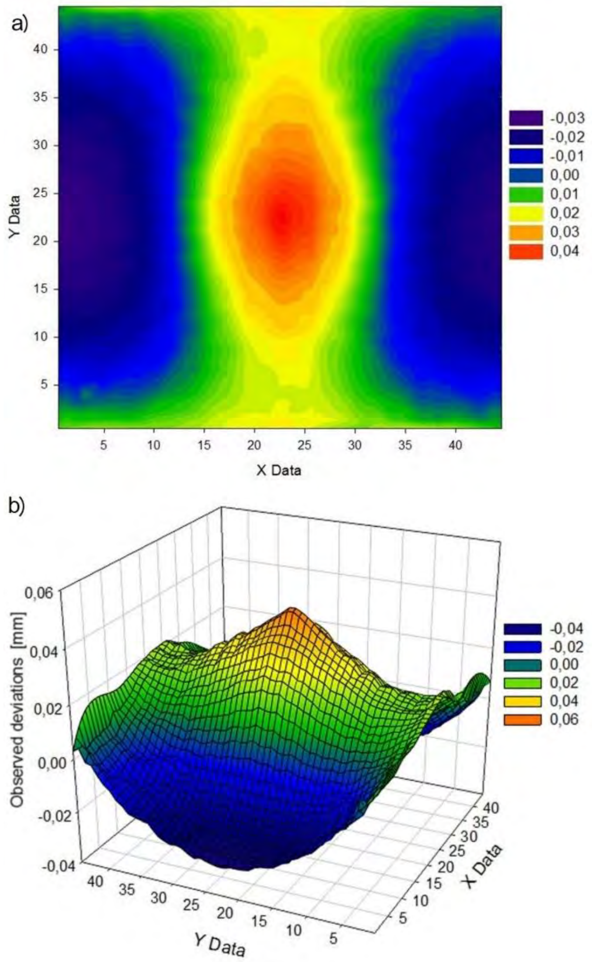


Fig. 7. Observed deviation: (a) deviation map, (b) spatial diagram

DOI 10.2478/ama-2022-0013 acta mechanica et automatica, vol.16 no.2 (2022)

Table 1 presents a summary of the coordinate measurements of the treated area. The obtained accuracy was at a relatively good level. At the same time, it leaves a margin to apply the method proposed in the article and to obtain better final effects of the treatment.

Tab. 1. Results of preliminary coordinate measurements

Deviations observed	Value (mm)
Max. negative	−0.030
Max. positive	+0.045
Average value	0.0021
Standard deviation	0.0173

3.1. Correction based on ‘raw’ measurement data

In accordance with the procedure described in the previous section of the article, the construction of the corrected part of the geometric model was started. Firstly, the nominal coordinates and coordinates of the 2025 measurement points were separated from the measurement program. Using equations (1) and (2), the corrected coordinates were determined for each point. Based on these, in the MASTERCAM system, a corrected geometric model of the produced surface was created. In the beginning, a series of 45 interpolated curves was created on the grid of points (Fig. 8a). In the next step, the surface patch was spread on the series of curves obtained (Fig. 8b).

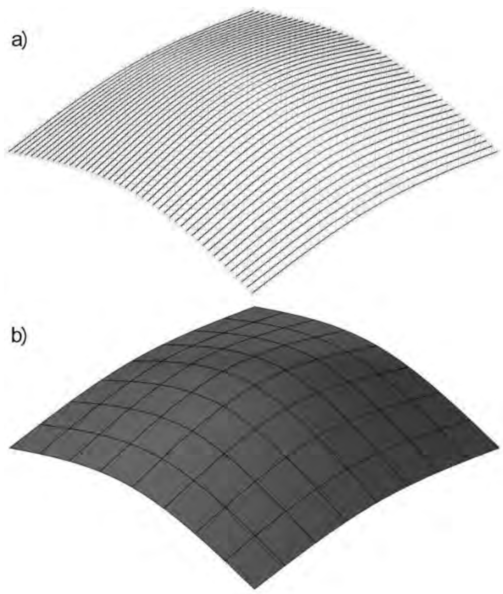


Fig. 8. Creation of the corrected geometric model of the object: (a) corrected points and series of curves, (b) patch of corrected surface

Based on the corrected geometric model of the parts, machining programs were re-created. The same tools and parameters used for the previous treatment of the object were used. The re-machined surface patch was subjected to coordinate measurements. The maps of obtained machining deviations (Fig. 9a) and their spatial distribution (Fig. 9b) indicate a significant improvement in the accuracy of manufacturing.

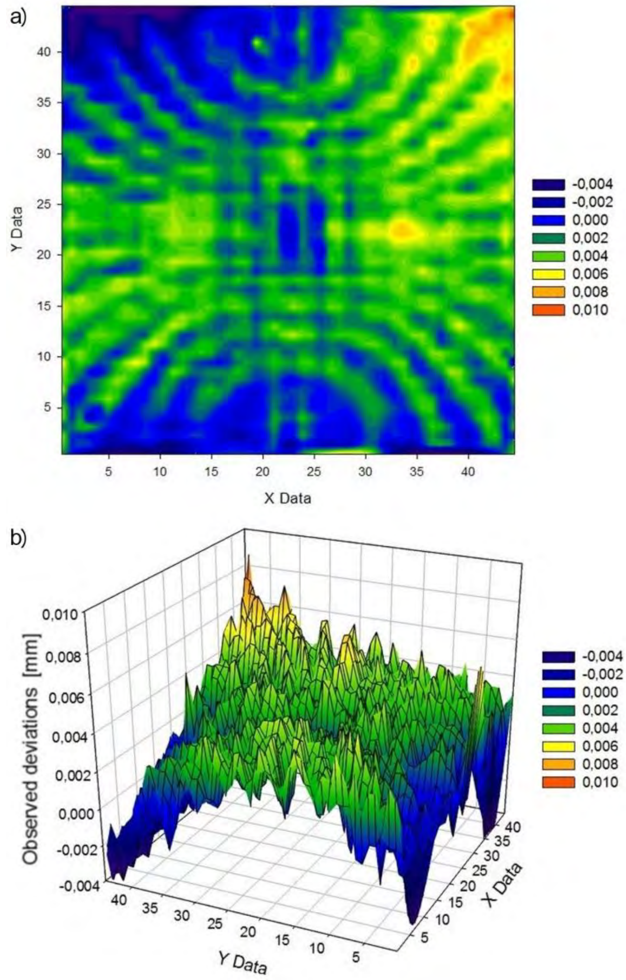


Fig. 9. Deviations observed after correction on ‘raw’ measurement data: (a) deviation map, (b) spatial diagram

Table 2 contains the most important information on the results obtained using the method of increasing accuracy proposed in the article. A significant decrease in the maximum positive and negative deviations is observed. Also, the mean value and standard deviation indicate a small spread of the results obtained with respect to nominal data.

Tab. 2. Results of machining deviations correction

Deviations observed	Value (mm
Max. negative	−0.005
Max. positive	+0.009
Average value	0.0023
Standard deviation	0.0018

3.2. Correction based on filtered measurement data

Another approach in the construction of a revised surface model of the manufactured object is the introduction of a filtration procedure for measurement data. Its purpose is to minimise the influence of random components of observed machining deviations on the final effect of the procedure of increasing the accuracy of manufacturing.

This article proposes the use of procedures used in image filtration [24, 25]. The use of filters in the processing of measurement data means that the values of points from its surroundings are taken into account in calculating the new point value. Each measuring point from the environment contributes a weight during the calculation.

These weights are saved in the form of a mask. Typical mask sizes are 3×3, 5×5 and 7×7. The dimensions of the masks are usually odd because the measuring point in the middle represents the point at which the filter conversion operation is performed. The following is an example of data filtration based on a 3×3 filter.

$f_{-1,-1}$	$f_{0,-1}$	$f_{1,-1}$
$f_{-1,0}$	$f_{0,0}$	$f_{1,0}$
$f_{-1,1}$	$f_{0,1}$	$f_{1,1}$

The deviations observed in the measurement points have the form of a grid consisting of *n* columns and *m* rows. The new value of the $d_{i,j}$ elements is calculated with the coordinates (i, j) according to the following procedure. First, the weighted sum of the point component and all neighbours is calculated according to the weights indicated by the filter mask.

$$d'_{ij} = f_{-1,-1} * d_{i-1,j-1} + f_{0,-1} * d_{i,j-1} + f_{1,-1} * d_{i+1,j-1} + f_{-1,0} * d_{i-1,j} + f_{0,0} * d_{i,j} + f_{1,0} * d_{i+1,j} + f_{-1,1} * d_{i-1,j+1} + f_{0,1} * d_{i,j+1} + f_{1,1} * d_{i+1,j+1} \tag{4}$$

The sum obtained in this way is divided by the sum of all mask weights, if it is different from 0.

$$d_{ij}^{f} = \frac{d'_{ij}}{f_{-1,-1}+f_{0,-1}+f_{1,-1}+f_{-1,0}+f_{0,0}+f_{1,0}+f_{-1,1}+f_{0,1}+f_{1,1}} \tag{5}$$

The process of normalising the component value of the observed machining deviation results in a smoother deviation distribution being obtained and minimises the influence of random components on the final result of the machining error correction. In order to test the procedure proposed in the article, a 5×5 mask was used. Due to the even distribution of the measurement points, it was assumed that the impact of all points surrounding the deviation being processed is the same (all weights equal to 1). The format of the mask used is presented below.

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

The effect of data filtration is presented in Fig. 10. A modified map of machining deviations is presented in Fig. 10a. Comparing it with the map of 'raw' deviations (Fig. 7), a significant smoothing of contours, representing individual levels of machining deviations, can be seen. This is due to the separation of random components (Fig. 10b) generated during the machining and measurement process.

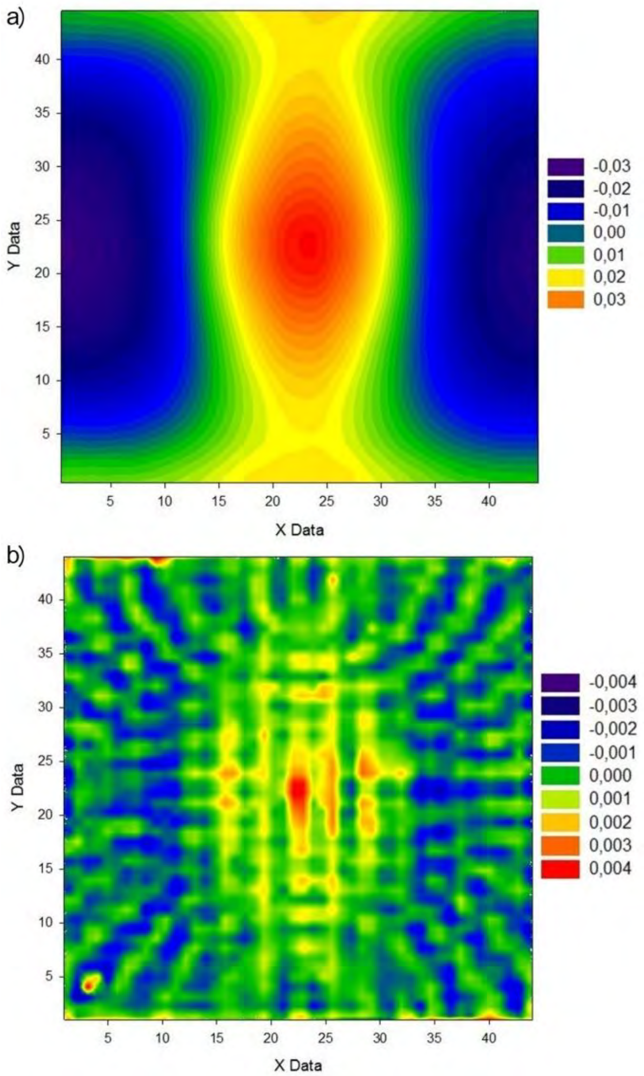


Fig. 10. Filtering effect of observed machining deviations:
(a) map of components after filtration,
(b) map of filtered components

Table 3 contains numerical values illustrating the change of 'raw' deviations after applying data filtration. A change in the maximum values of deviations after filtration can be observed. The filtered components are in the interval (−0.0027; 0.005), and their dispersion relative to the mean value (standard deviation) is insignificant.

Tab. 3. Filtration results of observed machining deviations

Deviations observed (mm)	'Raw' deviations	Corrected deviations	Filtered components
Max. negative	−0.030	−0.028	−0.00271
Max. positive	+0.045	+0.040	+0.005
Average value	0.0021	0.0021	−2.1E−05
Standard deviation	0.0173	0.0168	0.0009

On comparing the map of machining deviations obtained after the correction of machining based on 'raw' deviations (Fig. 9a) and the map of filtered components (Fig. 10b), one can notice some similarity. This indicates the effect of random components of machining deviations on the final effect of increasing machining accuracy. Therefore, it can be assumed that the filtration of the coordinate measurement results carried out, and the correction of manufacturing errors based on the corrected deviations, will positively influence the final result of the procedure.

After the data filtration procedure, the adjusted machining deviations were determined using equations (1) and (3). The further course of action was analogous to that of ‘raw’ measurement data. On the basis of 2025 corrected points, a series of 45 curves was created, on which the surface patch was spread. On the basis of the corrected surface patch, the machining programs and the surface patch were regenerated. The process ended with coordinate control measurements. The final results are presented in Fig. 11. The map view of the obtained machining deviations (Fig. 11a), in combination with the post-correction deviation map on the raw data base (Fig. 9a) indicates an improvement in surface accuracy. The results obtained after the correction of the filtered data are better than in the case of correction for ‘raw’ measurement data. The map of machining deviations does not show such a similarity to the map of filtered components (Fig. 10b). This indicates minimisation of the influence of random components of observed machining deviations on the final effect of increasing the accuracy of manufacturing.

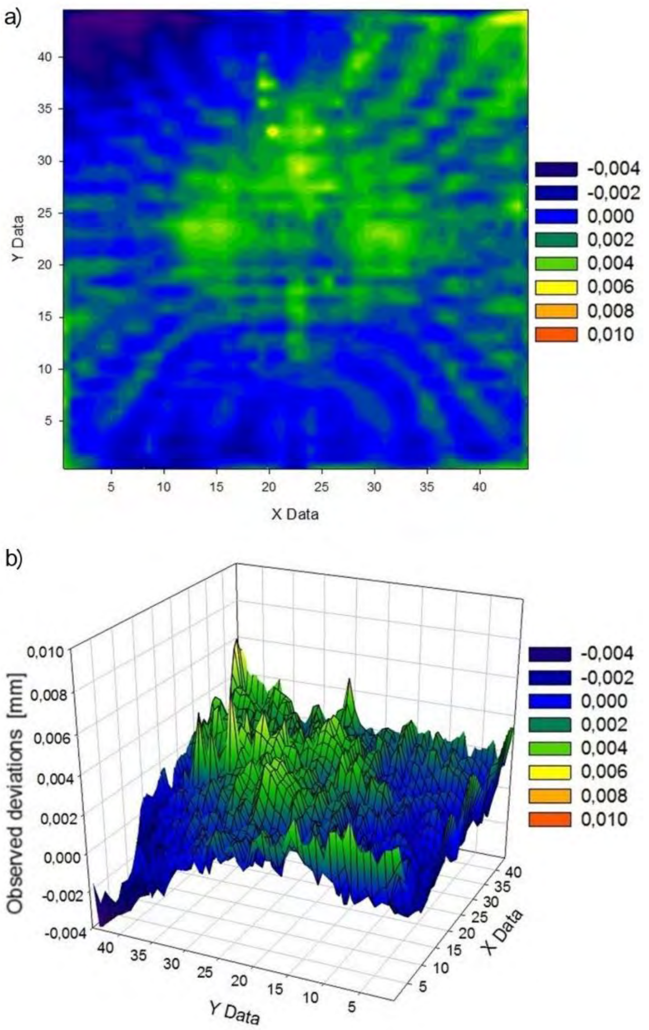


Fig. 11. Deviations observed after correction of filtered measurement data: (a) deviation map, (b) spatial diagram.

Table 4 presents the results of repeated corrections of machining deviations. The obtained machining deviations are in this range in the interval (−0.004; 0.007) and are smaller than those obtained in the correction for ‘raw’ measurement data. Other values, that is, the mean value and standard deviation, also decreased. This indicates a greater convergence of the surface area produced with its nominal CAD model. They also show a smaller spread over the nominal data, as illustrated in Fig. 11b.

Tab. 4. Results of correction of machining deviations based on filtered data

Deviations observed	Value (mm
Max. negative	−0.004
Max. positive	+0.007
Average value	0.0013
Standard deviation	0.0014

4. CONCLUSIONS

The implementation of the method of increasing the accuracy of manufacturing of shaped surfaces presented in this article allowed the accuracy of manufacturing to be significantly increased. Table 5 presents the results observed before and after the correction process using two methods. In both cases, they indicate the effectiveness of the proposed procedure. The presented results show a clear decrease in the maximum observed machining deviations. The introduction of filtration of measurement data made it possible to improve the final result. The observed maximum machining deviations are the smallest in this case. The deviation of standard deviations and their mean indicates an additional positive effect. The surface produced shows the greatest similarity to the nominal CAD model. The measurement data filtration makes it possible to reduce the influence of random components of observed machining deviations on the process of increasing the accuracy of manufacturing.

Tab. 5. List of obtained results

	‘Raw’ deviations (mm)	Correction 1 – ‘raw’ deviations (mm)	Correction 2 – filtered deviations (mm)
Max. negative	−0.030	−0.005	−0.004
Max. positive	0.045	0.009	0.007
Average value	0.0022	0.0023	0.0013
Standard deviation	0.0173	0.0018	0.0014

The implementation of the presented method of correction of manufacturing errors is relatively simple. It is based on typical hardware and software used in enterprises (CAD/CAM systems, CNC machine tools, CMMs). The implementation of machining error correction is additionally facilitated by the parametric linking of technological and geometric data in modern CAD/CAM systems. This means that once developed, technological data do not need to be re-entered into the system. As a consequence, after rebuilding the geometric model of the object, the tool path is automatically rebuilt.

REFERENCES

1. Ramesh R, Mannan MA, Poo AN. Error compensation in machine tools - a review. Part I: geometric, cutting-force induced and fixture dependent errors. Int J Mach Tool Manu. 2000; 40: 1235–1256.
2. Ramesh R, Mannan MA, Poo AN. Error compensation in machine tools - a review. Part II: thermal errors. Int J Mach Tool Manu. 2000; 40: 1257–1284.
3. Wenjie T, Weiguo G, Dawei Z, et al. A general approach for error modeling of machine tools. Int J Mach Tool Manu. 2014; 79: 17–23.
4. Zuriani U, Ahmed A, Sarhan, Mardi NA, et al. Measuring of positioning, circularity and static errors of a CNC Vertical Machining Centre for validating the machining accuracy. Measurement. 2015; 61: 39–50.
5. Zhouxiang J, Bao S, Xiangdong Z et al. On-machine measurement of location errors on five-axis machine tools by machining tests and a laser displacement sensor. Int J Mach Tool Manu. 2015; 95: 1–12.
6. Ibaraki S, Sawada M, Matsubara A, et al. Machining tests to identify kinematic errors on five-axis machine tools. Prec Eng. 2010; 34: 387–398.
7. Zhengchun Du, Shujie Zhang, Maisheng H. Development of a multi-step measuring method for motion accuracy of NC machine tools based on cross grid encoder. Int J Mach Tool Manu. 2010; 50: 270–280.
8. Vahebi Nojedeh M, Habibi M, Arezoo B. Tool path accuracy enhancement through geometrical error compensation. Int J Mach Tool Manu. 2011; 51:471–482.
9. Xiaoyan Z, Beizhi L, Jianguo Y, et al. Integrated geometric error compensation of machining processes on CNC machine tool. Procedia CIRP. 2013; 8: 135–140.
10. Lasemi A, Xue D, Gu P. Accurate identification and compensation of geometric errors of 5-axis CNC machine tools using double ball bar. Measurement Science and Technology. 2016; 27 (5): 055004.
11. Zhang X, Zhang J, Zheng X, Pang B, Zhao W. Tool orientation optimization of 5-axis ball-end milling based on an accurate cutter/workpiece engagement model. CIRP Journal of Manufacturing Science and Technology. 2017; 19: 106-116.
12. Kim YJ, Elber G, Barton M, Pottmann H. Precise gouging-free tool orientations for 5-axis CNC machining. Computer-Aided Design. 2015; 58: 220-229.
13. Barton M, Bizzarri M, Rist F, Sliusarenko O, Pottmann H. Geometry and tool motion planning for curvature adapted CNC machining. ACM Transactions on Graphics. 2021 Aug; 40 (4): 1–16. https://doi.org/10.1145/3450626.3459837
14. Hansel A, Yamazaki K, Konishi K. Improving CNC machine tool geometric precision using manufacturing process analysis techniques. Procedia CIRP. 2014; 14: 263–268.
15. Habibi M, Arezoo B, Vahebi Nojedeh M. Tool deflection and geometrical error compensation by tool path modification, Int J Mach Tool Manu. 2011: 51: 439–449.
16. Ryu SH, Chu CN. The form error reduction in side wall machining using successive down and up milling. Int J Mach Tool Manu. 2005; 45: 1523–1530.
17. Yang MY, Choi JG. A tool deflection compensation system for end milling accuracy improvement. J Manuf Sci Eng. 1998; 120: 222–229.
18. Landon Y, Segonds S, Mousseigne M, et al. Correction of milling tool paths by tool positioning defect compensation. Proc Inst Mec. Eng B. 2003; 217: 1063–1073.
19. Myeong-Woo Cho, Tae-il Seo, Hyuk-Dong Kwon. Integrated error compensation method using OMM system for profile milling operations. J Mater Process Techol. 2003; 136: 88–99.
20. Poniatowska M, Werner A. Fitting spatial models of geometric deviations of free-form surfaces determined in coordinate measurements. Metrol Meas Syst. 2010; 17: 599–610.
21. Poniatowska M, Werner A. Simulation tests of the method for determining a CAD model of free-form surface deterministic deviations. Metrol Meas Syst. 2012; 19: 151-158.
22. Poniatowska M. Free-form surface machining error compensation applying 3D CAD machining pattern model. Comput Aided Design. 2015; 62: 227–235.
23. Werner A, Skalski K, Piszczatowski S, et al. Reverse engineering of free-form surfaces. J Mater Process Technol. 1998; 76: 128-132.
24. Kawasaki T, Jayaraman PK, Shida K, et al. An image processing approach to feature-preserving B-spline surface fairing. Comput Aided Design. 2018; 99: 1–10.
25. Wang Z, Wang H. Image smoothing with generalized random walks: Algorithm and applications. Appl Soft Comput 2016; 46: 792–804.

Acknowledgement: The work is supported by the Polish Ministry of Education and Science under the statute activity under the project No. WZ/WM-IIM/2/2020.

Andrzej Werner: https://orcid.org/0000-0002-3768-5395

DOI 10.2478/ama-2022-0014 acta mechanica et automatica, vol.16 no.2 (2022)

BEHAVIOUR OF A NON-NEWTONIAN FLUID IN A HELICAL TUBE UNDER THE INFLUENCE OF THERMAL BUOYANCY

Mohamed RAMLA*, Houssem LAIDOUDI, Mohamed BOUZIT****

*Faculty of Physics, Oran University of Science and Technology - Mohamed Boudiaf, BP 1505, El-Menaouer, Oran, 31000, Algeria
**Laboratory of Sciences and Marine Engineering, Faculty of Mechanical Engineering, Oran University of Science and Technology - Mohamed Boudiaf, BP 1505, El-Menaouer, Oran, 31000, Algeria

ramla_med@yahoo.fr, houssem.laidoudi@univ-usto.dz, bouzit-mohamed@yahoo.fr

received 1 December 2021, revised 4 February 2022, accepted 7 February 2022

Abstract: This work is an evaluative study of heat transfer in the helical-type heat exchanger. The fluid used is non-Newtonian in nature and is defined by Oswald's model. The work was performed numerically by solving each of the Navier–Stokes equations and the energy equation using the package ANSYS-CFX. Following are the aspects that have been dealt with in this paper: the effects of thermal buoyancy, fluid nature and the tube shape on the heat transfer, and the fluid comportment. The interpretation of the obtained results was done by analyzing the isotherms and the streamlines. The mean values of the Nusselt number were also obtained in terms of the studied parameters. The results of this research enabled us to arrive at the following conclusion: the intensity of thermal buoyancy and the nature of the fluid affect the heat transfer distribution but keep the overall rate of heat transfer the same.

Key words: helical heat exchanger, mixed convection, forced convection, power-law fluids, Nusselt number

1. INTRODUCTION

The helical shape of tubes is one of the techniques used in many industrial applications, primarily heat exchangers. It is also used in the production of industrial food stuffs such as jam, chocolates and others. Because of the importance of this type of tubes, many researchers have devoted a set of works showing the behaviour of the fluid inside the tube and the effect of this behaviouron the quality of heat transfer.

[1] demonstrated the possibility of improving heat transfer within a spiral tube by creating columns within the inner walls of the tube. The fluid investigated in this research is of the non-Newtonian type mixed with Al_2O_3 nanoparticles. [2] numerically simulated the passage of nanofluids and studied the heat transfer within aspiral tube,which contained ribs. The study showed that the presence of ribs enhances heat transfer by 12%. [3] exploited the finite volume method to compute the heat transfer in a helical tube. The geometrical dimensions of the tube were designed for a micro-system. The work mainly examined the effect of velocity and the length of the helix's pitch on the thermal behaviour. Through this work, it was concluded that the helical length of the tube has an effect on the thermal activity. [4] presented an entire experimental work on the behaviour of a two-phase stream in a helical tube. The use of this type of two-phase stream increases the heat transfer coefficient by up to 20%. [5] worked on a multiple-row helical tube, with the purpose of deriving the correlation that shows the changes in heat transfer in terms of the elements acting on it. [6] studied the effect of an external current around a helical tube,which was used for heat transfer. It has been found that this type of tube reduced the dead zone in the back. [7] determined the correlation between the Nusselt number and the friction factor for helical tubes. The thermal transfer was forced convection. [8] submitted a detailed report on a study conducted on helical tubes specially used in the heat pump. This work explained in detail the thickness of the resulting ice in terms of the conditions used. [9] did experimental research on helical tubes, where in they compared the geometry of the tubes. The new proposed shape included a group of corrugates on the tube. The obtained results were compared with those from smooth-walled tubes, and an increase in heat transfer was observed. [10] used the analytical method to study the flow within a helical tube of circular cross-section. The point studied in this work was the shape of the bend. [11] replaced the circular cross-section of the tube with a square one. Their work also included a study on the heat transfer of the forced convection type. [12] proposed an optimal design of a helical exchanger that can beused for absorption processes. The research was performed numerically using the commercial code ANSYS-Fluent. [13] presented a work on helical pipes for the measurement of frictional pressure losses. The results of this work were obtained experimentally and numerically as well. [14] further presented experimental results on the use of helical tubes for measuring the rheological properties of a complex fluid.Various researchers [15-17] also performed experiments on the curved tube, which showed an unusual disturbance of the flow.

In addition, the helical shape has a very important streamline compared to direct barriers, and this has been confirmed by most recent studies [18-21]. It is used in the structure of the heat exchangers to enhance the cooling process.

[22] made geometrical changes to the cross-section of the helical tube, from a simple circular shape to an organised circular shape of internal protuberances. In addition to this, a nanofluid

DOI 10.2478/ama-2022-0014

was used. Al_2O_3 nanoparticles were added to the basic fluid (water). There was an enhancement of heat transfer when the volume fraction reached 0.5%. [23] studied the turbulent regime of a flow in helical heat exchangers. The K-epsilon model was used for the mobilisation of the regime. The cross-section of the tube was semicircular. The results showed that this type of cross-section is highly recommended. Further, [24] performed numerical research on the turbulent flow in a sector-by-sector type helical tube . It was concluded that the semicircular form of the cross-section is more efficient for application in heat exchangers. [25] conducted studies on a double-cooling thermal exchanger,which is used in advanced aero-engine. This work aimed to find the best design of the tube that increases the effectiveness of cooling of the air that passes around the tube. [26] also introduced a numerical study on the helical tube, where the external and internal impacts of thermal transition are based on conjugate thermal boundary layers. [27] numerically compared a new type of helical tube with a single U-tube for application in heat exchangers. It was found that the helical type is more efficient than the U-tube. [28] used SiO_2/water and Al_2O_3/water nanofluids in the helical tube of a thermal exchanger.The research examined the concentration of the nanoparticles and a few geometrical configurations of the tube. The results of the work were given in the form of a correlation that relates the Nusselt number to the parameters considered in the study.

Because this type of tube is very important in thermal applications, several new works have been published on this topic to explain the scientific view on it [29-35]. In general, the main objective of these studies is to research the geometrical shape that increases the value of heat transfer.

[36] conducted an analytical study on double-helical tubes to achieve the most effective heat transfer. The studied elements were the taper angle and the cross-section. [37] simulated the flow of Al_2O_3/water nanofluids in a helical tube to increase the thermal transfer. The results proved that the presence of nanoparticles enhances thermal transfer. [38] performed a numerical simulation of the flow around a helical tube. The tube contained fins of various geometric shapes.The results showed that the presence of these fins increases thermal transfer.

Through our observation of these previous works, it becomes clear to us that these types of tubes have efficiency in increasing heat transfer, completely different from that of straight tubes. Moreover, most studies focused on forced heat transfer only. Further, the fluids used either had simple behaviour or were nanofluids. There is a lack of work that combines helical tubes and the non-Newtonian behaviour of a fluid in the presence of thermal buoyancy,which produces mixed heat transfer. Therefore, our work is a three-dimensional numerical simulation of a complex fluid inside a helical tube, with thermal buoyancy acting in the opposite direction of the fluid movement. The items considered here are as follows: the flow velocity at the inlet of the tube, the power law index that determines the rheological nature of the fluid, the intensity of thermal buoyancy and the pitch between the loops of the helical tube.

2. DESCRIPTION OF STUDIED DOMAIN

The tube under consideration is shown in Fig. 1. The tube has a helical shape containing three rings. The gap (the pitch) between the rings is constant and is given by (*B*). The inner diameter of the tube is given by the index (*d*),while the diameter of the tube ring is given by (*D*). These geometrical values are given in the form of ratios represented in Tab. 1. The studied fluid enters the tube at a constant velocity and low temperature (T_{in}). The walls of the tube have a high and constant temperature value (T_w). The difference in temperature between the fluid and the walls of the tube leads to heat transfer between the hot side and the cold side. The fluid velocity is given by the Reynolds number. The thermal buoyancy force acts in the direction (*Z*). Its intensity is described by the value of the Richardson number (Ri).

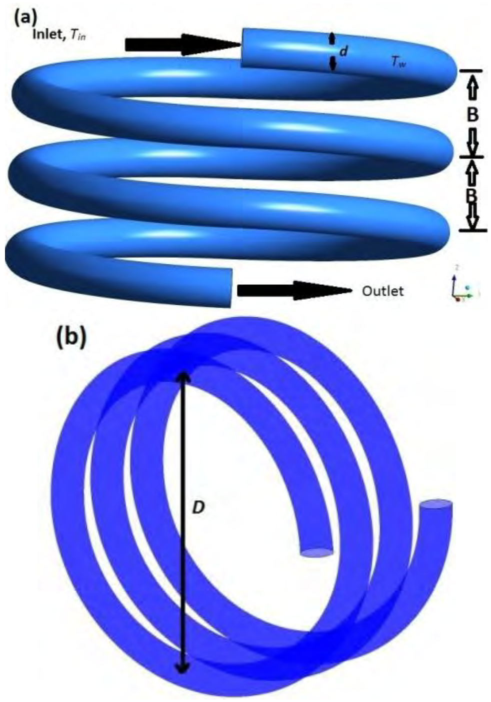


Fig. 1. Geometrical characteristics of the studied tube:
(a) first view; (b) second view (B, the pitch between the rings; d, inner diameter of the tube; D, diameter of the tube ring)

Tab. 1. Geometric rational values of the tube

Ratio	*d/D*	*B/D*
Value	0.1	0.2;0.4

3. MATHEMATICAL FORMULATIONS

In this section, we present the basic equations that must be solved numerically to obtain the numerical simulation of the desired study. These equations are as follows:

$$\frac{\partial U}{\partial X}+\frac{\partial V}{\partial Y}+\frac{\partial W}{\partial Z}=0 \quad (1)$$

$$U\frac{\partial U}{\partial X}+V\frac{\partial U}{\partial Y}+W\frac{\partial U}{\partial \mathrm{Z}}=-\frac{\partial P}{\partial \mathrm{X}}+\frac{1}{\mathrm{Re}}\left(\frac{\partial \tau^*_{xx}}{\partial X}+\frac{\partial \tau^*_{yx}}{\partial Y}+\frac{\partial \tau^*_{zx}}{\partial \mathrm{Z}}\right) \quad (2)$$

$$U\frac{\partial V}{\partial X}+V\frac{\partial V}{\partial Y}+W\frac{\partial V}{\partial \mathrm{Z}}=-\frac{\partial P}{\partial \mathrm{Y}}+\frac{1}{\mathrm{Re}}\left(\frac{\partial \tau^*_{xy}}{\partial X}+\frac{\partial \tau^*_{yy}}{\partial Y}+\frac{\partial \tau^*_{zy}}{\partial \mathrm{Z}}\right) \quad (3)$$

$$U\frac{\partial W}{\partial X}+V\frac{\partial W}{\partial Y}+W\frac{\partial W}{\partial \mathrm{Z}}=-\frac{\partial P}{\partial \mathrm{Z}}+\frac{1}{\mathrm{Re}}\left(\frac{\partial \tau^*_{xz}}{\partial X}+\frac{\partial \tau^*_{yz}}{\partial Y}+\frac{\partial \tau^*_{zz}}{\partial \mathrm{Z}}\right)+\mathrm{Ri}\theta \quad (4)$$

$$U\frac{\partial \theta}{\partial X}+V\frac{\partial \theta}{\partial Y}+W\frac{\partial \theta}{\partial \mathrm{Z}}=\frac{1}{\mathrm{Pr.Re}}\left(\frac{\partial^2 \theta}{\partial X^2}+\frac{\partial^2 \theta}{\partial Y^2}+\frac{\partial^2 \theta}{\partial Z^2}\right) \quad (5)$$

The dimensionless quantities of the variable in the above equations are given as follows:

$$(X,Y,Z)=\frac{(x,y,z)}{d},(U,V,W)=\frac{(u,v,w)}{\mathrm{u_{in}}},P=\frac{p}{\rho \mathrm{u_{in}^2}},\theta=\frac{(T-T_{\mathrm{in}})}{(T_{\mathrm{w}}-T_{\mathrm{in}})} \quad (6)$$

$$\tau^*=\frac{\tau}{m(\frac{u_{in}}{d})^n} \quad (7)$$

The dimensionless numbers Re, Ri and Pr are given by the following expressions:

$$\mathrm{Re}=\frac{\rho(u_{in})^{2-n}d^n}{m} \quad (8)$$

$$\mathrm{Pr}=\frac{mc_p}{k\,(u_{in}/d)^{n-1}} \quad (9)$$

$$\mathrm{Ri}=\frac{\mathrm{Gr}}{\mathrm{Re}^2}=\frac{g\beta_T\Delta T d^3}{(u_{in})^2} \quad (10)$$

$$\mathrm{Gr}=g\beta_T\Delta T d^3[\frac{\rho}{m}(\frac{u_{in}}{d})^{1-n}]^2 \quad (11)$$

Through the Ostwald model, the viscosity is defined as follows:

$$\mu=m(\frac{I_2}{2})^{(\frac{n-1}{2})} \quad (12)$$

where I_2 is given by the following expression:

$$\frac{I_2}{2}=2(\frac{\partial u}{\partial x})^2+2(\frac{\partial v}{\partial y})^2+2(\frac{\partial w}{\partial z})^2+(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}+\frac{\partial w}{\partial z})^2 \quad (13)$$

Generally, the fluid transfers heat in conductive and convective forms at the same time. The ratio between these two methods is determined by the Nusselt number, Nu. There is a local Nu number specific to each point in the analysed area, and there is an average value. These values of the Nusselt number are given respectively by the following two expressions:

$$\mathrm{Nu_L}=(\frac{\partial\theta}{\partial n_s})_{wall} \quad (14)$$

$$\mathrm{Nu}=\frac{1}{A}\int \mathrm{Nu_L}dA \quad (15)$$

where nsis the vector normal to the surface and *dA* is the elementary surface.

For every numerical work, a set of initial boundary conditions must be set for the simulation to occur. These boundary conditions are given as follows:

– At the inlet of the helical tube: cold temperature with uniform velocity:

$$U=0,V=1,W=0,\theta=0 \quad (16)$$

– On the tube walls: no-slip condition and hot temperature:

$$U=0,V=0,W=0,\theta=1 \quad (17)$$

– At the outlet tube: Neumann boundary condition is applied:

$$\frac{\partial U}{\partial X}=0,\frac{\partial V}{\partial X}=0,\frac{\partial W}{\partial X}=0,\frac{\partial\theta}{\partial X}=0 \quad (18)$$

4. NUMERICAL STEPS AND VALIDATION TEST

To achieve this simulation, several stages must be passed, which are arranged as follows: creating the grid for the studied domain, checking the effectiveness of this grid and checking the numerical methodology used for solving the mean equations.

The grid was created using Gambit. The grid cells were selected in quadratic form, as shown in Fig. 2. The grid contains 115.8000 elements. The effectiveness of the grid in finding the exact calculation depends mainly on the number of cells forming the entire grid. Three grids with different number of cells were created to calculate the Nu number for each case, so that the number doubles each time. The values of the elements together with the results of the calculation are shown in Tab. 2. The results are calculated for *n* = 1, Re = 100, Pr = 50 and Ri = 0. It is noted from Tab. 2 that there is stability for the Nu number when the number of grid elements exceeds the value 115.8000. Therefore, it can be concluded that the second grid G2 is suitable for this research.

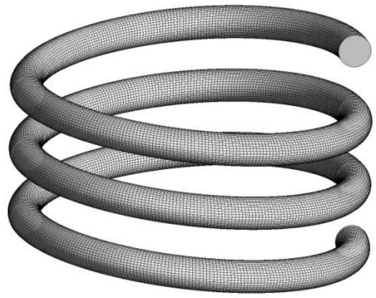


Fig. 2. The structural grid of the studied geometry

Tab. 2. Grid independency test

Grid	Elements	Nu	Difference %
G1	579000	1.478521	2.95
G2	115.8000	1.523521	0.79
G3	2316000	1.511453	–

Nu, Nusselt number

The numerical calculation was done using the finite volume method. This method transforms the differential equations into a matrix system and then solves them using numerical methods. The SIMPLEC algorithm was used for coupling pressure and velocity. However, the high-resolution discretisation scheme was used for solving the convective term.

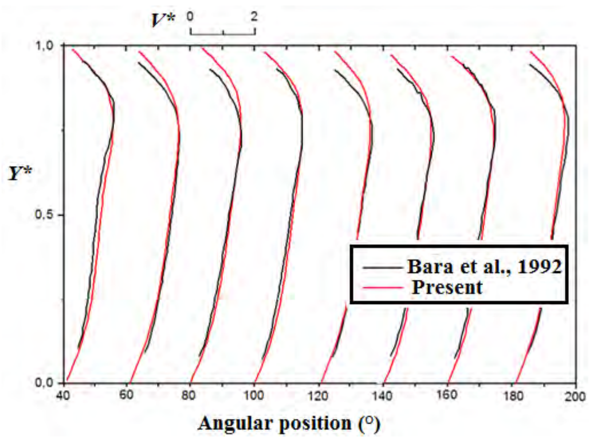


Fig. 3. Comparison of the results of this study with the experimental work of [39]

In addition to this, the accuracy of the results of this method has been proven by comparing the results of this method with the

DOI 10.2478/ama-2022-0014

results of previous works obtained under the same conditions. The first comparison was made with an experiment of the flow inside a curved tube at an angle of 180°. The work was previously done by [39]. Both results are represented in Fig. 3. A very noticeable agreement is found between the results of the experimental and the numerical works. The second comparison was made with the work carried out by [40]. It is a numerical simulation of a fluid in a 180° bent channel of square cross-section. The results of this comparison are represented in Fig. 4. Good agreement is found between the results. Through the results of the comparison study, it is possible to ascertain the effectiveness of this code in arriving at very accurate values.

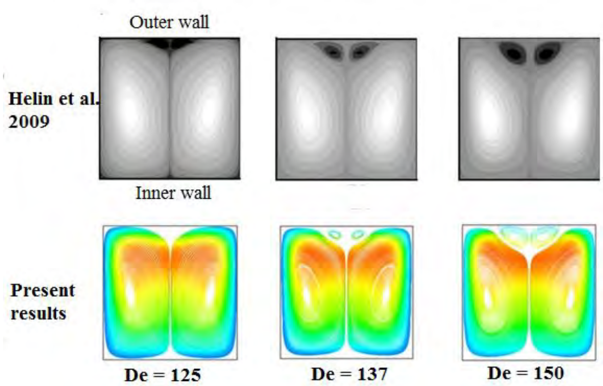


Fig. 4. Comparison of the results of this study with the numerical work of [40]

5. RESULTS AND DISCUSSION

We performed a set of numerical simulations to study a set of factors affecting the behaviour of a complex fluid inside a helical tube. The relevant parameters studied are as follows: the index (*n*),which changes the rheological properties of fluids, takes three values, viz., 0.6, 1 and 1.6; the value of the fluid's velocity determined by the Reynolds number, which varies in the range of 10–600; the intensity of the thermal buoyancy, which is determined by the Richardson number (Ri), which takes the values 0 and 1; and finally, the length of the step between the loops. Accordingly, two values have been tested.

It is worth noting that the analysis of the thermal and dynamic behaviour of the fluid is done by analysing and interpreting the streamlines and the isotherms.Moreover, the mean Nusselt values are presented in the form of curves in terms of the studied pertinent parameters.

5.1. Effect of the Re and Ri numbers

In Figs. 5, 6 and 7, the effects of Ri and Re on the dynamic behaviour of the fluid and its effect on the thermal pattern at $n = 1$ (Newtonian fluid) are presented.

Before we proceed to present the results, we note that the streamlines and isotherms are presented in cross-sections of the angles 90° and –90° together, and the angles 180° and –180° together.

Fig. 5 shows the behaviour of the fluid in the cross-section of angles 90° and -90° for all rings. The behaviour of the fluid is presented in terms of the Re and Ri numbers. It is noticed that when Ri = 0,on increasing the value of the Reynolds number, the flow becomes less stable and the flow is deflected towards the external tube wall. This, of course, is due to the effectof centrifugal force on the flow. The higher the value of Re, the higher is the velocity of the fluid, and this increases the deflection of the flow of the fluid due to the centrifugal force. On the other hand, in the presence of thermal buoyancy force (Ri = 1) effect, two steady vortices appear in this cross-section for Re = 10, but they quickly disappear by raising the value of Re. This phenomenon can be explained as follows: the presence of thermal buoyancy in the opposite direction to the movement of the fluid makes the diameter of the tube smaller, and this raises the seed of the flow in this position (angle 90° and –90°), and accordingly Dean vortices appear.

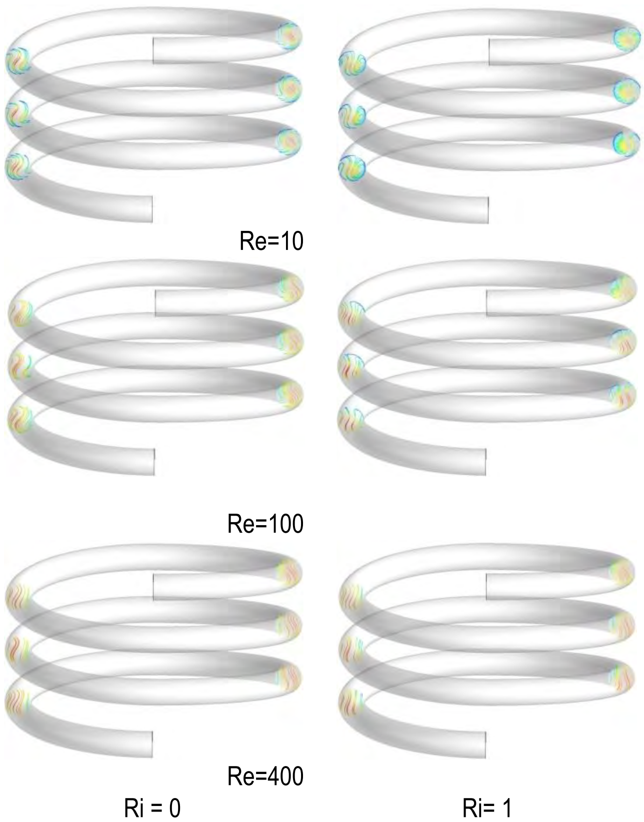


Fig. 5. Streamlines in the cross-sections of the angles 90° and –90° in terms of Re and Ri at $n = 1$ (Re, Reynolds number; Ri, Richardson number)

Fig. 6 shows the same phenomenon, but at the angles 180° and –180°. In this position and in the absence of thermal buoyancy (Ri = 0), the fluid flow is less stable than in the previous positions (angles 90° and –90°), and this is indicated by the presence of Dean vortices in all rings with these angles. It is also noted that the size of these vortices increases with the increase in the value of fluid velocity. In the presence of thermal buoyancy effect (Ri = 1), a slight rotation is observed at the location of the vortices at these angles. This is a result of the transfer of heated particles of the fluid upwards under the influence of thermal buoyancy force.

Fig. 7 shows the isotherms in terms of the studied parameters (Re and Ri) for the angles 90° and –90°. It is noted that the isotherm distribution is not uniform and this is due to the effect of

DOI 10.2478/ama-2022-0014

the centrifugal force on the flow, and from this, it can be deduced that the local heat transfer is different from one point to another. In the first case, i.e. without the effect of thermal buoyancy, it turns out that the outer wall transfers heat with fluid more than the remaining parts of the tube. In the second case, in the presence of thermal buoyancy, it becomes clear that the flow deviates a little to the bottom of the cross-section, making the bottom side more efficient in heat transfer compared to the other sides.

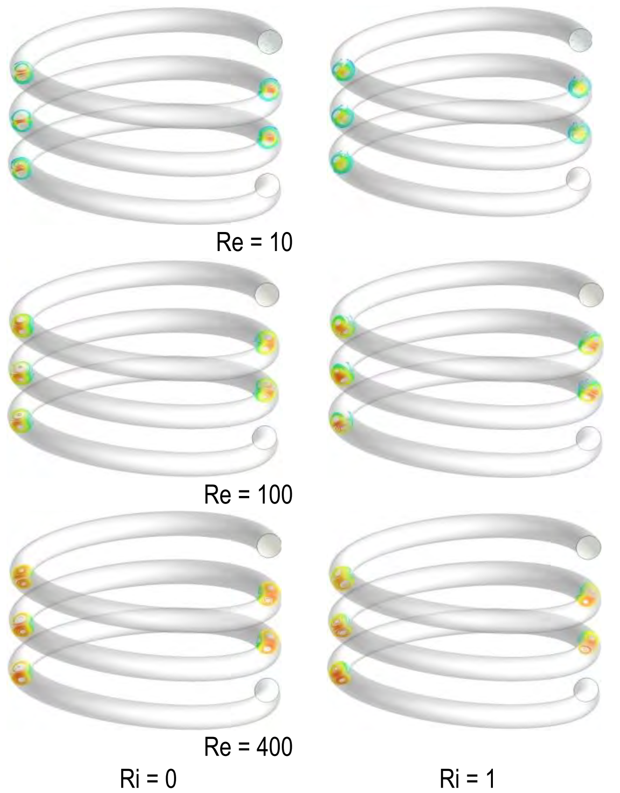


Fig. 6. Streamlines in cross-sections of the angles 180° and –180° in terms of Re and Ri at n = 1 (Re, Reynolds number; Ri, Richardson number)

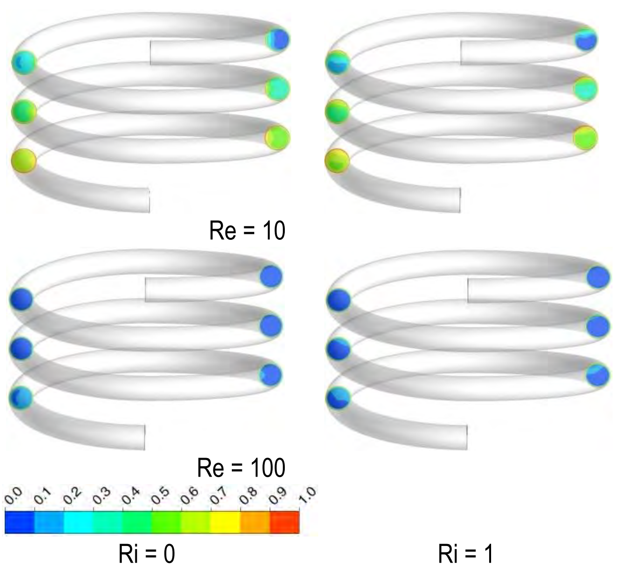


Fig. 7. Isotherms in cross-sections of the angles 90° and –90° in terms of Re and Ri at n = 1 (Re, Reynolds number; Ri, Richardson number)

It can be concluded that the dynamic behaviour of the fluid is affected significantly by the fluid velocity and the intensity of thermal buoyancy.

5.2. Effect of power law index

Through the law of Ostwald, the index n determines the rheological nature of the fluid; therefore, if the value is <1, this means that the fluid is of the shear-thinning type, i.e. a higher fluid velocity leads to a decrease in the viscosity of the fluid. Whereas, if the value of n = 1, the fluid is of the Newtonian type, meaning that the viscosity does not change with the velocity. Whereas, for values of n >1, the viscosity increases its value with velocity and here it is called shear-thickeningfluid. Some examples: blood is a shear-thinning fluid of n is <1; water is a Newtonian fluid of n = 1; chocolate is a shear-thickening fluid of n >1.

Fig.8 shows the distribution of the dimensionless velocity of the flow in the middle of the section of 90° angle in terms of the index n and Ri for Re = 100 and B/D = 0.1. It is noted from Fig. 4 that near the bottom, where the shear stress is very important, the higher the value of the index n, the lower is the fluid velocity, and this is clearly due to the rheological nature of the fluid. On the other hand, it is noted that the thermal effect actually can change the trajectory of the fluid flow.

These curves comprise only a sample, showing the presence of an effect of index n and the number Ri on the behaviour and velocity of fluid flow inside the tube. But this effect changes from one point to another inside the studied tube.

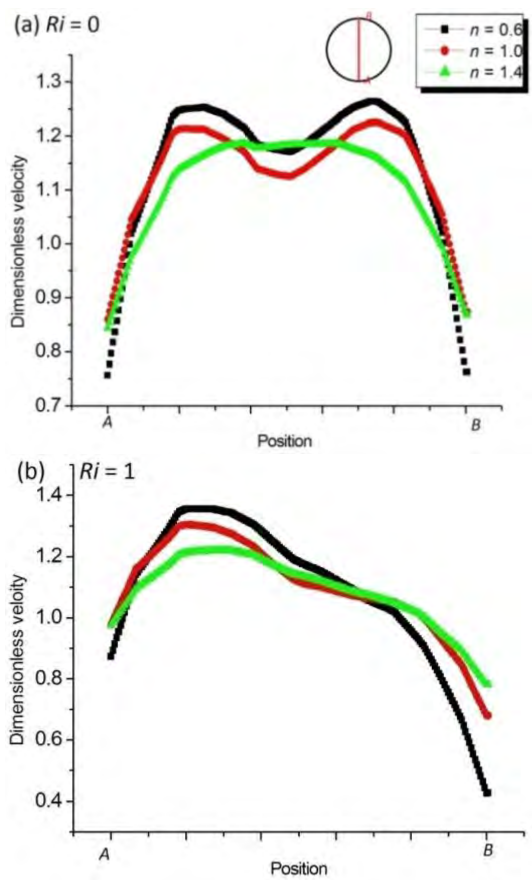


Fig. 8. Distribution of dimensionless velocity along the middle of the cross-section of 90° in terms of n and Ri at Re = 100 Ri, Richardson number.

DOI 10.2478/ama-2022-0014

5.3. Effect of distance between rings

Fig. 9 shows the behaviour of the fluid at the cross-section of angles 180° and –180°for different values of the ratio *B*/*D* and Re at Ri = 1 and *n* = 1. Fig.9 shows the appearance of steady vortices in this section. The shape and size of the vortices are almost not affected by the change of the *B*/*D* ratio, while there is a noticeable development of these vortices in terms of Re in both cases of the *B*/*D* ratio.

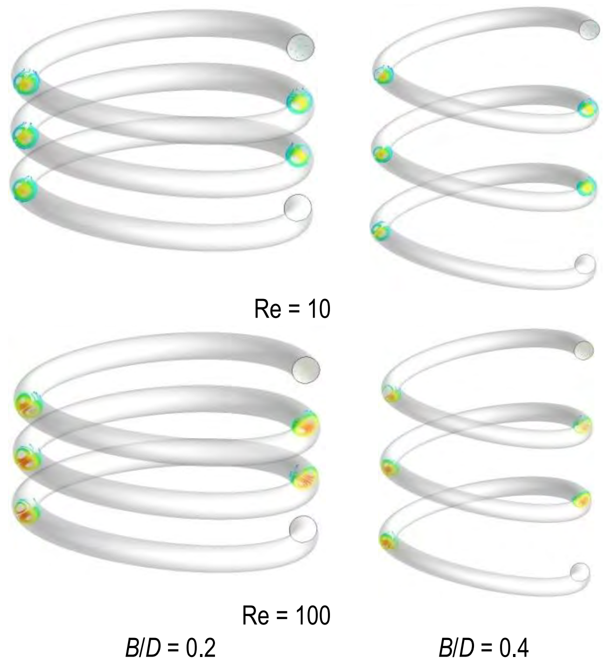


Fig. 9. Streamlines in cross-sections of the angles 180° and –180° in terms of *B*/*D* and Re at *n* = 1 and Ri =1 (B, the pitch between the rings; D, diameter of the tube ring. Re, Reynolds number)

5.4. The average Nusselt number

The main objective of all these studies is to derive the mean values of the Nusselt number in terms of all studied parameters, and accordingly, Fig. 10 explains the changes in the Nu values with *n*, Ri, *B*/*D* and Re. Through Fig. 10, it is shown that the Reynolds number positively affects the values of Nu for all values of *n*, Ri and *B*/*D*. While the changes in the remaining pertinentparameters slightly affect the value of Nu, these changes can be neglected. Finally, it can be concluded that the values of the index n and the number Ri affect the positional distribution of the Nu number, but the average remains approximately the same. In addition to this, the obtained results can be exploited to design helical tube heat exchangers.

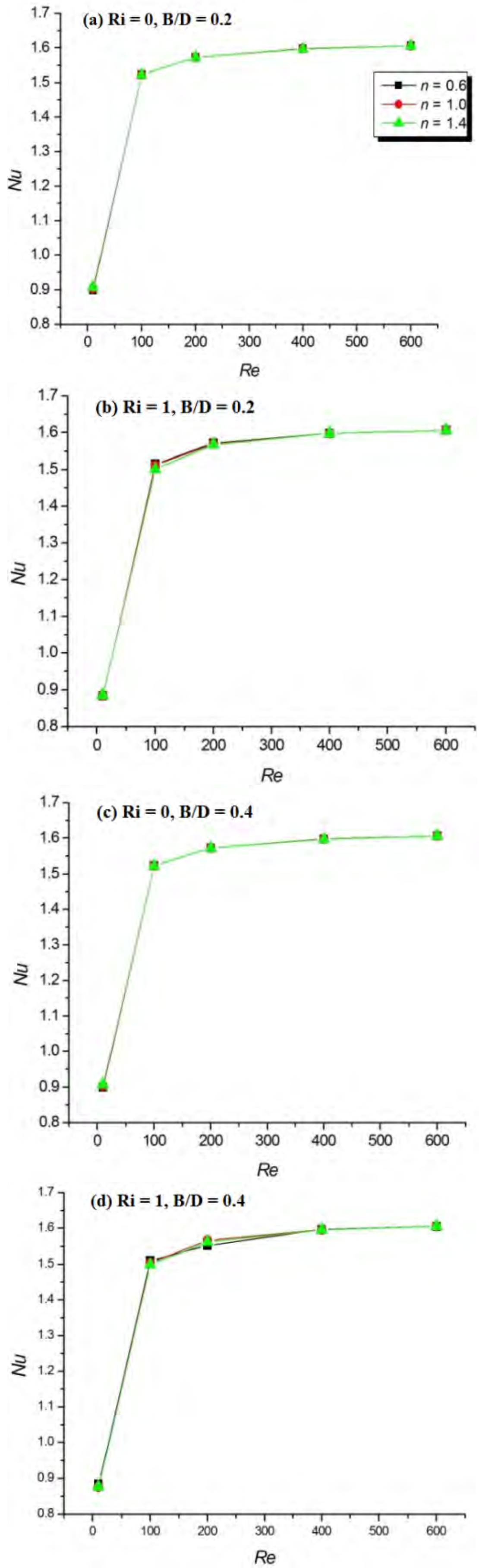


Fig.10. Average Nusselt number (Nu) versus Re, *n*, Ri and *B*/*D* for Pr = 50

DOI 10.2478/ama-2022-0014 acta mechanica et automatica, vol.16 no.2 (2022)

6. CONCLUSION

In this work, we studied the dynamic behaviour of a fluid inside a spiral tube and its effect on heat transfer. The fluid used in this research is defined by the Ostwald model. The work was performed numerically in a steady regime. Moreover, the relevant parameters considered are as follows: the Re number (=10–600), the power law index (=0.6, 1 and 1.4), Ri number (=0 and 1) and the pitch ratio (= 0.2 and 0.5). The most important results obtained are summarised as follows:

- The appearance of constant vortices in the cross-section of angles 180° and –180° for all rings and for Ri = 0; the size of the vortices is mainly related to the values of Re.
- For Ri = 1, early vortices appear in the sections for angles 90° and –90°.
- The values of the power law index and Ri affect only the local values of Nu; the mean rate remains approximately the same.
- Raising the values of the flow velocity increases the heat transfer.
- There is no effect of the *B*/*D* ratio on the behaviour of the flow and on the heat transfer for the considered values (0.2 and 0.4).

REFERENCES

1. Ibrahim M, Algehyne EA, Saeed T, Berrouk A S, Chu Y M, Cheraghian G. Assessment of economic, thermal and hydraulic performances a corrugated helical heat exchanger filled with non-Newtonian nanofluid. Scientific Reports. 2021; 11: 11568.
2. Naphon P, Wiriyasart S, Prurapark R, Srichat A. Numerical study on the nanofluid flows and temperature behaviors in the spirally coiled tubes with helical ribs. Case Studies in Thermal Engineering. 2021; 27: 101204.
3. Abu-Hamdeh NH, Alsulami RA, Rawa MJH, Aljinaidi AA, Alazwari M A, Eltaher MA, Almitani KA, Alnefaie KH, Abusorrah AM, Sindi HF, Goodarzi M, Safaei MR. A detailed hydrothermal investigation of a helical micro double-tube heat exchanger for a wide range of helix pitch length. Case Studies in Thermal Engineering. 2021; 28:101413.
4. Abdzadeh B, Hosainpour A, Jafarmadar S, Sharifian F. Thermo-entropic evaluation of the effect of air injection into horizontal helical tube. Journal of Energy Storage. 2021; 38:102542.
5. Zhou C, Yao Y, Ni L. Development of heat transfer correlations for multi-row helically coile d tub e heat exchangers use d in surface water heat pump systems. International Journal of Heat and Mass Transfer. 2020; 163: 120491.
6. Jha VK, Bhaumik SK. Enhanced cooling in compact helical tube cross-flow heat exchanger through higher area density and flow tortuosity. International Journal of Heat and Mass Transfer. 2020; 150: 119270.
7. Zhao H, Li X, Wu Y, Wu X. Friction factor and Nusselt number correlations for forced convection in helical tubes. International Journal of Heat and Mass Transfer. 2020; 155: 119759.
8. Zhou C, Zarrella A, Yao Y, Ni L. Analysis of the effect of icing on the thermal behavior of helical coil heat exchangers in surface water heat pump applications. International Journal of Heat and Mass Transfer. 2022; 183: 122074.
9. Cao Y, Ayed H, Anqi A E, Tutunchian O, Dizaji H S, Pourhedayat S. Helical tube-in-tube heat exchanger with corrugated inner tube and corrugated outer tube: experimental and numerical study. International Journal of Thermal Sciences. 2021; 170: 107139.
10. Ahn K, Lee KH, Lee JS, Won C, Yoon J. Analytic spring back prediction in cylindrical tube bending for helical tube steam generator. Nuclear Engineering and Technology. 2020; 52: 2100-2106.
11. Farnam M, Khoshvaght-Aliabadi M, Asadollahzadeh MJ. Intensified single-phase forced convective heat transfer with helical-twisted tube in coil heat exchangers. Annals of Nuclear Energy. 2021; 154: 108108.
12. Eisapour A H, Naghizadeh A, Eisapour M, Talebizadehsardari P. Optimal design of a metal hydride hydrogen storage bed using a helical coil heat exchanger along with a central return tube during the absorption process. International journal of hydrogen energy. 2021; 46: 14478-14493.
13. Gul S, Erge O, Oort E V. Frictional pressure losses of Non-Newtonian fluids in helical pipes: Applications for automated rheology measurements. Journal of Natural Gas Science and Engineering. 2020; 73: 103042.
14. Gul S, Erge O, Oort E V. Helical Pipe Viscometer System for Automated Mud Rheology Measurements. IADC/SPE International Drilling Conference and Exhibition. 2020; IADC/SPE-199572-MS.
15. Mokeddem M, Laidoudi H, Bouzit M. 3D Simulation of Dean vortices at 30 position of 180 curved duct of square cross-section under opposing buoyancy. Defect and Diffusion Forum. 2018; 389: 153-163.
16. Mokeddem M, Laidoudi H, Makinde OD, Bouzit M. 3D Simulation of incompressible poiseuille flow through 180° curved duct of square cross-section under effect of thermal buoyancy. Periodica Polytechnica Mechanical Engineering. 2019; 63: 257-269.
17. Mokeddem M, Laidoudi H, Bouzit M. Computational Analyses of Flow and Heat Transfer at 60° Position of 180° Curved Duct of Square Cross-Section. Diffusion Foundations. 2020; 26: 53-62.
18. Cao X, Du T, Liu Z, Zhai H. Experimental and numerical investigation on heat transfer and fluid flow performance of sextant helical baffle heat exchangers. International Journal of Heat and Mass Transfer. 2019; 142: 118437.
19. Cao X, Chen D, Du T, Liu Z, Ji S. Numerical investigation and experimental validation of thermo-hydraulic and thermodynamic performances of helical baffle heat exchangers with different baffle configurations. International Journal of Heat and Mass Transfer. 2020; 160: 120181.
20. Chen Y, Tang H, Wu J, Gu H, Yang S. Performance comparison of heat exchangers using sextant/trisection helical baffles and segmental ones. Chinese Journal of Chemical Engineering. 2019; 27: 2892–2899.
21. Chen D, Zhang R, Cao X, Chen L, Fan X. Numerical investigation on performance improvement of latent heat exchanger with sextant helical baffles. International Journal of Heat and Mass Transfer. 2021; 178: 121606.
22. Jamshidi N, Mosaffa A. Investigating the effects of geometric parameters on finned conical helical geothermal heat exchanger and its energy extraction capability. Geothermics. 2018; 76: 177–189.
23. Abu-Hamdeh NH, Almitani KH, Alimoradi A. Exergetic performance of the helically coiled tubeheat exchangers: Comparison the sector-by-sector with tube in tube types. Alexandria Engineering Journal. 2021; 60: 979–993.
24. Abu-Hamdeh NH, Bantan RAR, Tlili I. Analysis of the thermal and hydraulic performance of the sector-by-sector helically coiled tube heat exchangers as a new type of heat exchangers. International Journal of Thermal Sciences. 2020; 150: 106229.
25. Liu S, Huang W, Bao Z, Zeng T, Qiao M, Meng J. Analysis, prediction and multi-objective optimization of helically coiled tube-in-tube heat exchanger with double cooling source using RSM. International Journal of Thermal Sciences. 2021; 159: 106568.
26. Mirgolbabaei H. Numerical investigation of vertical helically coiled tube heat exchangers thermal performance. Applied Thermal Engineering. 2018; 136: 252–259.
27. Javadi H, Ajarostaghi S S M, Pourfallah M, Zaboli M. Performance analysis of helical ground heat exchangers with different configurations. Applied Thermal Engineering. 2019; 154: 24–36.
28. Maghrabie H M, Attalla M, Mohsen AAA. Performance assessment of a shell and helically coiled tube heat exchanger with variable orientations utilizing different nanofluids. Applied Thermal Engineering. 2021; 182: 116013.

DOI 10.2478/ama-2022-0014

29. Vivekanandan M, Venkatesh R, Periyasamy R, Mohankumar S, Devakumar L. Experimental and CFD investigation of helical coil heat exchanger with flower baffle. Materials Today: Proceedings. 2021; 37: 2174–2182.
30. Sadhasivam C, Murugan S, Manikandaprabu P, Priyadharshini SM, Vairamuthu J. Computational investigations on helical heat flow exchanger in automotive radiator tubes with computational fluid dynamics, Materials Today : Proceedings. 2021; 37: 2352–2355.
31. Gokulnathan E, Pradeep S, Jayan N, Bhatlu M L D, Karthikeyan S. Review of heat transfer enhancement on helical coil heat exchanger by additive passive method, Materials Today. Proceedings. 2021; 37: 3024–3027.
32. Padmanabhan S, Reddy OY, Yadav KVAK, Raja VKB, Palanikumar K. Heat transfer analysis of double tube heat exchanger with helical inserts, Materials Today: Proceedings. 2021; 46: 3588–3595.
33. Naik B, Hosmani A K, Kerur S M, Jadhav CC, Benni S, Annigeri S, Javali T, Aralikatti P. Numerical analysis of two tube helical heat exchanger using various nano-fluids, Materials Today: Proceedings. 2021; 47: 3137–3143.
34. Kumar PCM, Chandrasekar M. CFD analysis on heat and flow characteristics of double helically coiled tube heat exchanger handling MWCNT/water nanofluids. Heliyon. 2019; 5: e02030.
35. Dhumal G S, Havaldar SN, Numerical investigation of heat exchanger with inserted twisted tape inside and helical fins on outside pipe surface. Materials Today. Proceedings. 2021; 46: 2557–2563.
36. Kareem R. Optimisation of Double Pipe Helical Tube Heat Exchanger and its Comparison with Straight Double Tube Heat Exchanger. J. Inst. Eng. India Ser. C. 2017; 98, 587–593.
37. Zainith P, Mishra N K. Heat Transfer Enhancement of Al2O3-Based Nanofluid in a Shell and Helical Coil Heat Exchanger.Advances in Applied Mechanical Engineering. Lecture Notes in Mechanical Engineering. 2020. https://doi.org/10.1007/978-981-15-1201-8_18
38. Miansari M., Jafarzadeh A., Arasteh H. and Toghraie D. (2021), Thermal performance of a helical shell and tube heat exchanger without fin, with circular fins, and with V-shaped circular fins applying on the coil, Journal of Thermal Analysis and Calorimetry,143, 4273–4285.
39. Bara B, Nandakumar K, Masliyah J H. An experimental and numerical study of the Dean problem: flow development towards two-dimensional multiple solutions. Journal of Fluid Mechanics. 1992; 244: 339–376.
40. Helin L, Thais L, Mompean G. Numerical simulation of viscoelastic Dean vortices in a curved duct. Journal of Non-Newtonian Fluid Mechanics. 2009; 156: 84–94.

Mohamed Ramla: https://orcid.org/0000-0002-7144-5264

Houssem Laidoudi: https://orcid.org/0000-0001-8700-7077

Mohamed Bouzit: https://orcid.org/0000-0002-1417-7291

PERFORMANCE EVALUATION OF A SINGLE CYLINDER COMPRESSED AIR ENGINE: AN EXPERIMENTAL STUDY

Ahmad HANAN*, Tariq FEROZE*, Awais ARIF, Hasan IFTIKHAR**, Afzaal A. KHAN**, Sarmad JAVAID***

*Department of Civil Engineering, Military College of Engineering, National University of Sciences and Technology, 3X3Q+3Q6, Risalpur, Nowshera, Khyber Pakhtunkhwa, Pakistan
**Department of Mechanical and Aerospace Engineering, Air University, Main Campus PAF Complex E-9 Islamabad, Pakistan

ahmadhanan@live.com, taferoz@mce.nust.edu.pk, engrawaisarif@gmail.com, hasan101@live.com, afzaal.khan@mail.au.edu.pk, sarmadjvd91@gmail.com

received 14 December 2021, revised 3 February 2022, accepted 5 February 2022

Abstract: The quest to reduce dangerous environmental emissions has led to the research and use of alternate and renewable energy sources. One of the major contributors to the dangerous environmental emissions is the automotive industry. The world is, therefore, quickly moving towards hybrid and electric vehicles. An alternate pollution-free automotive engine is a compressed-air engine, which is powered by compressed air and is more efficient than the electric engine since it requires less charging time than a traditional battery-operated engine. Furthermore, the tanks used in compressed-air engines have a longer lifespan in comparison to the batteries used in electric vehicles. However, extensive research is required to make this engine viable for commercial use. The current study is a step forward in this direction and shows the performance analysis of a single-cylinder compressed-air engine, developed from a four-stroke, single-cylinder, 70 cc gasoline engine. The results show that compressed-air engines are economic, environmental friendly and efficient.

Key words: internal combustion engines, compressed air engines, engine analysis, single cylinder engine

1. INTRODUCTION

Engines, of both combustion and non-combustion types, are used to convert energy into useful mechanical motion. Combustion engines convert the energy of the burning fuel, viz. expansion of the high-pressure and -temperature gases produced during the burning of fuel, to mechanical work. The combustion process, besides producing mechanical work, generates and releases environment-polluting emissions, such as unburned hydrocarbons, carbon dioxide, nitrogen oxides, volatile organic compounds, particulate matter and so on [1]. The environmental challenges posed by combustion engines used in passenger cars have forced the automotive industry to look for environmental alternatives, such as compressed-air engines [2]. A compressed-air engine is powered by compressed air stored in a tank. Instead of creating an air–fuel mixture and burning it in the engine to drive pistons with the expansion of hot gases, compressed-air engines use the expansion of compressed air to drive their pistons. Hence, air-compressed engines have the inherent benefits of pollution-free character and economy in terms of fuel and lubrication (since ignition procedure is completely eliminated and a wide range of motor oils can be used for lubrication purposes for a longer period) [3]; moreover, there is no need for a separate air-conditioning system (the exhaust air can be used for air conditioning) [4].

The first compressed-air vehicle was developed for a locomotive by Bompas in England in 1828. There were two storage tanks between the frames, with conventional cylinders and cranks. The first documented compressed-air vehicle in France was built by the Frenchmen Tessie and Andraud of Motay in 1838. A compressed-air-run car was tested on a test track at Chaillot on 9 July 1840 and it worked well. A similar vehicle was made by Barin von Rathlen in 1848 and was driven from Putney to Wandsworth (London) at an average speed of 12 mph. During the 1850s, a French engineer called Julienne drove a self-made compressed-air car at Saint Denis in France, driven by air compressed at 350 psi. He used it in coal mines. In 1874, the famous Simplon tunnel was dug by using compressed-air locomotives and the cold exhaust of the engine was used for ventilation in the tunnel. Louis Mekarski built a standard-gauge self-contained tramcar, which was tested in February 1876 on the Courbevoie-Etoile Line of the Paris Tramways Nord (TN), which impressed the President and Minister of Transport Maurice de McMahon [5,6,7].

Various companies are financing the research, development and deployment of compressed-air locomotives. Overoptimistic reports indicating 'awaiting production' date back to at least 1999. For instance, an air car manufactured by the French company Motor Development International (MDI) made its public debut in South Africa in 2002, and it was predicted that production will start 'within 6 months' in January 2004. But until January 2009, the air car never went into production in South Africa. But the company started its production in France in 2009. And currently, MDI's air car is common on the roads of Paris, Orleans, Marsille and Lyon. Tata Motors, in collaboration with MDI, planned to develop a car, named AIRPod, which would run on compressed air and could travel 140 km. Tata planned to launch this car in 2020, but it has yet not been launched, although the first phase of the project has been completed successfully. The production model of the

10.2478/ama-2022-0015

AIRPod will have a maximum speed of >65 km/h [8,9]. However, more work is still needed in various areas related to compressed-air engines [10].

Compressed-air engines have been well studied, but their potential has not been understood enough to encourage mass production. Furthermore, less work is available on the description of its construction for academic developers and on the conversion of fuel engines to compressed engines. The current study was carried out to convert a four-stroke 70 cc petrol engine, which is cheap and easily available, to a two-stroke compressed-air engine. The redesigning and performance analysis has been presented here with the aim of understanding the potential and capabilities of a small compressed-air engine. The study will give confidence to researchers to carry out more research on compressed-air engines in collaboration with the automobile industry to pave the way for the development of vehicles run on such engines and reduce environmental challenges [11]. Furthermore, the novel idea may also encourage more research to study the prospects of converting existing fuel engines to compressed-air engines.

2. MATERIALS AND METHODS

For the conversion of a four-stroke engine to a two-stroke compressed-air engine, a single-cylinder four-stroke gasoline-operated 72 cc engine was used; its other important details are provided in Tab. 1 [12]. This particular engine was selected as it is cheap and easily available and has better relative performance parameters. A new operating cycle was devised, including a series of mechanical modifications and experimental verification, as explained in the ensuing paragraphs, in order to successfully achieve the aims and objectives [13].

Tab. 1. Important details of the 4-stroke engine

Property	Value
Bore length, cm	4.7
Stroke length, cm	4.12
Cylinders, n	1
Compression ratio	8.8:1
Displaced volume, cm^3	71.82
Clearance volume, cm^3	9.23
Total volume of cylinder, cm^3	81.05
Surface area of piston, cm^2	17.34

2.1. Cycle/mechanical/electrical modifications

A typical single-cylinder four-stroke gasoline engine operates on a standard Otto cycle, in which air at atmospheric pressure is introduced into the cylinder during the suction stroke, which is then compressed in the compression stroke with the sole purpose of raising the pressure of the introduced air. However, in a compressed-air engine, since the injected air is already at a high pressure, the compression stroke is no more needed. Furthermore, since no ignition is required inside the engine, the carburettor and spark plugs were removed from the engine [14]. The compressed-air cycle engine thus consists of only two strokes, namely intake/expansion and exhaust, to utilise the energy of the compressed air to the maximum. During the intake/expansion stroke, compressed air is injected into the cylinder and allowed to expand to move the piston down to the bottom dead centre (BDC) from the top dead centre (TDC). In the exhaust stroke, the air is expelled out of the cylinder.

2.1.1. Modification in Air Intake Circuit

Air must be injected into a compressed-air engine through the inlet that does not leak and that tolerates high pressure. To achieve this in the modified engine, compressed air is injected into the cylinder through an unorthodox opening (spark plug port) [15] during the intake/expansion stroke, while both inlet and outlet valves are closed; the air expands in the cylinder to move the piston down to the BDC. During the exhaust stroke, when the piston approaches the TDC, both inlet and outlet valves open one after another to let the expanded air out of the cylinder, as shown in Fig. 1. Therefore, the spark plug was replaced by a specially designed nozzle that is fixed in the threaded port of the spark plug. The fabricated nozzle and its port are depicted in Fig. 2.

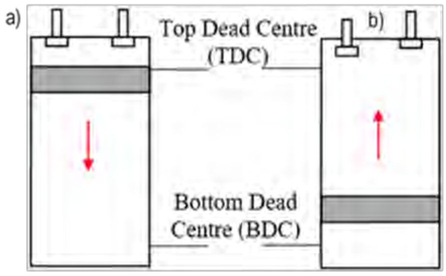


Fig. 1. Strokes in a compressed air cycle engine: a) Intake/expansion stroke; b) Exhaust Stroke

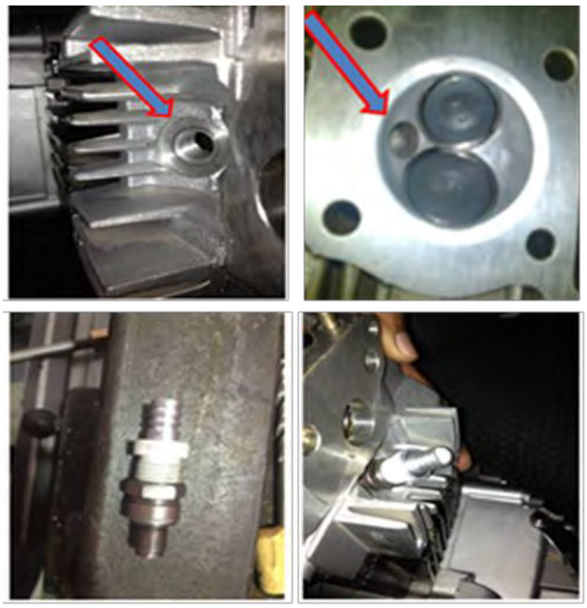


Fig. 2. Spark plug port and fabricated nozzle

2.1.2. Cam Crank Ratio

The sprockets at the end of the cam shaft and crank shaft are connected by a timing chain in the original engine. For the two-stroke operation of the compressed-air engine, the diameter and number of teeth on the cam sprocket must be equal to those of the crank sprocket to allow one revolution of the cam shaft during one complete revolution of the crankshaft. To achieve a 1:1 ratio, the crank sprocket was removed and the new sprocket was fabricated. Similarly, a new chain was designed using CAD software, which was fabricated and adjusted accordingly to allow for the change in length of the chain due to the change in the crankshaft sprocket [16]. The camshaft and crankshaft assembly and the modified crankshaft sprockets are shown in Figs. 3 and 4, respectively. The new dimensions of the chain and sprockets are given in Tab. 2.

Fig. 3. Crankshaft and camshaft assembly

Fig.4. Modified crankshaft sprocket

Tab. 2. Modified cam-crank dimensions

Property	Value
Distance – sprocket centres, cm	18
Teeth on camshaft sprocket, n	14
Teeth on crankshaft sprocket, n	14
Radius of camshaft sprocket, cm	4
Radius of crankshaft sprocket, cm	4
Length of chain, cm	61.5
Pitch, mm	17.8

2.1.3. Cam Lobe Positioning

The chain was wrapped in such a way to achieve the desired cycle, the piston must be at the TDC at the instant when both valves just close. However, in a four-stroke petrol engine, the piston is at the TDC at the beginning of the intake stroke and expansion strokes, and the cam position also varies accordingly. A graphical description of the cam positioning is shown in Fig. 5. The cam parameters and cam displacement are given in Tab. 3 and Fig. 6, respectively [17].

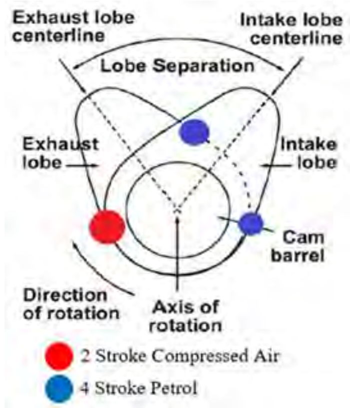


Fig. 5. Position of top dead centre (TDC)

Tab. 3. Cam parameters

Property	Value (mm)
Base diameter	21.1
Lobe lift	4.9
Lobe height	26

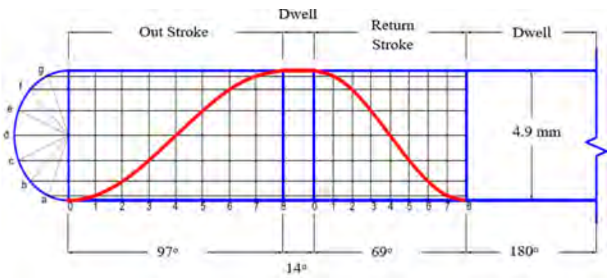


Fig. 6. Adjusted cam lobe displacement

2.1.4. Compressed Air Injection Control

The capacitor discharge ignition (CDI) box of the engine, along with other major electrical components – as shown in Fig. 7, was used to control the timing of injection of compressed air into the cylinder at the correct time (piston at the TDC) [18]. The inbuilt timing of the CDI was not altered; rather, mechanical modifications were made in a way that satisfies our operating cycle, described in Section 2.1. A solenoid valve was connected with the CDI through an electric circuit (Fig. 7), which controls the opening and closing of the valve with a pulse. This was done to stop the continuous supply of compressed air. The piston regains its position due to the crank mechanism operating within. Thus, the air injection angle was synchronised with the operating cycle; the cam profile is shown in Fig. 6 for reference.

- Solenoid valve: To synchronise the power stroke of compressed air with the piston's position.
- Relay switch: To control the timings of the solenoid valve.
- Switching transistor: To amplify the current from the CDI,

10.2478/ama-2022-0015

which is not enough to operate the switching relay.
– Resistor: To prevent excess flow of current to the transistor.
– Catch diode: To prevent damage to the switching transistor by capturing the reverse voltage from the solenoid and routing it to the battery.

Fig. 7. Electrical circuit for compressed air injection timings

2.2. Thermodynamics Performance Analysis

A detailed theoretical analysis of the thermodynamic performance was performed, based on the design parameters, to obtain the expected efficiency of the compressed-air engine. The important design parameters and the calculated performance parameters of the thermodynamic analysis are given in Tab. 4 [19].

Tab. 4. Calculated thermodynamic performance parameters/ design parameters

Parameter	Value
Pressure at the beginning of compression, bar	1.01352
Pressure at the end of compression, bar	6.2
Pressure at the end of expansion, bar	3.67
Volume of free air, m^3	0.37
Volume of air at the end of compression, m^3	0.103
Volume of air at the end of expansion, m^3	0.255
Temperature before compression, $^{\circ}$C	16
Temperature after compression, $^{\circ}$C	216
Temperature after expansion, $^{\circ}$C	16
Specific heat, kJ	63.62
Polytrophic constant	1.406
Length of connecting rod, mm	72.45
Area of piston, m^2	0.001734
Work done during compression, kJ	63.723
Work done during expansion, kJ	28.88
Efficiency	45%
Force available on piston, N	1,075
Torque, Nm	77.88
Revolutions per minute, rev/min	719.3
Angular velocity, rad/s	75

3. RESULTS AND DISCUSSION

3.1. Experimental Verification

A cylinder with a capacity of 50 L and a maximum pressure of 8 bar was used, with the engine installed on a specially designed test bench as shown in Fig. 8. The filling time of the cylinder using a compressor is 2.7 s. The compressed-air engine was operated at a pressure of 5.44 bar (80 psi) successfully at an average speed of 700 rev/min (recorded using a tachometer). The engine stopped running when the pressure in the tank reduced to 2.44 bar (30 psi).

The potential of the compressed-air engine was analysed, based on the values of the parameters applied and calculated during the experimental analysis. The average pressure used for the calculation was 5.9 bar, for which the results are given in Tab. 5.

A tachometer was used to measure the speed of the engine; the angular velocity was derived from it. A pressure gauge was attached at the outlet of the solenoid valve to monitor the inlet pressure, as shown in Figs. 7 and 8. Available force and torque were calculated with the help of gauge readings, and standard formulas of automotive engineering were used [20] for the calculations.

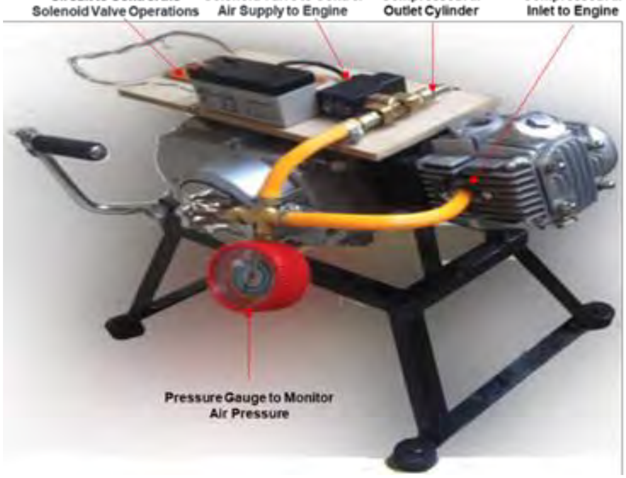


Fig. 8. Test bench used to test the compressed air engine

Tab. 5. Results of experimental analysis

Parameter	Value
Force, N	1,016
Torque, Nm	73.69
Angular velocity, rad/s	73.3

3.2. Economic Analysis

Despite numerous factors affecting the economy of a fuel, a preliminary comparative analysis was conducted on a 70 cc internal combustion (IC) engine idling at 700 rev/min. The study shows that, for a given mileage, using compressed air reduces the cost by nearly three times as compared to petrol [21]. The salient features of the study are given below [22]:

- The value of fuel consumption for a 70 cc IC engine is ~0.4 L/h under idle operating conditions.
- During our experiment, compressed air at 8 bar was used in a 50 L tank. The engine was kept running for around 2 min.
- With the help of thermodynamic analysis, it can be estimated that, to keep the engine running for 1 h, 1 m3 of compressed air at 30 bar is required.
- Cost of production of 1 m3 of compressed air at 30 bar is 0.2 USD.
- Currently, 1 L petrol costs nearly 0.9 USD in Pakistan.

4. CONCLUSIONS AND RECOMMENDATIONS

The compressed-air engine is economic and pollution free. By redesigning a small single-stroke petrol engine to a compressed-air engine, the current study has demonstrated the simplicity of the design of a compressed-air engine, its cost-effectiveness and potential. The study is expected to give confidence to researchers working on compressed-air engines and to pave the way for the development of vehicles run on such engines to reduce the fuel crises and environmental challenges; however, more research is needed in this field to utilise the technology with commercial viability [10]. Some of the areas recommended for further research, specifically in relation to the current study, are highlighted below:

- A more professional approach to the designing of such engines.
- Testing on a multiple-cylinder engine.
- Studying lightweight reinforced air storage tanks to render it possible for the engine to be used in the automotive industry.

REFERENCES

1. Zefaan H. Combustion chamber geometry effects in spark ignition engine exhaust emissions. Australian Journal of Mechanical Engineering. 2012;10(1):29-39. http://dx.doi.org/10.7158/M11-799.2012.10.1
2. Aravindhan N, Vasanth KM, Kumar RV, Jayasurya M, Prakash SS, Sabareeshwaran V. A novel approach for improving the performance of air engine to achieve zero-emission for a pollution-free environment. Materials Today: Proceedings. 2020;33(1):39-43. https://doi.org/10.1016/j.matpr.2020.02.930
3. Bossel U. Thermodynamic analysis of compressed air vehicle propulsion. Journal of KONES Internal Combustion Engines. 2005; 12(3):51-62.
4. Holovach I, Kasha L, Hudzii I. Individual drive of internal combustion engine lubrication system based on switched reluctance motor, Energy Engineering and Control Systems. 2020; 6(3):146-151. https://doi.org/10.23939/jeecs2020.02.146
5. Rząsa M, Łukasiewicz E, Wójtowicz D. Test of a new low-speed compressed air engine for energy recovery. Energies. 2021;14(4): 1-15. https://doi.org/10.3390/en14041179
6. Surwase AA, Date D, Patel A. Design of Pneumatic Powered Bicycle. International Journal of Recent Advances in Multidisciplinary Topics [Internet]. 2021 Aug [cited 2021 Dec. 15];2(8):58-60. Available from: https://www.journals.resaim.com/ijramt/article/view/1242
7. Wiley WK. Appliances for the use of compressed air [dissertation on the Internet]. Illinois: College of Engineering, University of Illinois; 1904. Available from: https://www.ideals.illinois.edu/bitstream/handle/2142/92605/5963444.pdf?sequence=1
8. Robertson S. Air car basics. Pneumatic Options Research Library; 1981. [cited 2021 Dec 15]. Available from: archive.org/details/aircarbasics
9. Thipse SS. Compressed air car. Tech Monitor. 2008; 1(2):33-37.
10. Fang Y, Lu Y, Roskilly AP, Yu X. A review of compressed air energy systems in vehicle transport. Energy Strategy Reviews. 2021;33: 1-13. https://doi.org/10.1016/j.esr.2020.100583
11. Korbut M, Szpica D. A Review of Compressed Air Engine in The Vehicle Propulsion System. Acta Mechanica et Automatica. 2021; 15(4): 215-226. https://doi.org/10.2478/ama-2021-0028
12. Szpica D, Korbut M. Modelling methodology of piston pneumatic air engine operation. Acta Mechanica et Automatica. 2019;13(4): 271-278. https://doi.org/10.2478/ama-2019-0037
13. Gajendra Babu MK, Murthy BS. Simulation and evaluation of a 4-stroke single-cylinder spark ignition engine. SAE Transactions. 1975;84(2):1631-1659
14. Seela CR, Raoa DV, Raoa MV. Performance Analysis of an Air Driven Engine Modified from SI Engine. Res. Artic. Int. J. Curr. Eng. Technol 2013; 3(4):1440-1446.
15. Szoka W, Szpica D. Adaptation of classic combustion engines to compressed air supply. Acta Mechanica et Automatica. 2012; 6(1):68-73.
16. Kumar A, Kumar N, Gupta D, Kumar V. Optimization Analysis of Injection Angle and Injector Nozzle of an Advanced Compressed Air Engine Kit. SAE Technical Paper; 2015 Apr. https://doi.org/10.4271/2015-01-1678
17. Kakaee AH, Zareei JA. Influence of varying timing angle on performance of an SI engine: An experimental and numerical study. Journal of Computational & Applied Research in Mechanical Engineering (JCARME). 2013;2(2):33-43. http://dx.doi.org/10.22061/jcarme.2013.51
18. Liu MY, Sun JJ, Zhao W, Zhang TT, Liu ZW. Research on a Motion Law Surveying Aparatus of a Cam Follower. Key Engineering Materials 2016;693:1758-1764. https://doi.org/10.4028/www.scientific.net/KEM.693.1758
19. Szpica D, Kusznier M. Modelling of the low-pressure gas injector operation. Acta Mechanica et Automatica. 2020;14(1):29-35. https://doi.org/10.2478/ama-2020-0005
20. Haywood J B. Internal Combustion Engine Fundamentals. New York: McGraw-Hill;1988.
21. Yu Q, Cai M, Shi Y, Xu Q. Optimization study on a single-cylinder compressed air engine. Chinese Journal of Mechanical Engineering. 2015;28(6):1285-1292. https://doi.org/10.3901/CJME.2015.0520.072
22. Nguyen YL, Le AT, Duc KN, Duy VN, Nguyen CD. A study on emission and fuel consumption of motorcycles in idle mode and the impacts on air quality in Hanoi, Vietnam. International Journal of Urban Sciences. 2021;25(4):522-541. https://doi.org/10.1080/12265934.2020.1871059

Ahmed Hanan: https://orcid.org/0000-0001-5165-431X

Tariq Feroze: https://orcid.org/0000-0003-0308-6922

Awais Arif: https://orcid.org/0000-0003-3054-5573

Hasan Iftikhar: https://orcid.org/0000-0002-4258-2892

Afzaal Ahmed Khan: https://orcid.org/0000-0003-1377-8754

Sarmad Javaid: https://orcid.o rg/0000-0001-7624-5651

DOI 10.2478/ama-2022-0016

LINEAR MOTION ERROR EVALUATION OF OPEN-LOOP CNC MILLING USING A LASER INTERFEROMETER

Agustinus WINARNO*, Benidiktus T. PRAYOGA*, Ignatius A. HENDARYANTO*

*Department of Mechanical Engineering, Vocational School, Universitas Gadjah Mada,
Jl. Yacaranda Sekip Unit IV, Yogyakarta 55281, Indonesia

beni@ugm.ac.id,winarno_agustinus@ugm.ac.id, aris.hendaryanto@ugm.ac.id

received 2 October 2021, revised 6 February 2022, accepted 10 February 2022

Abstract: The usage of computerised numerical control (CNC) machines requires accuracy verification to ensure the high accuracy of the processed products. This paper introduces an accuracy verification method of an open-loop CNC milling machine using a fringe counting of He–Ne laser interferometry to evaluate the best possible accuracy and functionality. The linear motion accuracy of open-loop CNC milling was evaluated based on the number of pulses from the controller against the actual displacement measured by the He–Ne fringe-counting method. Interval distances between two pulses are also precisely measured using the He–Ne interferometry. The linear motion error and controller error can be simultaneously evaluated in sub-micro accuracy. The linear positioning error due to the micro-stepping driver accuracy of the mini-CNC milling machine was measured with the expanded uncertainty of measurement and was estimated at 240 nm. The experimental results show that linear motion error of the open-loop CNC milling can reach up to 50 μm for 200 mm translation length.

Key words: accuracy, micro-stepping driver, CNC, milling, machine, interferometry, open loop

1. INTRODUCTION

Computerised numerical control (CNC) has been widely used in the manufacturing industry given its advantages of precision, accuracy, cutting productivity, and complexity of work that can be handled. CNC has a crucial role in the manufacturing industries nowadays [1-3]. Since the operation and maintenance of a CNC machine requires a highly skilled operator, numerous universities worldwide have developed several computer numerical machine courses to fulfil the demand for highly skilled human resources in the metal processing industry. Moreover, the computer numerical control course can also be used to introduce a control system and manufacturing technology teaching aid for elementary and junior high school students [4-6].

For low-cost digital manufacturing and educational purposes, a small open-loop controlled CNC milling machine, with relatively simpler instrumentation and cheaper than the close-loop types, can be a suitable choice. Since an open loop is not equipped with a feedback system such as a linear encoder, an open-loop control system is considered cheaper than a closed-loop system [4]. However, the customised milling machine has several limitations, such as a lack of components' dimension accuracy, inappropriate components' interaction, low rigidity and control system problems influencing the cutting performance. Korkut and Donertas stated that the machining performance is influenced by several parameters such as the cutting speed, feed rate, depth of cut and design of the machine [7]. Furthermore, Zmarzly reported that the influence of cutting parameters, especially cutting speed, on the surface quality is considered significant [8]. In order to provide the best possible accuracy and functionality, the accuracy of the developed open-loop CNC milling machine must first be verified.

The verification of an open-loop CNC milling machine can be started from geometric accuracy, which is one of the important evaluated parameters and depends on the translation accuracy of the axes and cutting accuracy [9-11]. The geometric accuracy of machine tools can be divided into two groups, single-axis geometric accuracy and geometric accuracy between axes [12]. The single-axis geometry accuracy can be determined by evaluating the single-axis translation errors. Hence, the potential machining error can be reduced, or an error compensation can be performed using the evaluation data.

The international standard of CNC accuracy evaluation is detailed in ISO 230, which consists of an accurate measurement without the stated load [13,14]. ISO 10791 also specifies several families of tests for machining centres with a horizontal spindle, standing alone or integrated into flexible manufacturing systems [15,16]. In general, linear positioning errors of mini-CNC milling machines can be identified by measuring the table displacement using laser interferometry, dial testers, ball-bars and other measuring devices [17,18]. However, a detailed correlation between the motor controller and actual translation still cannot be easily obtained. A method to simultaneously obtain the correlation of some parameters, such as motor, motor driver, controller, software, and translation table, is required to be developed for the small CNC milling machine with an open-loop control system.

Therefore, this paper proposes an experimental method to evaluate the correlation between the displacement and signal (pulse) of a micro-stepping motor driver using a He–Ne fringe-counting method for a small CNC milling machine. The paper first discusses the methodology, including the experimental setup and the test rig. Then, the effectiveness of the method and the accuracy of the calculation are discussed in the results and discussion sections.

DOI 10.2478/ama-2022-0016

2. METHODOLOGY

In this research, the open-loop CNC milling has been built in our laboratory [19]. The translation length of the open-loop CNC milling table is defined by the CNC part program (G-Code) that controls the number of pulses given to the motor controller (Fig. 1). Some error sources that contribute to the total error are shown in Fig. 2. The first error is the error due to the interface problem (error 1). The second error is the error due to the motor controller (error 2). Errors 3 and 4 are transmission errors and motion errors caused by an imperfect guide system, respectively.

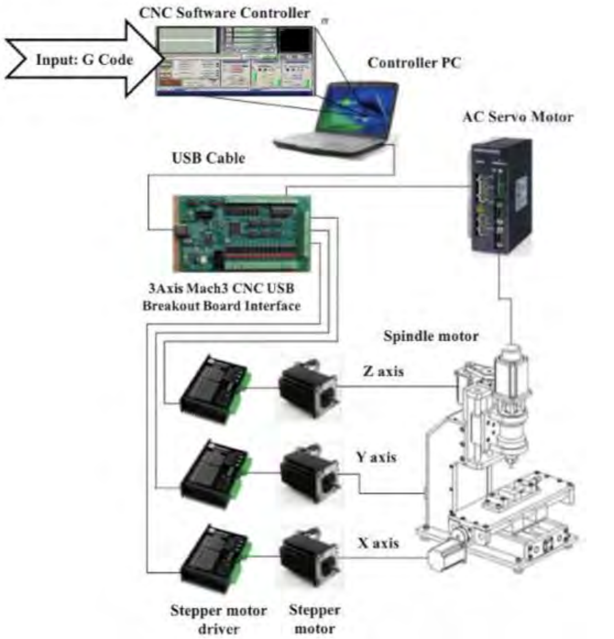


Fig. 1. Schematic diagram of an open-loop mini-CNC milling

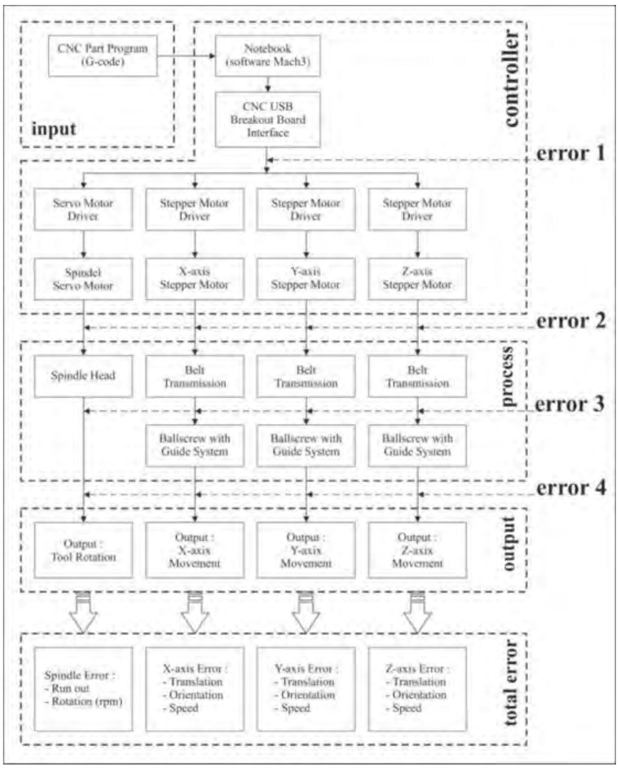


Fig. 2. Map of error sources in the open-loop CNC milling machine

There are six types of motion errors in a single axis milling machine table, grouped as either linear motion errors or rotational motion errors, as shown in Fig. 3. Linear motion errors consist of positioning errors (E_{XX}), vertical straightness errors (E_{ZX}) and horizontal straightness errors (E_{YX}). Rotational motion errors consist of pitch angular errors (E_{BX}), yaw angular errors (E_{CX}) and roll angular errors (E_{AX}). In this research, the method of evaluating the linear positioning error due to the micro-stepping driver accuracy of the mini-CNC milling machine consisted of applying fringe-counting He–Ne laser interferometry to measure the error displacement in E_{XX} direction. The translation of mini-CNC milling machines was compared with the translation measured by He-Ne laser interferometry [20]. Therefore, the correlation between the electric pulse that controls the stepper motor, and the real translation of the CNC table can be evaluated.

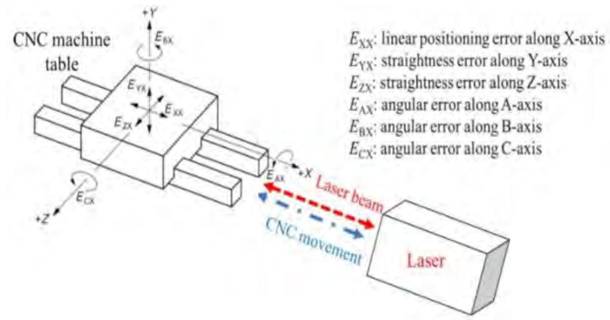


Fig. 3. Method for evaluating the accuracy of the CNC machine using laser interferometry by measuring the positioning error

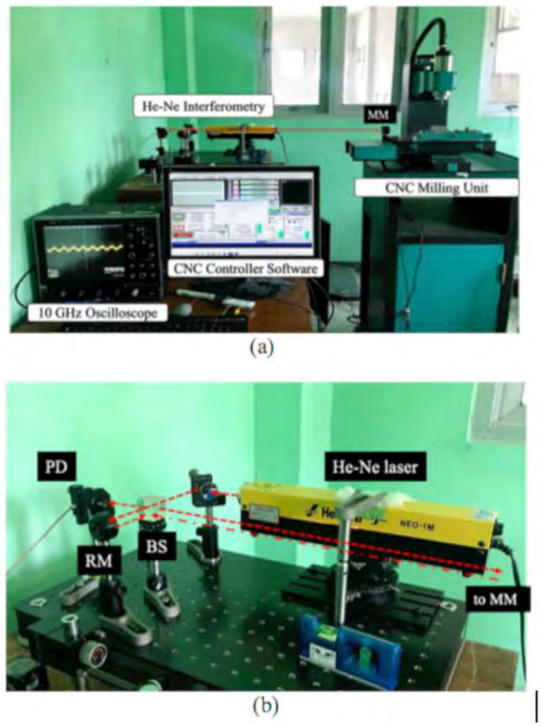


Fig. 4. Setup of evaluation of a mini-CNC milling machine using He–Ne interferometry, consisting of (a) the He–Ne interferometer and (b) retroreflector on the table of the mini-CNC [20]

In this study, the effect of the micro-stepping motor driver on the accuracy of the mini-CNC milling machine was evaluated. The CNC milling machine used in the experiment is a vertical fixed bed type developed in our laboratory (Fig 1). The configuration of

DOI 10.2478/ama-2022-0016

mechanical and electrical parts is shown in the schematic diagram of the open-looped CNC control system. It has a travel distance of 200 mm, 150 mm and 120 mm in X, Y and Z axes, respectively. The experimental setup consists of He–Ne interferometry, an open-loop mini-CNC milling unit and an oscilloscope, as shown in Fig. 4(a).

To minimise the geometrical error, the configuration of beam laser was set carefully, such that the laser beam was in alignment with the path of motion followed by the machine [Figs. 3 and 4(b)]. The open-loop CNC milling machine consists of three motor steppers controlled by micro-stepping drivers (M542, Leadshine Technology Co., Ltd.).

The G-codes are interpreted by Mach3 CNC software to control the stepper motor using an electric pulse generated by a micro-stepping driver. The micro-stepping driver divides one step of a stepper motor (1.8° or 1/200 revolution) into 25 micro-steps. Hence, the total pulses delivered to the motor stepper is 5,000 pulses/revolution. The milling table is driven by a double start ball-screw with 10 mm lead (5 mm effective pitch).
The distance between two pulses (L_P) is calculated as 1 μm by using Eq. (1), as follows:

$$L_p = \frac{P}{i_{ms} \times i_r} \tag{1}$$

where P is 5,000 μm, which is the effective pitch; i_{ms} is 25, which is the number of steps constituting one step of a stepper motor, 25 thus being the micro-step resolution (the number of micro-steps per motor stepper step); and i_r is 200, which is the number of steps per revolution of the motor stepper.

The translation of the CNC table was measured using unstabilised He–Ne fringe-counting interferometry. In this method, the unstabilised He–Ne laser beam from source (NEO-1M, Neoark) with a wavelength of 632.9908 nm incident to the beam splitter BS (BS004, Thorlabs) is divided by two, where one beam is incident to a reference mirror (RM), and another beam is incident to the moving retroreflector (MM) on the table of the mini-CNC milling, as shown in Fig. 4(b). Finally, the reflection beam from MM and RM was detected by a photodetector PD (SM1PD1, Thorlabs), and the fringe pattern was observed by a 10 GHz oscilloscope (Wave Runner 64Xi-A, LeCroy). The fringe-counting method has been widely used to measure the travelled distance [18]. To calculate the displacement length by He–Ne laser using the fringe-counting method, the total translation was calculated by Eq. (2):

$$d = \frac{\lambda \times i}{2n} \tag{2}$$

where d is the translation length in μm, λ is the wavelength (0.6329908 μm) of the unstabilised He–Ne laser in a vacuum [20, 21], i is the number of fringes recorded and n is the refractive index calculated by Ciddor's equation [23]. The measurement using wavelength is influenced by the refractive index, where the refractive index is also influenced by the medium [24-26].

The fractional uncertainty assigned to the vacuum wavelength of the unstabilised He–Ne laser is 1.5 × 10^{-6} (relative standard uncertainty) [21]. In order to simultaneously obtain the interference signal and pulse from the motor driver, the data was automatically recorded 2 s after the table moved to the designed positioned. The interference fringes and pulses were recorded by an oscilloscope with a 210 MS/s sampling rate. The data from the oscilloscope was analysed using an in-house program developed using Python to obtain the distance between pulses. Since the oscilloscope's memory was limited, evaluation of longer translation was also performed using the commercial laser measuring system Renishaw XL80 to obtain the error trend.

3. RESULTS AND DISCUSSION

Data of electric pulses are sorted to obtain each peak, as seen in Fig. 5, where the first pulse is set as a reference pulse. Afterward, relative distances from the reference pulse to the further pulses are calculated by counting the number of He–Ne fringes (integer and fractional parts). The electric pulses and He–Ne interference fringes provide detailed correlation between the micro-stepping driver and actual translation length. From Eqs (1) and (2), we ascertain that the ideal number of interference fringes between two pulses (i) in a vacuum is 3.16, because ideally, d = L_P = 1 μm. However, the number of He–Ne fringes between two pulses fluctuated based on the experimental data.

This fluctuation is considered to be caused primarily by the linear motion error, rather than changes in environmental conditions, since the measurement room was well controlled at 23 ± 0.5 °C and <70% humidity Maintaining humidity and temperature at an optimal level is important because controlled environmental conditions are necessary to rule out fluctuations caused by extrinsic causes [22]. As the second consideration, the fluctuation is caused by the accuracy of the micro-stepping driver in dividing one step of the stepper motor into 25 steps. In order to evaluate the effect of this fluctuation, the total number of pulses (i_P) were counted from the designed start point to the designed endpoint. From Eq. (1), the theoretical translation length in μm (L) is calculated as follows:

$$L = i_p \times L_p \tag{3}$$

From Eqs (2) and (3), the total translation error in m (E) can be calculated as

$$E = L - d = (i_p \times L_p) - \left(\frac{\lambda \times i}{2n}\right) \tag{4}$$

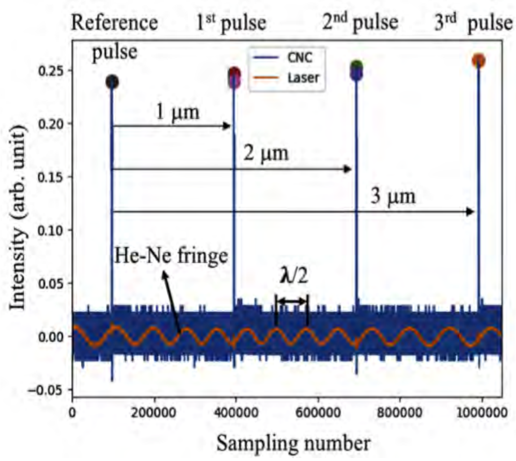


Fig. 5. The calculation result shows the correlation between pulse and distance

The data in Fig. 5 was calculated to obtain the actual distance between two pulses. The distance from the reference pulse to the second pulses (L_{P1}) was calculated as 0.819 μm, which differs by 181 nm from the theoretical value (1 μm). Based on 50 data of the distance measurement between two pulses, the average error from theoretical values was 70 nm, with a standard deviation of 170 nm. From this experiment, the maximum error of the micro-stepping driver can be estimated at about 170 nm. Due to the memory limitation of the oscilloscope in recording the data, only translation lengths up to 0.2 mm were evaluated using the proposed method, as shown in Fig 6. The result shows that the translation error is due to the fact that micro-stepping drivers linearly increase proportionally to the translation length. However, this data falls too short of predicting the error of longer translation. Therefore, the evaluation of longer translation was performed using the commercial laser measuring system Renishaw XL80, where the sampling point was taken every 5 mm. The distance between two pulses cannot be evaluated using Renishaw XL80 due to the limitation of the Renishaw software development kit (SDK) that only allows sampling points up to 10 sampling/s.

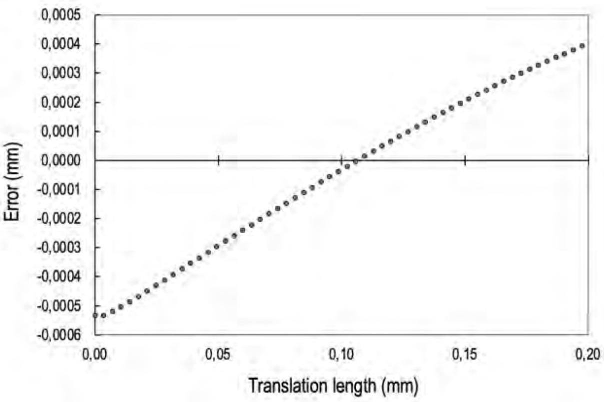


Fig. 6. The translation error of the open-loop CNC milling measured by the He–Ne fringe-counting method

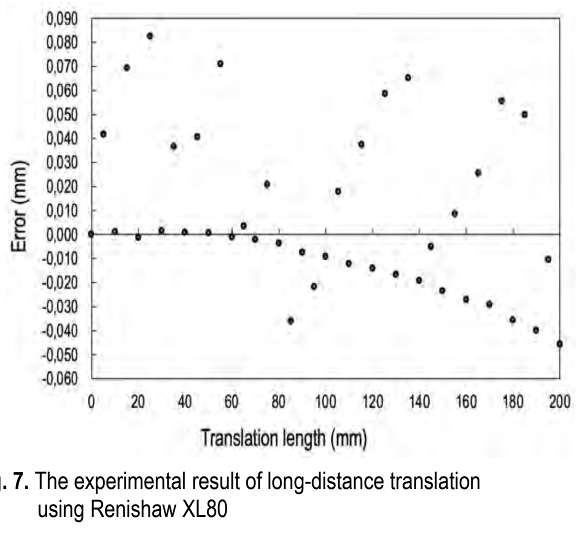


Fig. 7. The experimental result of long-distance translation using Renishaw XL80

The measurement using Renishaw XL80 shows that translation error increases until 30 mm translation length and then decreases to minus direction. The trend line in Fig. 7 shows that the error up to 200 mm translation length is about 50 µm. This error is considered as the error of the carriage table (error 4). From Fig. 7, we observe that there is a random error of about 85 µm, which could be attributed to problems associated with the transmission, mechanical, software or other aspects as causative factors. In the measurement of the table travel (i.e. CNC milling table), the periodic error, such as errors caused by the pitch of the ball-screw and pitch of linear or rotating scales, was wrecked [7]. The effect of motion between elements involved in the construction of machines is one of the most important errors that need to be considered [27]

From the experimental results obtained using the He–Ne fringe-counting method and the commercial length measuring system Renishaw XL80, we derive several observations, as follows: The measurement system should have at least ¼ resolution of the measurand, or its maximum error should not exceed one-quarter of the resolution of the measurand. The He–Ne fringe-counting method achieved a measurement standard deviation of 170 nm. Therefore, it is considered enough to be used in evaluating the micro-stepping driver with a 1 µm resolution. Based on the experimental data, we can consider that the error contribution of the micro-stepping driver to overall error is low. The He–Ne fringe-counting method can only measure short translation, but this method can be applied for longer translation length by adding a data acquisition system to obtain longer data.

The uncertainty budget of measurement up to 0.2 mm using the proposed method is shown in Tab. 1. The standard deviation of 170 nm, obtained from 50 repeated measurements, contributes 24 nm towards standard uncertainty. The uncertainty of distance measurement between two pulses is influenced by the accuracy of the fringe-counting method, which depends on the refractive index of air, wavelength and length.

The geometrical error attributable to imperfect alignment is estimated to contribute 100 nm towards standard uncertainty. Since the measurement room was well controlled at 23 ± 0.5 °C and <70% humidity, it is estimated that temperature fluctuations contribute towards standard uncertainty only by 5 nm. The number of pulses from the micro-stepping motor driver has been estimated to contribute to the standard uncertainty of 50 nm. The combined uncertainty is calculated to be 120 nm. Hence, the expanded uncertainty of the measurement is estimated at about 240 nm with k = 2.

Tab. 1. Uncertainty budget of the fringe-counting method

Uncertainty component	Source	Uncertainty contribution (nm)
u_c (L_d)	Distance between two pulses	
u (λ)	(He–Ne wavelength)	0.3
u (i)	(Number of waves)	32
u (n)	(Refractive index)	6
u (L_p)	Repeatability	24
u (i_p)	Number of pulses	50
u (G)	Geometrical error	100
u (L_t)	Temperature	5
Combined standard uncertainty (u_c)		120
Expanded uncertainty (k = 2)		240

The experiment allowed us to evaluate the linear positioning error attributable to the micro-stepping driver accuracy of the mini-CNC milling machine, and the expanded uncertainty of measurement was estimated at about 240 nm. This research is slightly different from the study of Begović et al [17]. Their research shows the results of linear displacement error measurement and indicates that as a consequence of increasing the lengths from 400 mm to 800 mm, error increases to a maximum of 140 µm

DOI 10.2478/ama-2022-0016

while using commercial CNC. It is considered to reduce the error by re-arranging the interferometer to minimise the dead-path error. The linear translation accuracy of the open loop milling machine has been evaluated using Renishaw XL 80, and the results show that the error up to 200 mm translation length was about 50 µm. From the specification of Renishaw XL, linear measurement accuracy is an assured ±0.5 ppm. For education purposes, this machine is quite promising for use in real applications. Norhadi and Tarng stated that the open-loop system for CNC milling has shown good performance but is limited to use only in non-precise machining [4].

However, further improvement to reduce other errors such as transmission error and motion error caused by imperfect guide systems is highly demanded. The error compensation mechanism maybe good choices to be added. Xu and Dai stated that the error compensation method can be used to improve machining precision [11]. In brief, the accuracy of the micro-stepping driver motion is one of the important factors that affect the linier positioning of the CNC table, and thus warrants further consideration. Further accuracy evaluation such as cutting accuracy, the effect of cutting speed and feed rate and other parameters are considered possible future works.

4. CONCLUSIONS

An open-loop controlled CNC milling machine has been developed for low-cost digital manufacturing and educational purposes. The CNC with an open-loop control system is considered more suitable for educational purposes in terms of a lower price than a closed-loop system. An experimental method was proposed to evaluate the accuracy of an open-loop mini-CNC milling machine using the He–Ne fringe counting. The correlation between the translation error and the accuracy of the motor controller has been investigated. Using fringe counting He–Ne laser, the accuracy of the micro-stepping driver could be precisely evaluated by measuring the distance between two electric pulses. Linear motion error, especially the translation error up to 0.2 mm, was measured with the expanded uncertainty of 240 nm. The longer translation up to 200 mm was measured in order to obtain the overall accuracy of the open-loop CNC milling machine by using the commercial laser measuring system Renishaw XL80. The experiment shows that the linear motion error of the open-loop CNC milling can reach up to 50 µm for 200 mm translation length. From the perspective of educational requirements, the model of machine discussed in the present research offers promising results in terms of good machine design. Simulation to find the correlation between the displacement value for pulses of micro-stepping driver and error for longer displacement is considered as a candidate for future work. Further accuracy evaluation such as cutting accuracy, the effect of cutting speed and feed rate and other parameters are also considered as candidates for future works.

REFERENCES

1. Liu C, Xiang S, Lu C, Wu C, Du Z, Yang J. Dynamic and static error identification and separation method for three-axis CNC machine tools based on feature workpiece cutting. International Journal of Advanced Manufacturing Technology. 2020; 107(5–6): 2227-2238. https://doi.org/10.1007/s00170-020-05103-5
2. Martinov G M, Ljubimov A B, Martinova L I. From classic CNC systems to cloud-based technology and back. Robotics and Computer-Integrated Manufacturing. 2020; 63: 101927 https://doi.org/10.1016/j.rcim.2019.101927
3. Zhao W, Chen M, Xi W, Xi X, Zhao F, Zhang Y. Reconstructing CNC platform for EDM machines towards smart manufacturing. Procedia CIRP. 2020: 95: 161–177 https://doi.org/10.1016/j.procir.2020.03.134
4. Nurhadi H, Tarng Y S. Open-and closed-loop system of computer integrated desktop-scale CNC machine, IFAC Proceedings Volumes. 2010: 42(24):222–226. https://doi.org/10.3182/20091021-3-JP-2009.00041
5. Andersen H V, Pitkänen K. Empowering educators by developing professional practice in digital fabrication and design thinking. International Journal of Child-Computer Interaction. 2019: 21: 1-16. https://doi.org/10.1016/j.ijcci.2019.03.001
6. Ropin H, Pfleger-Landthaler A. Irsa W A. FabLab as integrative part of a learning factory. Procedia Manufacturing. 2020;45: 355–360. https://doi.org/10.1016/j.promfg.2020.04.033
7. Korkut I, Donertas M A. The influence of feed rate and cutting speed on the cutting forces, surface roughness and tool-chip contact length during face milling. Materials and Design. 2007; 28(1): 308-312. https://doi.org/10.1016/j.matdes.2005.06.002
8. Zmarzły P. Technological heredity of the turning process, Tehnicki Vjesnik. 2020; 27(4): 1194–1203.
9. Mori M, Yamazaki K, Fujishima M, Liu J, Furukawa N. A study on development of an open servo system for intelligent control of a CNC machine tool. CIRP Annals - Manufacturing Technology. 2001; 50(1): 247–250. http://dx.doi.org/10.1016/S0007-8506(07)62115-5
10. Zhou Q. Application of PLC in the CNC machine tool control system. Applied Mechanics and Materials. 2012; 182-183: 902–905. https://doi.org/10.4028/www.scientific.net/AMM.182-183.902
11. Xu HH, Dai C. Research on precision detection and error compensation technology for 3-axis CNC milling machine, Applied Mechanics and Materials. 2014: 455; 505–510. https://doi.org/10.4028/www.scientific.net/AMM.455.505
12. Ibaraki S, Oyama C, Otsubo H. Construction of an error map of rotary axes on a five-axis machining center by static R-test. International Journal of Machine Tools and Manufacture. 2011; 51(3): 190–200. http://dx.doi.org/10.1016/j.ijmachtools.2010.11.011
13. ISO 230-1. Test code for machine tools — Part 1: Geometric accuracy of machines operating under no-load or quasi-static conditions; 2012.
14. Blackshaw D M S. Machine tool accuracy and repeatability-a new approach with the revision of ISO 230-2. Transactions on Engineering Sciences. 1997; 16: 91-100. https://doi.org/10.2495/LAMDAMAP970081
15. ISO 10791-4. Test conditions for machining centres-Part 4: Accuracy and repeatability of positioning of linear and rotary axes. 1998
16. ISO 10791-6:2014. Test conditions for machining centres-Part 6: Accuracy of speeds and interpolations. 2014
17. Begović E, Plančić I, Ekinović S, Ekinov E. Laser Interferometry-Measurement and Calibration Method for Machine Tools, Proc 3rd Conference "MAINTENANCE 2014", 2014; 19–28.
18. Zhang Y, Chu X, Yang S. Research of error detection and compensation of CNC machine tools based on laser interferometer, Proc in 2nd International Conference on Machinery, Materials Engineering, Chemical Engineering and Biotechnology, 2016; 285–289.
19. Lasiyah, S., Development of accuracy measurement for mini Milling CNC with Helium-Neon Laser (in Indonesian). Final Project, Department of Mecahnical Engineering, Vocational College, Gadjah Mada University. 2019
20. Winarno A, Lasiyah S, Prayoga B T, Hendaryanto I A, Sukidjo F X. Development of accuracy evaluation method for open loop educational CNC Milling Machine. Jurnal Rekayasa Mesin. 2021; 12(1): 217-225. https://doi.org/10.21776/ub.jrm.2021.012.01.23

DOI 10.2478/ama-2022-0016 acta mechanica et automatica, vol.16 no.2 (2022)

21. Stone J A, Decker J E, Gill P, Juncar P, Lewis A, Rovera G D, Viliesid M. Advice from the CCL on the use of unstabilized lasers as standards of wavelength: The helium-neon laser at 633 nm, Metrologia. 2009; 46(1): 11–18. https://doi.org/10.1088/0026-1394/46/1/002
22. Haitjema H. Calibration of displacement laser interferometer systems for industrial metrology, Sensors. 2019;19(19):1-21. https://dx.doi.org/10.3390%2Fs19194100
23. Ciddor P E, Hill R J. Refractive index of air 2 Group index, Applied Optics. 1999; 38(9): 1663-1667. https://doi.org/10.1364/AO.38.001663
24. Dobosz M, Iwasinska-Kowalska O. A new method of non-contact gauge block calibration using a fringe-counting technique: I. Theoretical basis, Optics and Laser Technology, 2010; 42(1): 141–148. https://doi.org/10.1016/j.optlastec.2009.05.012
25. Iwasinska-Kowalska O, Dobosz M. A new method of noncontact gauge block calibration using the fringe counting technique: II. Experimental verification, Optics and Laser Technology, 2010;42(1):149–155. https://doi.org/10.1016/j.optlastec.2009.05.011
26. Winarno A, Takahashi S, Matsumoto H, Takamasu K. A new measurement method to simultaneously determine group refractive index and thickness of a sample using low-coherence tandem interferometry. Precision Engineering, 2019; 55:254–259. https://doi.org/10.1016/j.precisioneng.2018.09.013
27. Ni Y, Zhou H, Shao C, Li J. Research on the Error Averaging Effect in A Rolling Guide Pair. Chinese Journal of Mechanical Engineering (English Edition). 2019; 32(72). https://doi.org/10.1186/s10033-019-0386-y

Acknowledgment: This research is funded by Ministry of Education and Culture of the Republic of Indonesia, by Research Grants of Penelitian Terapan Unggulan Perguruan Tinggi (PTUPT) 2020 (No. 1871/UN1/DITLIT/DIT-LIT/PT/2020). The authors are grateful to Sus Lasiyah for the prepare the mini CNC milling

Agustinus Winarno: https://orcid.org/0000-0001-7134-1142

Benidiktus Tulung Prayoga: https://orcid.org/0000-0002-0054-1777

Ignatius Aris Hendaryanto: https://orcid.org/0000-0003-0204-8406

DOI 10.2478/ama-2022-0017

BEARING FAULT DETECTION AND DIAGNOSIS BASED ON DENSELY CONNECTED CONVOLUTIONAL NETWORKS

Julius NIYONGABO*, Yingjie ZHANG, Jérémie NDIKUMAGENGE*****

*Doctoral School, University of Burundi, UNESCO Road No 2, Bujumbura 1550, Burundi
**College of Computer Science and Electronic Engineering, Hunan University, Lushan Road (S), Changsha 410082, China
***Faculty of Engineering Science, Department of Information and Communication Technology, University of Burundi, UNESCO Road No 2, Bujumbura 1550, Burundi

niyojule@gmail.com, zhangyj@hnu.edu.cn, jeremie.ndikumagenge@ub.edu.bi

received 18 December 2021, revised 3 February 2022, accepted 12 February 2022

Abstract: Rotating machines are widely used in today's world. As these machines perform the biggest tasks in industries, faults are naturally observed on their components. For most rotating machines such as wind turbine, bearing is one of critical components. To reduce failure rate and increase working life of rotating machinery it is important to detect and diagnose early faults in this most vulnerable part. In the recent past, technologies based on computational intelligence, including machine learning (ML) and deep learning (DL), have been efficiently used for detection and diagnosis of bearing faults. However, DL algorithms are being increasingly favoured day by day because of their advantages of automatically extracting features from training data. Despite this, in DL, adding neural layers reduces the training accuracy and the vanishing gradient problem arises. DL algorithms based on convolutional neural networks (CNN) such as DenseNet have proved to be quite efficient in solving this kind of problem. In this paper, a transfer learning consisting of fine-tuning DenseNet-121 top layers is proposed to make this classifier more robust and efficient. Then, a new intelligent model inspired by DenseNet-121 is designed and used for detecting and diagnosing bearing faults. Continuous wavelet transform is applied to enhance the dataset. Experimental results obtained from analyses employing the Case Western Reserve University (CWRU) bearing dataset show that the proposed model has higher diagnostic performance, with 98% average accuracy and less complexity.

Key words: bearing, deep learning, machine learning, transfer learning, fault detection and diagnosis, CWRU dataset

1. INTRODUCTION

With the rapid growing of industrialisation, electrical machines are always present in industries and electrified transportation systems. Often, these machines operate under harsh environments and complex conditions, such as high ambient temperature, high moisture and overload, which can eventually result in motor malfunctions that lead to high maintenance costs, severe financial losses and safety hazards [1]. Many research projects have revealed that rolling element bearings, also known as bearings, are the most vulnerable parts of electrical machines, accounting for high failure rate and downtime. These studies have also shown that bearing fault is the most common fault type and is responsible for 30–40% of all electrical machine failures [1,2]. Fig. 1. illustrates the structure of a rolling element bearing with four types of common scenarios of misalignment that are likely to cause bearing failures.

Since bearing is the most vulnerable component in a motor drive system, fault detection and diagnosis of bearings have become an essential part of development and engineering research [3]. Therefore, signal processing based–methods, machine learning (ML)–based techniques and deep learning (DL)–based approaches have been proposed and implemented. Signal processing–based methods implement fault diagnosis by detecting characteristic frequencies which are related to the faults [2]. For example, Li et al. [4] developed a variational mode decomposition–based bearing fault diagnosis method, which can effectively identify fault frequencies. However, the signal processing–based methods need to rely on professional knowledge, and it is arduous for these methods to realise accurate fault diagnosis under an actual strong noise environment [5]. On the other hand, ML-based approaches and DL, known also as intelligent methods, can perform the fault diagnosis task without the fault-related characteristics' frequencies and prior physical knowledge [2].

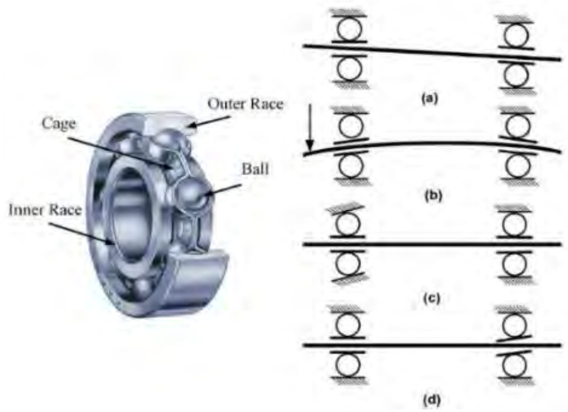


Fig. 1. Structure of a rolling element bearing with four types of common scenarios of misalignment that are likely to cause bearing failures: (a) Misalignment (out-of-line), (b) shaft deflection, (c) crooked or titled out race and (d) crooked or titled inner race [1]

Recently, various methods based on ML algorithms, including artificial neural networks (ANN), principal component analysis (PCA) and support vector machines (SVM), have been used successfully to make intelligent decisions regarding the presence of bearing faults [6-8]. Most of the literature applying these ML algorithms report satisfactory results with a classification accuracy of >90% [1]. However, one of the disadvantages of classical ML techniques is that they usually demand sophisticated manual feature engineering, which unavoidably requires expert domain knowledge and numerous human efforts.

To achieve better performance, DL methods with automated feature extraction capabilities have recently received much interest for bearing faults diagnosis [9,10]. DL is a subfield of ML that is inspired by ANN, which in turn are inspired by biological neural networks [11-15]. In the recent decade, deep Convolutional Neural Networks (CNN) have been classified as the most utilised models of DL and have been successfully applied to identify bearing faults. Moreover, many variations of CNN are employed in bearing fault diagnosis.

In this paper, after applying transfer learning techniques on DenseNet-121 architecture, which is one of the latest discoveries in neural network architectures and solves challenges caused by the depth of CNN, we propose a new approach to detect and diagnose bearing faults. We show the effectiveness of our model by conducting its evaluation and compare the experimental results to other previous CNN-based models.

The rest of this article is structured as follows: In section II, we briefly explain CNN architecture. Section III contains an overview of DenseNet algorithm and details of the proposed approach. In section IV, the most popular bearing dataset used in this work is presented. In section V, results are discussed. Finally, the conclusion is presented in section VI.

2. CNN ARCHITECTURE

CNN are feed-forward neural networks in which information flow takes place in one direction only, from their inputs to their outputs. They are composed by an input layer, convolutional layers, pooling layers, fully connected layers and an output layer [16]. The typical architecture of a CNN-based bearing fault classifier is illustrated in Fig. 2. The convolutional layer processes data input and passes the result to the next layer. The pooling layer has a function of compressing data in order to reduce its size and parameters. It is also charged to control overfitting. Often, max-pooling is the most-intensively used convolution operation in CNN. The fully connected layer is on the last position and looks like the perceptron. Its inputs come from the final convolutional layer. At the end of CNN, the SoftMax operator is used for classification problems. Since 2016, many papers applying CNN bearing fault diagnosis [17-25] have proliferated in the research domain.

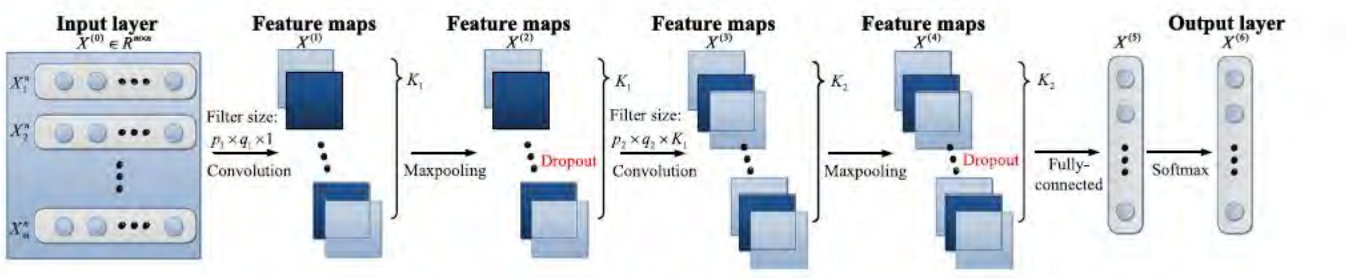


Fig. 2. Architecture of CNN-based fault diagnosis model (1). CNN, convolutional neural networks

The most common CNN architectures developed in the last decade are Alexnet (2012), Inception (2014), VGGNet (2014), ResNet (2015) and DenseNet (2017). Since AlexNet, the number of layers has increased and CNN architecture went deeper and deeper. However, deep networks become complex and difficult to be trained because of the vanishing gradient problem or dead neurons; the accuracy starts saturating and then degrades also. DenseNet has been proposed [26] to overcome this problem and easily train deep CNN.

3. OVERVIEW OF DENSENET AND THE PROPOSED MODEL

3.1. Overview of DenseNet

DenseNets or Densely Connected Convolution networks were introduced in 2017 by Gao Huang Liu, Zhuang Liu, Laurens van der Maaten and Kilian Q. Weinberger in their paper “Densely Connected Convolutional Networks” [26].

In comparison with traditional convolutional networks, which have L layers with L connections, one between each layer and its subsequent layer, DenseNet has L(L+1)/2 direct connections, and each layer is directly connected to every other layer.

DenseNet architecture is mainly constituted by Dense Block and the transition layers. The transition layers consist of a Batch-Normalisation layer, and 1 × 1 convolution followed by a 2 × 2 average pooling layer. The first part of DenseNet architecture is constituted by 7 × 7 convolution, and a stride 2 layer followed by a 3 × 3 maxpool, stride 2 layer. After the four dense blocks comes a classification layer. Inside each Dense Block and transition layer, the convolution operations are performed [26].

The main advantages of DenseNet are that it overcomes the vanishing gradient problem and does not require many parameters to train the model. Moreover, variation in the input of layers as a result of concatenated feature maps prevents the model from succumbing to the overfitting problem. The most popular DenseNet architectures are DenseNet-121, DenseNet-169, DenseNet-201 and DenseNet-264. Tab. 1 shows that each DenseNet architecture has four dense blocks with a varying number of layers, and Transition Layers are added between these blocks. Thus, DenseNet-121 consists of [6, 12, 24, 16] layers in the four dense blocks, DenseNet-169 has [6, 12, 32, 32] layers and DenseNet-201 consists of [6, 12, 48, 32] layers, whereas DenseNet-161 has [6, 12, 36, 24] layers.

DOI 10.2478/ama-2022-0017

Tab. 1. DenseNet Architectures for ImageNet [26]

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112×112	7×7 conv, stride 2			
Pooling	56×56	3×3 max pool, stride 2			
Dense Block (1)	56×56	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times6$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{vonv}\end{bmatrix}\times6$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times6$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times6$
Transition Layer (1)	56×56	1×1 conv			
	28×28	2×2 average pool, stride 2			
Dense Block (2)	28×28	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times12$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times12$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times12$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times12$
Transition Layer (2)	28×28	1×1 conv			
	14×14	2×2 average pool, stride 2			
Dense Block (3)	14×14	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times24$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times32$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times48$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times36$
Transition Layer (3)	14×14	1×1 conv			
	7×7	2×2 average pool, stride 2			
Dense Block (4)	7×7	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times16$	$\begin{bmatrix}1\times1\ \text{conv}\\1\times1\ \text{conv}\end{bmatrix}\times32$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times32$	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times24$
Classification Layer	1×1	7×7 global average pool			
		1000 D fully − connected, softmax			

3.2. The Proposed Model

In this article, by using transfer learning techniques, we fine-tuned DenseNet-121 [26]. To be precise, the following strategies are used:

- DenseNet-121's original architecture is implemented.
- Weights are initialised randomly.
- The last fully connected layer of DenseNet-121 is replaced with another new layer adapted to our classification problem.
- A flatten layer is added before the last fully connected layer.

Before training and evaluation of the proposed model, a transformation of the time domain data to wavelet domain is realised, and thereafter, we train and evaluate the new architecture on wavelet domain image for fault diagnosis. As the number of classes for Case Western Reserve University (CWRU) bearing data is 10, in the proposed approach we use a 10 Dense fully connected layer to detect and diagnose different faults of bearing. SoftMax is utilised for a classification layer. A comparison with other CNN-based models is also done. The architecture of the proposed model is shown in Tab. 2 and Fig. 3.

Tab. 2. DenseNet Architecture of the proposed model

Layers	Output Size	Proposed Model
Convolution	112×112	7×7 conv, stride 2
Pooling	56×56	3×3 max pool, stride 2
Dense Block (1)	56×56	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times6$
Transition Layer (1)	56×56	1×1 conv
	28×28	2×2 average pool, stride 2
Dense Block (2)	28×28	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times12$
Transition Layer (2)	28×28	1×1 conv
	14×14	2×2 average pool, stride 2
Dense Block (3)	14×14	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times24$
Transition Layer (3)	14×14	1×1 conv
	7×7	2×2 average pool, stride 2
Dense Block (4)	7×7	$\begin{bmatrix}1\times1\ \text{conv}\\3\times3\ \text{conv}\end{bmatrix}\times16$
Classification Layer	1×1	7×7 global average pool
	Flattening	
		10 Dense fully − connected, softmax

4. DATASET AND IMPLEMENTATION

4.1. Dataset

Data is the fundamental unit and the foundation for all ML or DL architectures. Deep networks and DL algorithms are influential ML structures that work most excellently if trained on vast amounts of data. With a small training dataset, we get limited sample variations, and the network efficiency decreases. So, the quantity and the frequency of data availability have a vital role in DL applications. Generally, the more the data, the better the accuracy [2].

To evaluate the performance of the model, in this research work, we used the popular bearing dataset from CWRU Bearing Data Center. The CWRU bearing dataset has become one of the

sciendo

most popular datasets for machine fault diagnosis since it was published. The experimental setup for collecting the CWRU bearing dataset, consisting of an electric motor on left, a torque transducer/encoder in the middle and a dynamometer on the right, is illustrated in Fig. 4.

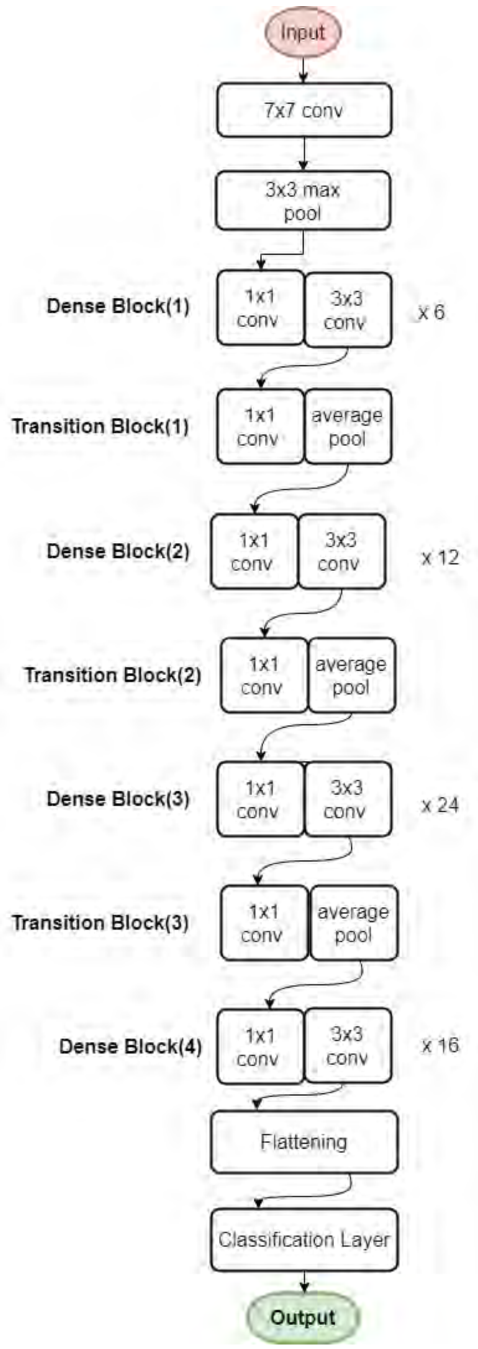


Fig. 3. Structure of the proposed model

The single point motor bearing faults simulated by the electro-discharge machining were tested in this platform, including inner-race fault (IF), outer-race fault (OF) and ball fault (BF). Each fault has three types: 7 mils, 14 mils, 21 mils (1 mil = 0.001 inches). The sampling frequencies of 12 kHz and 48 kHz were used for the collection of data. For the drive-end bearing experiments, data were collected at 12,000 and 48,000 samples/s. Fan-end data were collected at 12,000 samples/s. For the normal-baseline, the data collection rate was 48,000 samples/s [2].

Fig. 4. Experimental setup for collecting the CWRU bearing dataset. CWRU, Case Western Reserve University

In this work, we selected 48 kHz drive-end bearing fault data. Normal data collected with 1 hp load have also been used. We applied continuous wavelet transform on the dataset consisting of 4,600 data samples and 10 classes are considered: Ball defect (0.007 inch, load: 1 hp), Ball defect (0.014 inch, load: 1 hp), Ball defect (0.021 inch, load: 1 hp), IF (0.007 inch, load: 1 hp), IF (0.014 inch, load: 1 hp), IF (0.021 inch, load: 1 hp), Normal (load: 1 hp), OF (0.007 inch, load: 1 hp, data collected from 6 O'clock position), OF (0.014 inch, load: 1 hp, 6 O'clock) and OF (0.021 inch, load: 1 hp, 6 O'clock) [13].

4.2. Implementation

In this research paper, we implemented our model with Tensorflow (2.3.0) environments using Jupyter Notebooks and Python on a computer with the following properties:

- Processor: Intel® Core (TM) i5-8250U CPU @ 1.60GHz 1.80 GHz
- Installed RAM: 12.0 Go (11.9 Go Usable)

The number of parameters of the model is as follows: total: 7.047.370, trainable: 6.965.898 and non-trainable: 81.472. The following hyperparameters are used: number of filters: 128, kernel size:1, strides: 2, zero-padding: same, batch size: 64 and learning rate: 0.007.

The activation function used in the model is ReLU. The size of the data is (4600, 32, 32). Eighty percent of the total data is taken as training data and 20% as test set [13,24]. The number of epochs for all models in this paper is 50, the optimiser is SGD and momentum is initialised at 0.5.

5. RESULTS AND DISCUSSION

In this part, we are going to present and discuss the experimental results. After 10 iterations of running the proposed model under the above-mentioned conditions, an average accuracy of 98.57% is obtained. The curves of the training and validation accuracy are shown in Fig. 5. In Fig. 6 the train and validation loss curves are illustrated.

For the reliability of the experiment, confusion matrix of the considered model is created and can be seen in Fig. 7. The comparison of fault diagnosis accuracy with three previous DL Algorithms based on CNN is presented in Tab. 3, from which it is clear that the accuracy of the proposed model is higher when compared with the AlexNet, VGG-16 and ResNet-50 models. Moreover, from Fig. 8, we can confirm that the proposed approach achieves high accuracy.

DOI 10.2478/ama-2022-0017

Tab. 3. Fault diagnosis average accuracy results of AlexNet, VGG-16, ResNet-50 and the proposed model

Method	Iter1	Iter2	Iter3	Iter4	Iter5	Iter6	Iter7	Iter8	Iter9	Iter10	Average
AlexNet	0.9110	0.9430	0.9620	0.9040	0.9280	0.9690	0.9710	0.9280	0.9690	0.9440	0.9349
VGG-16	0.9010	0.9390	0.9560	0.9640	0.9640	0.9650	0.9680	0.9680	0.9680	0.9660	0.9559
ResNet-50	0.9340	0.9440	0.9030	0.9630	0.9760	0.9700	0.9780	0.9840	0.9700	0.9760	0.9603
Proposed model	**0.9750**	**0.9620**	**0.9730**	**0.9670**	**0.9780**	**0.9810**	**0.9790**	**0.9850**	**0.9760**	**0.9810**	**0.9857**

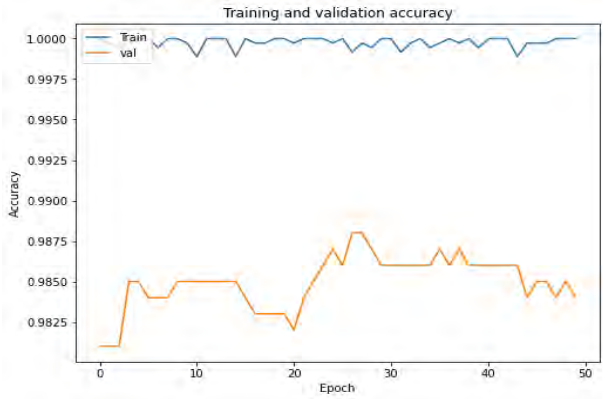


Fig. 5. Curves of training and validation accuracy during training

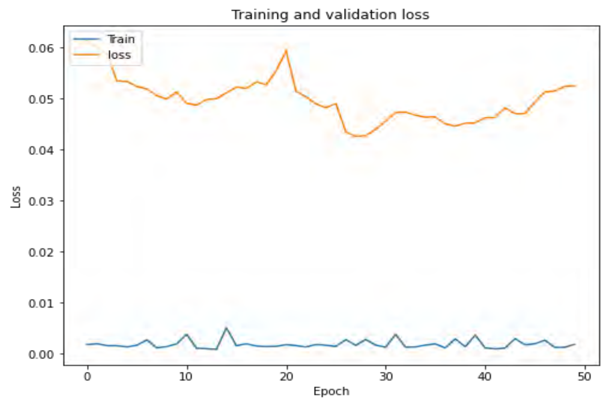


Fig. 6. Curves of training and validation loss during training

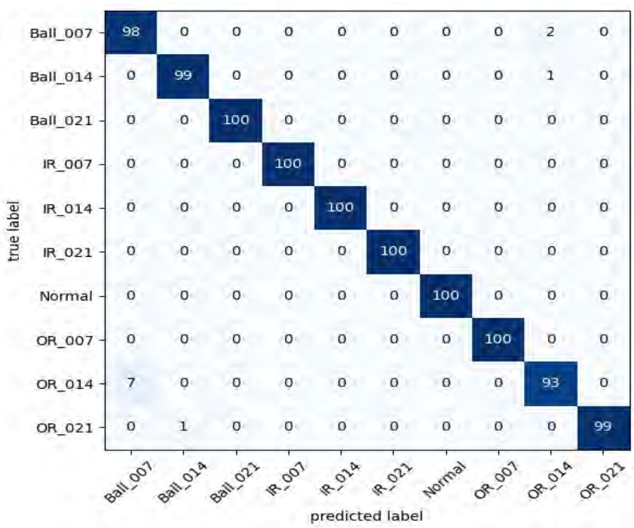


Fig. 7. Confusion matrix of DenseNet model

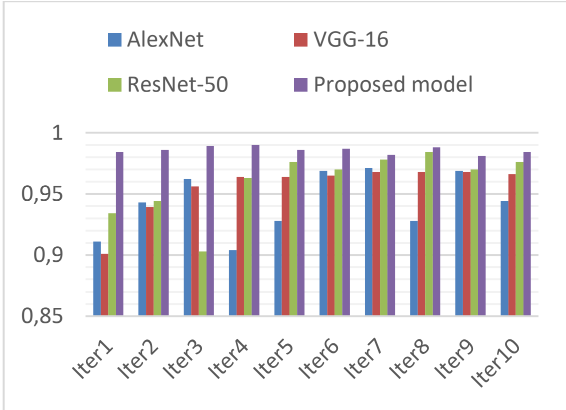


Fig. 8. The average classification accuracies

6. CONCLUSION

The early detection, diagnosis and visualisation of bearing faults can significantly reduce downtime and maintenance costs in industry. In this research, using DL technologies and transfer learning, we presented an intelligent fault diagnosis method based on deep CNN applied to rolling element bearings. We applied the model to the CWRU bearing dataset, and the simulation results reveal that the fine-tuned DenseNet-121 model performs an efficient classification for bearing faults. Moreover, it is observed that, in comparison with three previous DL models of the same family, this method has a higher classification accuracy and identifies different faults of bearing correctly.

Future researches will consist of improving the structure of the model and applying the new model on other bearing datasets.

REFERENCES

1. Zhang S, Zhang S, Wang B, Habetler TG. Deep Learning Algorithms for Bearing Fault Diagnosticsx - A Comprehensive Review. IEEE Access. 2020;8:29857–81.
2. Zhang Z, Li H, Chen L, Han P. Shrinkage Networks. 2021;2021(DI).
3. Neupane D, Seok J. Bearing fault detection and diagnosis using case western reserve university dataset with deep learning approaches: A review. IEEE Access. 2020;8:93155–78.
4. Li G, Tang G, Luo G, Wang H. Underdetermined blind separation of bearing faults in hyperplane space with variational mode decomposition. Mech Syst Signal Process. 2019;120:83–97. https://doi.org/10.1016/j.ymssp.2018.10.016
5. Huang T, Fu S, Feng H, Kuang J. Bearing fault diagnosis based on shallow multi-scale convolutional neural network with attention. Energies. 2019;12(20).
6. Awadallah MA, Morcos MM. Application of AI tools in fault diagnosis of electrical machines and drives - An overview. IEEE Trans Energy Convers. 2003;18(2):245–51.

7. Batista L, Badri B, Sabourin R, Thomas M. A classifier fusion system for bearing fault diagnosis. Expert Syst Appl [Internet]. 2013;40(17):6788–97. http://dx.doi.org/10.1016/j.eswa.2013.06.033
8. Liu R, Yang B, Zio E, Chen X. Artificial intelligence for fault diagnosis of rotating machinery: A review. Mech Syst Signal Process. 2018;108:33–47. https://doi.org/10.1016/j.ymssp.2018.02.016
9. Bansal N, Sharma A, Singh RK. A Review on the Application of Deep Learning in Legal Domain. IFIP Adv Inf Commun Technol. 2019;559:374–81.
10. Zhao R, Yan R, Chen Z, Mao K, Wang P, Gao RX. Deep Learning and Its Applications to Machine Health Monitoring: A Survey. 2016;14(8):1–14. Available from: http://arxiv.org/abs/1612.07640
11. Brownlee J. What is Deep Learning? August 14, 2020 . Available from: https://machinelearningmastery.com/what-is-deep-learning/, October 15 2021.
12. Great Learning Team. Introduction to Resnet or Residual Network. Sep 28. 2020, Available online: https://www.mygreatlearning.com/blog/resnet/, October 15, 2021. 2021;2021.
13. Sahoo B. Fault Diagnosis using Deep Learning on raw time domain data. July 15 2020. Available on line: https://github.com/biswajitsahoo1111/cbm_codes_open/blob/master/notebooks/Deep_learning_based_fault_diagnosis_using_CNN_on_raw_time _domain_data. 2021.
14. Chen Z, Cen J, Xiong J. Rolling Bearing Fault Diagnosis Using Time-Frequency Analysis and Deep Transfer Convolutional Neural Network. 2020;8.
15. Singhal G. Introduct ion to DenseNet with TensorFlow. May 6, 2020, Available online: https://www.pluralsight.com/guides/introduction-to-densenet-with-tensorflow, October 25, 2021.
16. Si L, Xiong X, Wang Z. Tan C. A Deep Convolutional Neural Network Model for Intelligent Discrimination between Coal and Rocks in Coal Mining Face. Math Probl Eng. 2020.
17. Guo X, Chen L, Shen C. Application To Bearing Fault Diagnosis. Measurement [Internet]. 2016; Available from: http://dx.doi.org/10.1016/j.measurement.2016.07.054
18. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Loccufier M, Verstockt S, et al. Convolutional Neural Network Based Fault Detection for Rotating Machinery. J Sound Vib. 2016;377:331–45.
19. Liu R, Meng G, Yang B, Sun C, Chen X. Dislocated Time Series Convolutional Neural Architecture: An Intelligent Fault Diagnosis Approach for Electric Machine. IEEE Trans Ind Informatics. 2017;13(3):1310–20.
20. Lu C, Wang Z, Zhou B. Intelligent fault diagnosis of rolling bearing using hierarchical convolutional network based health state classification. Adv Eng Informatics. 2017;32:139–51.
21. Wang H, Xu J, Yan R, Sun C, Chen X. Intelligent bearing fault diagnosis using multi-head attention-based CNN. Procedia Manuf. 2020;49:112–8. https://doi.org/10.1016/j.promfg.2020.07.005
22. Zilong Z, Wei Q. Intelligent fault diagnosis of rolling bearing using one-dimensional multi-scale deep convolutional neural network based health state classification. ICNSC 2018 - 15th IEEE Int Conf Networking, Sens Control. 2018;(April):1–6.
23. Li S, Liu G, Tang X, Lu J, Hu J. An ensemble deep convolutional neural network model with improved D-S evidence fusion for bearing fault diagnosis. Sensors (Switzerland). 2017;17(8).
24. Magar R, Ghule L, Li J, Zhao Y, Farimani AB. FaultNet: A Deep Convolutional Neural Network for Bearing Fault Classification. IEEE Access. 2021;9(October):25189–99.
25. Guo S, Yang T, Gao W, Zhang C, Zhang Y. An intelligent fault diagnosis method for bearings with variable rotating speed based on pythagorean spatial pyramid pooling CNN. Sensors (Switzerland). 2018;18(11).
26. Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition. CVPR 2017. 2017:2261–9.

Acknowledgements: This research was jointly funded by National Key Research and Development Program of China, grant number 2019YFE0105300.

Julius Niyongabo: https://orcid.org/0000-0002-1429-9918

Yingjie Zhang: https://orcid.org/0000-0002-4170-6152

Jérémie Ndikumagende: https://orcid.org/0000-0003-1094-0066

DOI 10.2478/ama-2022-0018

DESIGN OF THREE CONTROL ALGORITHMS FOR AN AVERAGING TANK WITH VARIABLE FILLING

Michał KOLANKOWSKI*, **Robert PIOTROWSKI***

*Faculty of Electrical and Control Engineering, Gdańsk University of Technology,
ul. G. Narutowicza 11/12, 80-233 Gdańsk, Poland

michkola1999@gmail.com, robert.piotrowski@pg.edu.pl

received 3 January 2022, revised 28 February 2022, accepted 6 March 2022

Abstract: An averaging tank with variable filling is a nonlinear multidimensional system and can thus be considered a complex control system. General control objectives of such object include ensuring stability, zero steady-state error, and achieving simultaneously shortest possible settling time and minimal overshoot. The main purpose of this research work was the modeling and synthesis of three control systems for an averaging tank. In order to achieve the intended purpose, in the first step, a mathematical model of the control system was derived. The model was adapted to the form required to design two out of three planned control systems by linearization and reduction of its dimensions, resulting in two system variants. A multivariable proportional-integral-derivative (PID) control system for the averaging tank was developed using optimization for tuning PID controllers. State feedback and output feedback with an integral action control system for the considered control system was designed using a linear-quadratic regulator (LQR) and optimization of weights. A fuzzy control system was designed using the Mamdani inference system. The developed control systems were tested using theMATLAB environment. Finally, the simulation results for each control algorithm (and their variants) were compared and their performance was assessed, as well as the effects of optimization in the case of PID and integral control (IC) systems.

Key words: control system, fuzzy control system, integral control system, LQR, mathematical model, PID control system, state feedback controller, tank with variable filling

1. INTRODUCTION

An averaging tank with variable filling is a tank containing a substance with variable component concentration. The concentration is assumed to be even in the entire volume of the substance. The considered tank is with variable filling and thus the volume of substance contained in the tank is variable and the system's performance is not affected by it not being constant. Averaging in the tank may be achieved by mixing.

Averaging tanks are widely used, primarily in wastewater treatment for stabilizing the composition of wastewater. It is important because the technological parameters of the treatment process are determined based on the average composition of wastewater. Moreover, the averaging tank ensures a steady flow of wastewater to the further stages of the treatment process despite the input flow of wastewater to the tank being variable [1].

Because averaging tank is a nonlinear multiple-input–multiple-output (MIMO) system, it may be considered a complex control system. Traditional control methods may thus be insufficient, and modern, more advanced control algorithms should be applied.

Designing control algorithms for tank systems is a widely researched and relevant topic. In Astrom and Hagglund [2], proportional-integral-derivative (PID) controllers were applied for tank control. Multivariable PID was proposed for the control of a tank with heating [3]. The Control system for the quadruple tank was designed in Johansson [4] and Saeed et al. [5] using multi-loop, decentralized proportional-integral (PI) control. Moreover, in Saeed et al. [5], it was compared with generalized predictive control (GPC), which is an optimal control method. In Meenatchi Sundaram and Venkateswaran [6], a Smith predictor for a system composed of three tanks was implemented. In Janani [7], the control of a two-tank system was achieved using a state feedback structure. In Bojan-Dragos et al. [8] and Berk et al. [9], fuzzy PID control systems for a vertical two-tank system and a single tank were designed.

The paper is a further development of the research works presented in Kolankowski and Piotrowski [10], which describes the modeling of the system in abridged form, the design of integral control (IC) system, and control results assessment. This paper also includes the design of multivariable PID control (using optimization for PID controllers tuning) and fuzzy control algorithms. Moreover, it compares the results of the developed control systems and summarizes their performance.

The structure of this paper is as follows. The derivation and implementation of the mathematical model of the averaging tank are described in Section 2. The design of control systems is presented in Section 3. In Section 4, the control results are discussed. The last section presents the conclusions.

2. DESCRIPTION AND MODELLING OF TANK

A substance of variable component concentration flows into the averaging tank (see Fig. 1). A substance of component concentration equal to the average concentration in the tank flows out of the tank. It is assumed that both inflow and outflow are forced.

Input (1–3) and output (4, 5) variables of the system are described in Tab. 1. The considered system is a MIMO system with 3 input variables and 2 output variables.

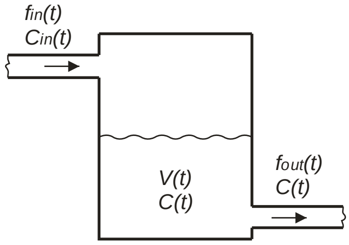


Fig. 1. Scheme of the averaging tank

Tab. 1. Symbols of variables and their units

No.	Description	Symbol	Unit
1.	Inflow rate of substance with component concentration $C_{in}(t)$	f_{in}	m^3/s
2.	Outflow rate of substance with component concentration $C(t)$	f_{out}	m^3/s
3.	Concentration of the component in the substance flowing into the tank	C_{in}	kg/m^3
4.	Volume of the substance of concentration $C(t)$ in tank	V	m^3
5.	Averaged component concentration in the tank	C	kg/m^3

2.1. Nonlinear model of tank

It is assumed that there are no disturbances and that mixing is instantaneous, and complete, lumped-parameter model can thus be applied. The model is based on the conservation of mass law (1) and the conservation of impurity law (2).

$$\frac{dV(t)}{dt} = f_{in}(t) - f_{out}(t) \tag{1}$$

$$\frac{d(V(t)C(t))}{dt} = f_{in}(t)\cdot C_{in}(t) - f_{out}(t)\cdot C(t) \tag{2}$$

The result of product $V(t) \cdot C(t)$ in equation (2) is:

$$\frac{dV(t)}{t}\cdot C(t) + V(t)\cdot\frac{dC(t)}{dt} = f_{in}(t)\cdot C_{in}(t) - f_{out}(t)\cdot C(t) \tag{3}$$

Substituting $V(t)$ from equation (1) in (3):

$$\left(f_{in}(t) - f_{out}(t)\right)\cdot C(t) + V(t)\cdot\frac{dC(t)}{t} =$$
$$= f_{in}(t)\cdot C_{in}(t) - f_{out}(t)\cdot C(t) \tag{4}$$

After transformation, equation (4) takes the form:

$$V(t)\cdot\frac{dC(t)}{dt} = f_{in}(t)\cdot C_{in}(t) - f_{in}(t)\cdot C(t) \tag{5}$$

The final model of the system takes the form (equations (5) and (1)):

$$\begin{cases} V(t)\cdot\frac{dC(t)}{dt} = f_{in}(t)\cdot(C_{in}(t) - C(t)) \\ \frac{dV(t)}{dt} = f_{in}(t) - f_{out}(t) \end{cases} \tag{6}$$

The derived model is described by a set of first-order differential equations (input/output model). For control purposes, this model is continuous, dynamic, nonlinear, deterministic, stationary, and a lumped-parameter model.

2.2. Linearization of tank model

Many control system structures and analysis methods are intended for linear systems exclusively. Thus, linearization of the control system model was necessary. It was performed by using a Taylor series expansion in the neighborhood of the equilibrium point (f_{in0}, f_{out0}, C_{in0}, C_0, V_0). It can be performed, provided that the function is differentiable in a given point [11].

The nonlinear model was written as:

$$\dot{C}(t) = f_c(V(t), \dot{V}(t), C(t), \dot{C}(t), C_{in}(t), f_{in}(t), f_{out}(t)) \tag{7}$$

$$\dot{V}(t) = f_V(V(t), \dot{V}(t), C(t), \dot{C}(t), C_{in}(t), f_{in}(t), f_{out}(t)) \tag{8}$$

where

$$f_c\left(V(t), \dot{V}(t), C(t), \dot{C}(t), C_{in}(t), f_{in}(t), f_{out}(t)\right) = \frac{f_{in}(t)}{V(t)}\cdot$$
$$C_{in}(t) - \frac{f_{in}(t)}{V(t)}\cdot C(t) \tag{9}$$

$$f_V\left(V(t), \dot{V}(t), C(t), \dot{C}(t), C_{in}(t), f_{in}(t), f_{out}(t)\right) = f_{in}(t) -$$
$$f_{out}(t) \tag{10}$$

In the equilibrium point, the static equations of the system can be written as ($\dot{V}(t) = 0;\ \dot{C}(t) = 0$):

$$f_C(V_0, 0, C_0, 0, C_{in0}, f_{in0}, f_{out0}) = \frac{f_{in0}}{V_0}\cdot C_{in0} - \frac{f_{in0}}{V_0}\cdot C_0 = 0 \tag{11}$$

$$f_V(V_0, 0, C_0, 0, C_{in0}, f_{in0}, f_{out0}) = f_{out0} - f_{out0} = 0 \tag{12}$$

The static characteristics of the system were obtained:

$$\begin{cases} \frac{f_{in0}}{V_0}\cdot(C_{in0} - C_0) = 0 \\ f_{in0} - f_{out0} = 0 \end{cases} \tag{13}$$

Hence

$$\begin{cases} C_0 = C_{in0} \\ f_{in0} = f_{out0} \end{cases} \tag{14}$$

Introducing deviation variables

$$\begin{cases} \Delta C(t) = C(t) - C_0 \\ \Delta\dot{C}(t) = \dot{C}(t) - \dot{C}_0 = \dot{C}(t) \\ \Delta V(t) = V(t) - V_0 \\ \Delta\dot{V}(t) = \dot{V}(t) - \dot{V}_0 = \dot{V}(t) \\ \Delta C_{in}(t) = C_{in}(t) - C_{in0} \\ \Delta f_{in}(t) = f_{in}(t) - f_{in0} \\ \Delta f_{out}(t) = f_{out}(t) - f_{out0} \end{cases} \tag{15}$$

The final result linearized model takes the form

$$\begin{cases} V_0\cdot\Delta\frac{dC(t)}{dt} = f_{in0}\cdot\left(\Delta C_{in}(t) - \Delta C(t)\right) + (C_{in0} - C_0)\cdot\Delta f_{in}(t) \\ \Delta\frac{dV(t)}{dt}(t) = \Delta f_{in}(t) - \Delta f_{out}(t) \end{cases} \tag{16}$$

The coordinates of the equilibrium point (V_0, C_0, C_{in0}, and f_{in0}) appear in the equations of the linearized model as parameters.

The linearized model in input/output form was converted to transfer function form, assuming zero initial conditions. The linearized model in the s-domain is

$$\begin{cases} C(s) = \frac{1}{T\cdot s+1}\cdot(k\cdot(C_{in0} - C_0)\cdot f_{in}(s) + C_{in}(s)) \\ V(s) = \frac{1}{s}\cdot(f_{in}(s) - f_{out}(s)) \end{cases} \tag{17}$$

where $k = \frac{1}{f_{in0}}; T = \frac{V_0}{f_{in0}}$.

The control algorithms presented in subsections 3.1 and 3.2 use the linearized model of the system in state space in two versions: with the assumptions of constant outflow rate ($f_{out}(t) = f_{out0}$) and constant inflow rate ($f_{in}(t) = f_{in0}$). It is thus necessary to convert the model to state space form. In order to do so, state variables were chosen as follows:

$$\Delta x_1(t) = \Delta C(t) \tag{18}$$

$$\Delta x_2(t) = \Delta V(t) \tag{19}$$

They can be written as the state vector:

$$\Delta \boldsymbol{x}(t) = \begin{bmatrix} \Delta x_1(t) \\ \Delta x_2(t) \end{bmatrix} \tag{20}$$

Input variables are

$$\Delta u_1(t) = \Delta f_{in}(t) \tag{21}$$

$$\Delta u_2(t) = \Delta f_{out}(t) \tag{22}$$

$$\Delta u_3(t) = \Delta C_{in}(t) \tag{23}$$

They can be written as the input vector:

$$\Delta \boldsymbol{u}(t) = \begin{bmatrix} \Delta u_1(t) \\ \Delta u_2(t) \\ \Delta u_3(t) \end{bmatrix} \tag{24}$$

Output variables are

$$\Delta y_1(t) = \Delta C(t) \tag{25}$$

$$\Delta y_2(t) = \Delta V(t) \tag{26}$$

They can be written as the output vector:

$$\Delta \boldsymbol{y}(t) = \begin{bmatrix} \Delta y_1(t) \\ \Delta y_2(t) \end{bmatrix} \tag{27}$$

The values of variables in the equilibrium point can be converted to state space form as follows:

$$x_{1,0} = C_0; x_{2,0} = V_0; u_{1,0} = f_{in0}; u_{2,0} = f_{out0}; u_{3,0} = C_{in0} \tag{28}$$

Assuming $f_{out}(t)$=const, the input vector is changed and takes the form:

$$\Delta \boldsymbol{u}(t) = \begin{bmatrix} \Delta u_1(t) \\ \Delta u_3(t) \end{bmatrix} \tag{29}$$

Hence, the obtained linearized model is as follows:

$$\Delta \dot{\boldsymbol{x}}(t) = \boldsymbol{A} \cdot \Delta \boldsymbol{x}(t) + \boldsymbol{B} \cdot \Delta \boldsymbol{u}(t) \tag{30}$$

$$\Delta \boldsymbol{y}(t) = \boldsymbol{C} \cdot \Delta \boldsymbol{x}(t) + \boldsymbol{D} \cdot \Delta \boldsymbol{u}(t) \tag{31}$$

where

$$\boldsymbol{A} = \begin{bmatrix} \frac{-u_{1,0}}{x_{2,0}} & 0 \\ 0 & 0 \end{bmatrix}; \boldsymbol{B} = \begin{bmatrix} \frac{u_{3,0}-x_{1,0}}{x_{2,0}} & \frac{u_{1,0}}{x_{2,0}} \\ 1 & 0 \end{bmatrix}$$

$$\boldsymbol{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{32}$$

A is state matrix; ***B*** is input matrix; ***C*** is output matrix; ***D*** is feedthrough matrix.

Assuming $f_{in}(t)$=const, the input vector is changed and takes the form:

$$\Delta \boldsymbol{u}(t) = \begin{bmatrix} \Delta u_2(t) \\ \Delta u_3(t) \end{bmatrix} \tag{33}$$

The obtained linearized model is described by equations (30) and (31), where matrices ***A***, ***B***, ***C***, and ***D*** take the form:

$$\boldsymbol{A} = \begin{bmatrix} \frac{-u_{1,0}}{x_{2,0}} & 0 \\ 0 & 0 \end{bmatrix}; \boldsymbol{B} = \begin{bmatrix} 0 & \frac{u_{1,0}}{x_{2,0}} \\ -1 & 0 \end{bmatrix}; \boldsymbol{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \boldsymbol{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{34}$$

A static equilibrium point was selected for the linearized model: (f_{in0}, f_{out0}, C_{in0}, C_0, V_0) = (0.2 m^3/s; 0.2 m^3/s; 5 kmol/m^3; 5 kmol/m^3; 2 m^3) and the initial values of variables $C(t)$ and $V(t)$ were assumed as C_p = 5 kmol/m^3; V_p = 2 m^3.

3. DESIGN OF CONTROL SYSTEMS

3.1. Multivariable PIDs control system

A control system with two PID controllers is a closed-loop system with output feedback from both process variables. The control error of each process variable is fed into an input of one of the PID controllers. The control law of a PID controller in Ideal Standard Algorithm (ISA) form is as follows [2]:

$$u(t) = K_p \cdot \left(e(t) + \frac{1}{T_i} \cdot \int_{t_0}^{t_0+t_f} e(t)dt + T_d \cdot \frac{de(t)}{dt} \right) \tag{35}$$

where $u(t)$ is control variable, $e(t)$ is control error, K_p is proportional gain, T_i is integral time, T_d is derivative time, t_o is initial time, and t_f is final time.

The block diagram of the control system is presented in Fig. 2.

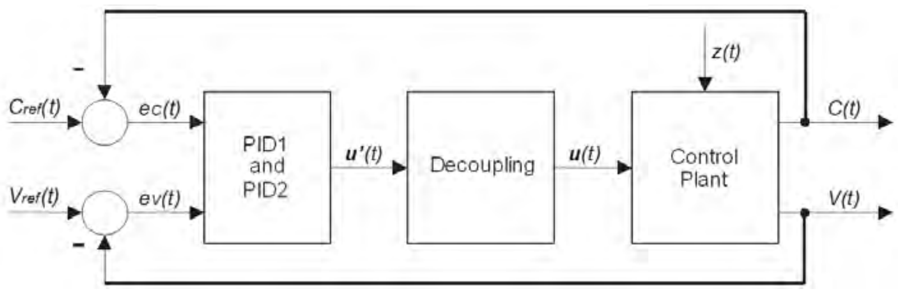


Fig. 2. Block diagram of multivariable PID control system. PID, proportional-integral-derivative

The output signals of the controllers are fed into the decoupling block. Then the signal is fed into the input of the control plant. The variable $\boldsymbol{u'}(t)$ is a vector of signals generated by PID controllers. The variable $z(t)$ represents disturbances affecting the control plant. An example of a disturbance in a physical plant is the evaporation of a substance from the tank, which affects the volume of the substance in the tank.

The multivariable PID control system was designed based on the linearized model of the system (Eq. (16)) because PID control is a method dedicated to linear systems. A PID control structure for the MIMO system was designed according to the following methodology [17]:

I. Determination of transfer function of the system

The transfer function of the system was determined based on the state space form of the linearized model, using the equation

$$\boldsymbol{G_p}(s)=\boldsymbol{C}\cdot\left[s\cdot\boldsymbol{I}^{(n)}-\boldsymbol{A}\right]^{-1}\cdot\boldsymbol{B}+\boldsymbol{D} \quad (36)$$

II. Determination of relative gain matrix (RGA) of control plant and, according to it, determination of the desired coupling of inputs and outputs

$$\boldsymbol{\Lambda}(\mathbf{G})=\boldsymbol{K}\circ(\boldsymbol{K}^{-1})^T \quad (37)$$

where $\boldsymbol{\Lambda}(\mathbf{G})$ is relative gain matrix, **K** is static gain matrix, and $\circ$ is Hadamard product.

There are two possible pairings for a Two-Input Two-Output (TITO) system: 1-1/2-2 and 1-2/2-1. Generally, a system of *n*×*n* dimensions has *n*! possible pairings.

The transfer function matrix of controllers for decentralized control takes the form:

$$\boldsymbol{G_c}(s)=\begin{bmatrix}G_{c1}(s) & 0\\ 0 & G_{c2}(s)\end{bmatrix}\text{ for 1-1/2-2 coupling} \quad (38)$$

or

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & G_{c2}(s)\\ G_{c1}(s) & 0\end{bmatrix}\text{ for 1-2/2-1 coupling} \quad (39)$$

III. Determination of decoupler structure, if necessary, i.e., if transfer function matrix is not diagonal

The transfer function of the decoupler in all control loops can be written as

$$\boldsymbol{T}(s)=\begin{bmatrix}T_{12}(s) & T_{22}(s)\\ T_{11}(s) & T_{21}(s)\end{bmatrix} \quad (40)$$

The transfer function of the open-loop system can be written as

$$\boldsymbol{G_{OL}}(s)=\boldsymbol{G_p}(s)\cdot\boldsymbol{T}(s)\cdot\boldsymbol{G_c}(s) \quad (41)$$

IV. Determination of the type of PID controller and controller's parameters which ensure the stability of the closed-loop system and the required performance of the system

The transfer function of a closed-loop control system with negative feedback is as follows

$$\boldsymbol{G_{CL}}(s)=\boldsymbol{G_p}(s)\cdot\boldsymbol{T}(s)\cdot\boldsymbol{G_c}(s)\cdot[\boldsymbol{I}+\boldsymbol{G_p}(s)\cdot\boldsymbol{T}(s)\cdot\boldsymbol{G_c}(s)]^{-1} \quad (42)$$

The error transfer function of a closed-loop system (the quotient of Laplace transform of control error and input) is as follows

$$\boldsymbol{G_E}(s)=[\boldsymbol{I}+\boldsymbol{G_p}(s)\cdot\boldsymbol{T}(s)\cdot\boldsymbol{G_c}(s)]^{-1} \quad (43)$$

The displacement error in steady state is

$$\boldsymbol{e_{ss}}=\mathbf{lim}_{s\to 0}\, s\cdot\boldsymbol{G_E}(s)\cdot\frac{A}{s} \quad (44)$$

1.1.1. Selection of controller type and parameters for $f_{out}(t)$=const system

Ad. I. The transfer function of the control plant takes the form

$$\boldsymbol{G_p}(s)=\begin{bmatrix}\frac{u_{3,0}-x_{1,0}}{u_{1,0}+s\cdot x_{2,0}} & \frac{u_{1,0}}{u_{1,0}+s\cdot x_{2,0}}\\ \frac{1}{s} & 0\end{bmatrix} \quad (45)$$

In the assumed operating point, it is

$$\boldsymbol{G_p}(s)|_{s_0}=\begin{bmatrix}0 & \frac{1}{10s+1}\\ \frac{1}{s} & 0\end{bmatrix} \quad (46)$$

Ad. II. The RGA for the considered control plant is

$$\boldsymbol{\Lambda}(\boldsymbol{G})=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \quad (47)$$

It means that only the couple of transfer functions $G_{21}(s)$ and $G_{12}(s)$ affect the value of the process variable in steady state. Hence, the appropriate pairing is 1-2/2-1 in this case. The transfer function matrix of the controller takes the form:

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & G_{c2}(s)\\ G_{c1}(s) & 0\end{bmatrix} \quad (48)$$

Ad. III. The transfer function matrix of the control plant is not diagonal, but triangular. Hence, a structure which removes the influence of the input U_1 on the output Y_1 was implemented. In order to do that, the transfer function of the decoupler $T_{11}(s)$, whose task is to compensate this influence, was determined

$$T_{11}(s)\cdot G_{p12}(s)\cdot U_{21}(s)+G_{p11}(s)\cdot U_{21}(s)=0 \quad (49)$$

From equation (49), $T_{11}(s)=-\frac{G_{p11}(s)}{G_{p12}(s)}$ was determined. A block diagram of the control system with decoupling is shown in Fig. 3.

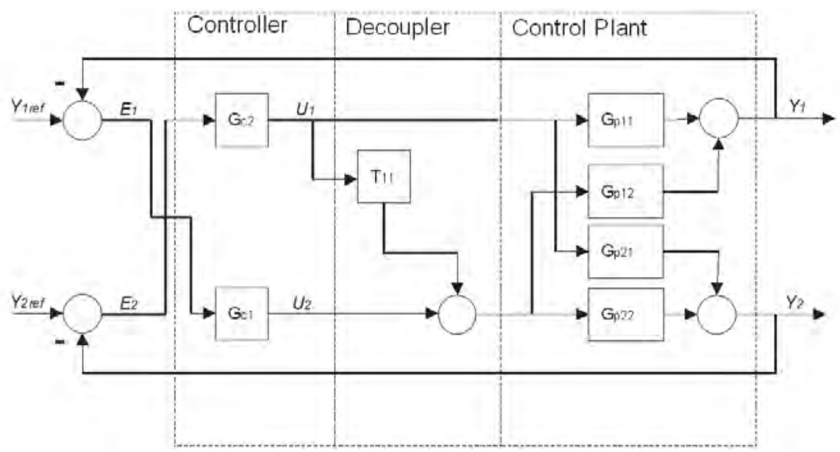


Fig. 3. Block diagram of the control system with decoupling

DOI 10.2478/ama-2022-0018

The decoupling matrix takes the form

$$\boldsymbol{T}(s)=\begin{bmatrix}1 & 0\\ -\frac{G_{p11}(s)}{G_{p12}(s)} & 1\end{bmatrix}=\begin{bmatrix}1 & 0\\ -\frac{u_{3,0}-x_{1,0}}{u_{1,0}} & 1\end{bmatrix}|_{s_0}=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} \quad (50)$$

Thus, for the selected equilibrium point, the introduction of additional decoupling does not affect the performance of the control system.

Ad. IV. Determination of the type of PID controller and controller's parameters

For type P (proportional) controllers, the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & K_{p2}\\ K_{p1} & 0\end{bmatrix} \quad (51)$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s)=\begin{bmatrix}\frac{K_{p1}\cdot u_{1,0}}{s\cdot x_{2,0}+u_{1,0}\cdot(1+K_{p1})} & 0\\ 0 & \frac{K_{p2}}{s+K_{p2}}\end{bmatrix} \quad (52)$$

The closed-loop system is stable if $K_{p1}>0$ and $K_{p2}>0$, because poles in both of its transfer functions are located in the left half-plane of the s-plane. The error transfer function of the closed-loop system is

$$\boldsymbol{G_E}(s)=\begin{bmatrix}\frac{s\cdot x_{2,0}+u_{1,0}}{s\cdot x_{2,0}+u_{1,0}\cdot(1+K_{p1})} & 0\\ 0 & \frac{s}{s+K_{p2}}\end{bmatrix} \quad (53)$$

The steady-state displacement error is

$$\boldsymbol{e_{ss}}=\begin{bmatrix}\frac{A}{1+K_{p1}} & 0\\ 0 & 0\end{bmatrix} \quad (54)$$

The control system with type P controllers thus does not ensure zero steady-state error of the process variable $y_1(t)$, which is $C(t)$.

For type proportional-integral (PI) controllers the matrix $\boldsymbol{G_R}(s)$ takes the form:

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & K_{p2}\cdot(1+\frac{1}{T_{i2}\cdot s})\\ K_{p1}\cdot(1+\frac{1}{T_{i1}\cdot s}) & 0\end{bmatrix} \quad (55)$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s)=\begin{bmatrix}\frac{K_{p1}\cdot u_{1,0}(1+T_{i1}\cdot s)}{s^2\cdot T_{i1}\cdot x_{2,0}+s\cdot T_{i1}\cdot u_{1,0}\cdot(1+K_{p1})+u_{1,0}\cdot K_{p1}} & 0\\ 0 & \frac{K_{p2}\cdot(1+T_{i2}\cdot s)}{s^2\cdot T_{i2}+s\cdot T_{i2}\cdot K_{p2}+K_{p2}}\end{bmatrix} \quad (56)$$

The closed-loop system is stable if $K_{p1}>0$, $K_{p2}>0$, $T_{i1}>0$, and $T_{i2}>0$, because poles in both of its transfer functions are located in the left half-plane of the s-plane. The error transfer function of the closed-loop system is

$$\boldsymbol{G_E}(s)=\begin{bmatrix}\frac{T_{i1}\cdot x_{2,0}\cdot s^2+T_{i1}\cdot u_{1,0}\cdot s}{s^2\cdot T_{i1}\cdot x_{2,0}+s\cdot T_{i1}\cdot u_{1,0}\cdot(1+K_{p1})+u_{1,0}\cdot K_{p1}} & 0\\ 0 & \frac{T_{i2}\cdot s^2}{s^2\cdot T_{i2}+s\cdot T_{i2}\cdot K_{p2}+K_{p2}}\end{bmatrix} \quad (57)$$

The steady-state displacement error is

$$\boldsymbol{e_{ss}}=\begin{bmatrix}0 & 0\\ 0 & 0\end{bmatrix} \quad (58)$$

The control system with PI-type controllers thus ensures zero-steady state error for both process variables.

For type proportional-integral-derivative (PID) controllers, the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & K_{p2}\cdot(1+\frac{1}{T_{i2}\cdot s}+T_{d2}\cdot s)\\ K_{p1}\cdot(1+\frac{1}{T_{i1}\cdot s}+T_{d2}\cdot s) & 0\end{bmatrix} \quad (59)$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s)=\begin{bmatrix}\frac{K_{p1}\cdot u_{1,0}(1+T_{i1}\cdot s+T_{d1}\cdot T_{i1}\cdot s^2)}{s^2\cdot T_{i1}(x_{2,0}+T_{d1}\cdot K_{p1}\cdot u_{1,0})+s\cdot T_{i1}\cdot u_{1,0}(1+K_{p1})+u_{1,0}\cdot K_{p1}} & 0\\ 0 & \frac{K_{p2}\cdot(1+T_{i2}\cdot s+T_{d2}\cdot T_{i2}\cdot s^2)}{s^2\cdot T_{i2}\cdot(1+T_{d1}\cdot K_{p1})+s\cdot T_{i2}\cdot K_{p2}+K_{p2}}\end{bmatrix} \quad (60)$$

The closed-loop system is stable if $K_{p1}>0$, $K_{p2}>0$, $T_{i1}>0$, $T_{i2}>0$, $T_{d1}>0$, and $T_{d2}>0$, because poles in both of its transfer functions are located in the left half-plane of the s-plane. The addition of a derivative term does not affect the steady-state error; thus, it is zero for PI controllers.

The general form of the control law for this system is

$$\Delta\boldsymbol{u}(t)=\begin{bmatrix}\Delta f_{in}(t)\\ \Delta C_{in}(t)\end{bmatrix}=\begin{bmatrix}K_{p2}\cdot\left(\Delta e_V(t)+\frac{1}{T_{i2}}\cdot\int_{t_0}^{t_0+t_f}\Delta e_V(t)dt+T_{d2}\cdot\frac{d\Delta e_V(t)}{dt}\right)\\ K_{p1}\cdot\left(\Delta e_C(t)+\frac{1}{T_{i1}}\cdot\int_{t_0}^{t_0+t_f}\Delta e_C(t)dt+T_{d1}\cdot\frac{d\Delta e_C(t)}{dt}\right)\end{bmatrix} \quad (61)$$

where

$$\Delta e_C(t)=\Delta C_{ref}(t)-\Delta C(t);\ \Delta e_V(t)=\Delta V_{ref}(t)-\Delta V(t). \quad (62)$$

1.1.2. Selection of controller type and parameters f or $f_{in}(t)$=const system

Ad. I. The transfer function of the control plant is described by equation (63)

$$\boldsymbol{G_p}(s)=\begin{bmatrix}0 & \frac{u_{1,0}}{u_{1,0}+s\cdot x_{2,0}}\\ -\frac{1}{s} & 0\end{bmatrix} \quad (63)$$

In the assumed equilibrium point, it is

$$\boldsymbol{G_p}(s)|_{s_0}=\begin{bmatrix}0 & \frac{1}{10\cdot s+1}\\ -\frac{1}{s} & 0\end{bmatrix} \quad (64)$$

Ad. II. The RGA for the considered control plant is as follows

$$\boldsymbol{\Lambda}(\boldsymbol{G})=\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix} \quad (65)$$

It means that only the couple of transfer functions $G_{21}(s)$ and $G_{12}(s)$ affect the value of the process variable in steady state. Hence, the appropriate pairing is 1-2/2-1 in this case. The transfer function matrix of the controller takes the form:

$$\boldsymbol{G_c}(s)=\begin{bmatrix}0 & G_{c2}(s)\\ G_{c1}(s) & 0\end{bmatrix} \quad (66)$$

DOI 10.2478/ama-2022-0018 *acta mechanica et automatica, vol.16 no.2 (2022)*

Ad. III. The transfer function matrix of the control plant is diagonal. Hence, decoupling may be skipped in this case. The decoupler matrix ***T(s)*** is an identity matrix

Ad. IV. Determination of the type of PID controller and controller's parameters.

For type P controllers, the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_c}(s) = \begin{bmatrix} 0 & K_{p2} \\ K_{p1} & 0 \end{bmatrix} \tag{67}$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s) = \begin{bmatrix} \frac{K_{p1}u_{1,0}}{s \cdot x_{2,0} + u_{1,0}(1+K_{p1})} & 0 \\ 0 & \frac{K_{p2}}{-s+K_{p2}} \end{bmatrix} \tag{68}$$

In order to ensure the stability of the closed-loop system, it is necessary to place the pole of the transfer function $G_{CL22}(s)$, which is the transfer function of the control loop for *V(t)*, in the left half-plane of the s-plane. Hence, the condition for system's stability is $K_{p2} < 0$. It can be fulfilled by placing a gain of -1 in the control loop of *V(t)*. Then the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_c}(s) = \begin{bmatrix} 0 & -K_{p2} \\ K_{p1} & 0 \end{bmatrix} \tag{69}$$

The closed-loop system is stable if $K_{p1} > 0$ and $-K_{p2} > 0$. The error transfer function of the closed-loop system is

$$\boldsymbol{G_E}(s) = \begin{bmatrix} \frac{s \cdot x_{2,0} + u_{1,0}}{s \cdot x_{2,0} + u_{1,0} \cdot (1+K_{p1})} & 0 \\ 0 & \frac{s}{s+K_{p2}} \end{bmatrix} \tag{70}$$

The steady-state displacement error is

$$e_{ss} = \begin{bmatrix} \frac{A}{1+K_{p1}} & 0 \\ 0 & 0 \end{bmatrix} \tag{71}$$

The control system with P-type controllers thus does not ensure zero steady-state error of the process variable $y_1(t)$, which is *C(t)*.

For type PI controllers, the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_R}(s) = \begin{bmatrix} 0 & -K_{p2} \cdot (1 + \frac{1}{T_{i2} \cdot s}) \\ K_{p1} \cdot (1 + \frac{1}{T_{i1} \cdot s}) & 0 \end{bmatrix} \tag{72}$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s) = \begin{bmatrix} \frac{K_{p1} \cdot u_{1,0}(1+T_{i1} \cdot s)}{s^2 \cdot T_{i1} \cdot x_{2,0} + s \cdot T_{i1} \cdot u_{1,0} \cdot (1+K_{p1}) + u_{1,0} \cdot K_{p1}} & 0 \\ 0 & \frac{K_{p2} \cdot (1+T_{i2} \cdot s)}{s^2 \cdot T_{i2} + s \cdot T_{i2} \cdot K_{p2} + K_{p2}} \end{bmatrix} \tag{73}$$

The closed-loop system is stable if $K_{p1} > 0$, $-K_{p2} > 0$, $T_{i1} > 0$, and $T_{i2} > 0$. The error transfer function of the closed-loop system is

$$\boldsymbol{G_E}(s) = \begin{bmatrix} \frac{T_{i1} \cdot x_{2,0} \cdot s^2 + T_{i1} \cdot u_{1,0} \cdot s}{s^2 \cdot T_{i1} \cdot x_{2,0} + s \cdot T_{i1} \cdot u_{1,0} \cdot (1+K_{p1}) + u_{1,0} \cdot K_{p1}} & 0 \\ 0 & \frac{T_{i2} \cdot s^2}{s^2 \cdot T_{i2} + s \cdot T_{i2} K_{p2} + K_{p2}} \end{bmatrix} \tag{74}$$

The steady-state displacement error is

$$\boldsymbol{e_{ss}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{75}$$

The control system with PI-type controllers thus ensures zero steady-state error for both process variables.

For type PID controllers, the matrix $\boldsymbol{G_c}(s)$ takes the form

$$\boldsymbol{G_c}(s) = \begin{bmatrix} 0 & -K_{p2} \cdot (1 + \frac{1}{T_{i2} \cdot s} + T_{d2} \cdot s) \\ K_{p1} \cdot (1 + \frac{1}{T_{i1} \cdot s} + T_{d2} \cdot s) & 0 \end{bmatrix} \tag{76}$$

The closed-loop transfer function takes the form

$$\boldsymbol{G_{CL}}(s) = \begin{bmatrix} \frac{K_{p1} \cdot u_{1,0}(1 + T_{i1} \cdot s + T_{d1} \cdot T_{i1} \cdot s^2)}{s^2 \cdot T_{i1} \cdot (x_{2,0} + T_{d1} \cdot K_{p1} \cdot u_{1,0}) + s \cdot T_{i1} \cdot u_{1,0} \cdot (1+K_{p1}) + u_{1,0} \cdot K_{p1}} & 0 \\ 0 & \frac{K_{p2} \cdot (1+T_{i2} \cdot s + T_{d2} \cdot T_{i2} \cdot s^2)}{s^2 \cdot T_{i2} \cdot (1+T_{d1} \cdot K_{p1}) + s \cdot T_{i2} \cdot K_{p2} + K_{p2}} \end{bmatrix} \tag{77}$$

The closed-loop system is stable if $K_{p1} > 0$, $-K_{p2} > 0$, $T_{i1} > 0$, $T_{i2} > 0$, $T_{d1} > 0$, and $T_{d2} > 0$. The addition of a derivative term does not affect the steady-state error; thus, it is zero as in the case of PI controllers.

The general form of the control law for this system is

$$\Delta\boldsymbol{u}(t) = \begin{bmatrix} \Delta f_{out}(t) \\ \Delta C_{in}(t) \end{bmatrix} = \begin{bmatrix} -K_{p2}\left(\Delta e_V(t) + \frac{1}{T_{i2}}\int_{t_0}^{t_0+t_k} \Delta e_V(t)dt + T_{d2}\frac{d\Delta e_V(t)}{dt}\right) \\ K_{p1}\left(\Delta e_C(t) + \frac{1}{T_{i1}}\int_{t_0}^{t_0+t_k} \Delta e_C(t)dt + T_{d1}\frac{d\Delta e_C(t)}{dt}\right) \end{bmatrix}. \tag{78}$$

1.1.3. Optimization of controller parameters

In order to acquire the optimal parameters of PID controllers, optimization was performed based on integral performance indices of control variables *C(t)* and *V(t)*:

$$IE = \int_0^\infty e(t)dt \tag{79a}$$

$$ISE = \int_0^\infty e^2(t)dt \tag{79b}$$

$$IAE = \int_0^\infty |e(t)|dt \tag{79c}$$

$$ITAE = \int_0^\infty t|e(t)|dt \tag{79d}$$

$$ISEG = \int_0^\infty (e^2(t) + a\dot{e}^2(t))dt;\ a=0.5 \tag{79e}$$

where *e(t)* is control error.

Optimization was performed using the method of minimizing the maximum of a set of objective functions [13]. In this case, the objective functions are the integral performance indices of process variables *C(t)* and *V(t)*. The decision variables are the parameters of both controllers. It was assumed that all decision variables must be nonnegative (as constraints). The optimization was performed for the nonlinear model of the tank, and absolute variables were converted to deviation variables, on which PID controllers operate. The performance index Integral of Squared Error Generalized (ISEG) was not used in the case of multivariable PID control because the results for it were invalid.

Given that the classical tuning methods, such as I and II Ziegler-Nichols method, are dedicated to single-input single-output (SISO) systems, rough values of PID constants, which serve as initial values of decision variables in the optimization process, were experimentally determined so that the system is stable and its performance is fine.

DOI 10.2478/ama-2022-0018

The optimization was performed for three types of the controller: P, PI, and PID and for the system with $f_{out}(t)$ = const and the system with $f_{in}(t)$ = const. Optimization results were assessed based on the values of the integral indices and the graphs of process and control variables. Best results were achieved for PI-type controllers for both versions of the control system. For the system with $f_{out}(t)$ = const, the initial constants (determined experimentally) were as given in (80) and the optimal values of the constants (optimal with respect to IE) were as given in (81).

$$\begin{cases} K_{pC} = 50 \\ T_{iC} = 1 \\ K_{pV} = 8 \\ T_{iV} = 2 \end{cases} \tag{80}$$

$$\begin{cases} K_{pC} = 74.37 \\ T_{iC} = 8.325 \cdot 10^{-4} \\ K_{pV} = 8.000 \\ T_{iV} = 0.500 \end{cases} \tag{81}$$

For the system with $f_{in}(t)$ = const, the initial constants (determined experimentally) were as given in (82) and the optimal values of the constants (optimal with respect to IE) were as shown in (83).

$$\begin{cases} K_{pC} = 50 \\ T_{iC} = 10 \\ K_{pV} = 5 \\ T_{iV} = 30 \end{cases} \tag{82}$$

$$\begin{cases} K_{pC} = 61.5 \\ T_{iC} = 7.641 \cdot 10^{-4} \\ K_{pV} = 5.034 \\ T_{iV} = 3.208 \cdot 10^{-2} \end{cases} \tag{83}$$

3.2. State feedback control system

In order to design a control system structure with output feedback and state feedback, the linearized model of the tank was converted to state-space form (see Section 2), because the control algorithm is dedicated to linear systems. It was necessary in order to build an expanded model and determine gain matrices in output feedback loop $\boldsymbol{F_i}$ and in state feedback loop $\boldsymbol{F}$. The values of these matrices may be determined using multiple methods, for instance, pole placement method or optimal regulator method. In this case, the linear-quadratic regulator (LQR) method was used [14]. Due to the fact that the number of control inputs must not exceed the number of state variables of the control, reduction of the number of control inputs of the system was necessary. It was accomplished in two ways: by assumption of $f_{out}(t)=f_{out0}$ and by assumption of $f_{in}(t)=f_{in0}$. The reduced state-space model is described by the equations (30), (31), and additionally (29) and (32) in the $f_{out}(t)$=const system and (33) and (34) in the $f_{out}(t)$=const system. The assumption of $C_{in}(t)=C_{in0}$ is inexpedient for this model because it would prevent the change of $C(t)$ since there is only one substance inflowing into the tank.

A control system structure with output feedback and full-state feedback can ensure zero steady-state error only given that the exact model of the system and the dynamics of control inputs are known. In order to ensure greater insensitivity to the uncertainty of the model of the system and the dynamics of its input variables and thus improve the system's robustness, integral action must be applied. It should be implemented in the output feedback loop, in series with the control plant. The introduction of integral action causes an increase in the rank of astatism of the control system.

IC structure is a control structure with state feedback and output feedback with integral action (see Fig. 4).

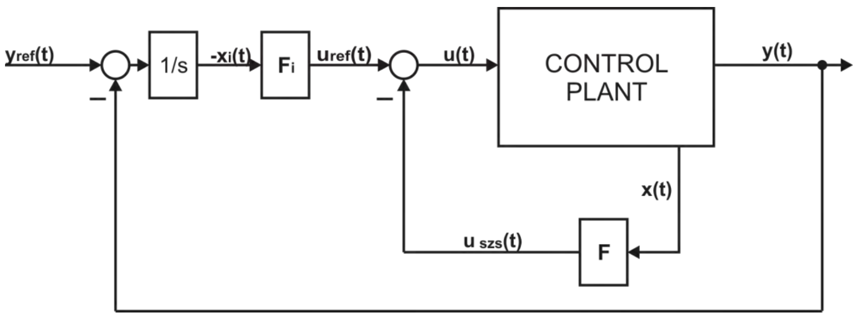


Fig. 4. Block diagram of IC system. IC, integral control

where $\boldsymbol{y_{ref}}(t)$, $\boldsymbol{x_i}(t)$, $\boldsymbol{u_{ref}}(t)$, $\boldsymbol{u_{szs}}(t)$, $\boldsymbol{F_i}$, and $\boldsymbol{F}$ are reference value vector, integral state variable vector, the component of control variable vector influenced by output feedback, the component of control variable vector influenced by state feedback, gain in output feedback loop matrix, and gain in state feedback loop matrix, respectively.

In order to formulate a description of an expanded system, it is necessary to introduce a deviation integral state variable $\Delta\boldsymbol{x_i}(t)$. It consists of integrated errors, which allow for output variable following its reference value. Then the equations of the expanded system take the form (84)–(85). The state vector of the expanded system is as given in (86) or in expanded form as shown in (87).

$$\Delta\dot{\boldsymbol{x}}(t) = \boldsymbol{A} \cdot \Delta\boldsymbol{x}(t) + \boldsymbol{B} \cdot \Delta\boldsymbol{u}(t) \tag{84}$$

$$\Delta\dot{\boldsymbol{x}}_{\boldsymbol{i}}(t) = \boldsymbol{C} \cdot \Delta\boldsymbol{x}(t) + \boldsymbol{M} \cdot \Delta\boldsymbol{u}(t) - \boldsymbol{y_{ref}}(t) \tag{85}$$

$$\Delta\boldsymbol{x_{exp}}(t) = \begin{bmatrix} \Delta\boldsymbol{x}(t) \\ \Delta\boldsymbol{x_i}(t) \end{bmatrix} = \begin{bmatrix} \Delta x_1(t) \\ \Delta x_2(t) \\ \Delta x_{i1}(t) \\ \Delta x_{i2}(t) \end{bmatrix} \tag{86}$$

$$\Delta\dot{\boldsymbol{x}}_{\boldsymbol{exp}}(t) = \boldsymbol{A_{exp}} \cdot \Delta\boldsymbol{x_{exp}}(t) + \boldsymbol{B_{exp}} \cdot \Delta\boldsymbol{u}(t) - \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot \boldsymbol{y_{ref}}(t) \tag{87}$$

where $\boldsymbol{M}$ is direct input on the output impact matrix.

The state matrix $\boldsymbol{A_{exp}}$, the input matrix $\boldsymbol{B_{exp}}$, and $\boldsymbol{M}$ of the expanded system are

$$\boldsymbol{A_{exp}} = \begin{bmatrix} \boldsymbol{A} & 0 \\ \boldsymbol{C} & 0 \end{bmatrix};\ \boldsymbol{B_{exp}} = \begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{M} \end{bmatrix};\ \text{and } \boldsymbol{M} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{88}$$

It is possible to design an IC system only if the expanded system is controllable. An expanded system can be considered controllable only if the following two conditions are met:

1. The primary system is controllable,
2. The number of control variables is no less than that of the number of process variables.

Control law takes the form

$$\Delta \boldsymbol{u}(t) = -\boldsymbol{F_{exp}} \cdot \Delta \boldsymbol{x_{exp}}(t) = -[\boldsymbol{F} \quad \boldsymbol{F_i}] \cdot \begin{bmatrix} \Delta \boldsymbol{x}(t) \\ \Delta \boldsymbol{x_i}(t) \end{bmatrix} \tag{89}$$

where $\boldsymbol{F_{exp}}$ is the gain matrix of the expanded system.

In the case of the considered system, the gain matrix takes the form

$$\boldsymbol{F_{exp}} = [\boldsymbol{F} \quad \boldsymbol{F_i}] = \begin{bmatrix} F_{11} & F_{12} & F_{i11} & F_{i12} \\ F_{21} & F_{22} & F_{i21} & F_{i22} \end{bmatrix} \tag{90}$$

The LQR method, which was selected to find the value of the $\boldsymbol{F_{exp}}$ matrix, is an optimal linear state feedback controller with a quadratic performance criterion. In general, for a linear, stationary system described by equation (30) and for the infinite time horizon, the cost function is as follows:

$$\boldsymbol{J}(\boldsymbol{u}) = \frac{1}{2} \cdot \int_0^\infty [\boldsymbol{x^T}(t) \cdot \boldsymbol{Q} \cdot \boldsymbol{x}(t) + \boldsymbol{u^T}(t) \cdot \boldsymbol{R} \cdot \boldsymbol{u}(t)]dt \tag{91}$$

where $\boldsymbol{Q}$ is a positive semidefinite, symmetric weight of state deviation $n \times n$ matrix (n is the dimension of the $\boldsymbol{x}(t)$ vector); and $\boldsymbol{R}$ is a positive definite, symmetric weight of input usage $r \times r$ matrix (r is the dimension of the $\boldsymbol{u}(t)$ vector). The first term of (91) is associated with control error. The second term is associated with the value of control variables. Infinite time was used in order to ensure that the controller stays near the equilibrium point after the initial transition state. An infinite-time LQR can be used under the condition that the system is completely controllable. Finding proper values of weight matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ allows to achieve a compromise between the control error and the control cost [14, 15].

It was necessary to design a control system in two configurations: with $f_{out}(t)$=const and with $f_{in}(t)$=const. In the case of the first configuration, the state matrix and the input matrix of the expanded system were achieved as follows:

$$\boldsymbol{A_{roz}} = \begin{bmatrix} \frac{-u_{1,0}}{x_{2,0}} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} -0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{92}$$

$$\boldsymbol{B_{roz}} = \begin{bmatrix} \frac{u_{3,0}-x_{1,0}}{x_{2,0}} & \frac{u_{1,0}}{x_{2,0}} \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{93}$$

Hence, the controllability of the expanded system was determined.

Condition I

In order to determine the controllability of the primary system, a Kalman controllability matrix for this system was created.

$$\boldsymbol{M_{CK}} = [\boldsymbol{B} \quad \boldsymbol{AB}] = \begin{bmatrix} \frac{u_{3,0}-x_{1,0}}{x_{2,0}} & \frac{u_{1,0}}{x_{2,0}} & \frac{-u_{1,0}\cdot(u_{3,0}-x_{1,0})}{{x_{2,0}}^2} & \frac{{u_{1,0}}^2}{{x_{2,0}}^2} \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{94}$$

For the previously chosen operating point, $\mathrm{rank}(\boldsymbol{M_{CK}}) = \mathrm{rank}(\boldsymbol{A}) = 2$ and thus the primary system is controllable.

Condition II

For the considered system, $\mathrm{rank}(\boldsymbol{u}) = 2$ and $\mathrm{rank}(\boldsymbol{y}) = 2$, thus the condition is met. Since both conditions are met, the expanded system is controllable. For $f_{in}(t)$=const, the system configuration state matrix of the expanded system is the same as for the $f_{out}(t)$=const configuration (see equation (92)), while the input matrix takes the form

$$\boldsymbol{B_{exp}} = \begin{bmatrix} 0 & \frac{u_{1,0}}{x_{2,0}} \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0.1 \\ -1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \tag{95}$$

Hence, the controllability of the expanded system was determined.

Condition I

In order to determine the controllability of the primary system, a Kalman controllability matrix for this system was created:

$$\boldsymbol{M_{CK}} = [\boldsymbol{B} \quad \boldsymbol{AB}] = \begin{bmatrix} 0 & \frac{u_{1,0}}{x_{2,0}} & 0 & \frac{-{u_{1,0}}^2}{{x_{2,0}}^2} \\ -1 & 0 & 0 & 0 \end{bmatrix} \tag{96}$$

For the previously chosen equilibrium point, $\mathrm{rank}(\boldsymbol{M_{CK}}) = \mathrm{rank}(\boldsymbol{A}) = 2$ and thus the primary system is controllable.

Condition II

For the considered system, $\mathrm{rank}(\boldsymbol{u}) = 2$ and $\mathrm{rank}(\boldsymbol{y}) = 2$, thus the condition is met. Since both conditions are met, the expanded system is controllable.

Manual tuning of matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ was performed based on simulation tests, which allowed to assess weight matrices' influence on control quality and performance of control system. The achieved values served as the initial values for optimization.

The assessment was conducted based on the properties of graphs (overshoot, settling time, rise time, maximum value, and change rate of control variables) and the values of integral performance indices (79a)–(79e).

For the $f_{out}(t)$=const system, the achieved values of matrices $\boldsymbol{Q}$, $\boldsymbol{R}$, $\boldsymbol{F}$, and $\boldsymbol{F_i}$ were as follows:

$$\boldsymbol{Q} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 10 \end{bmatrix}; \quad \boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 0 & 0.1 \end{bmatrix} \tag{97a}$$

$$\boldsymbol{F} = \begin{bmatrix} 0 & 3.0536 \\ 13.8661 & 0 \end{bmatrix}; \quad \boldsymbol{F_i} = \begin{bmatrix} 0 & 3.0536 \\ 10 & 0 \end{bmatrix} \tag{97b}$$

For the $f_{in}(t)$=const system, the results were

$$\mathbf{Q} = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 10 & 0 \\ 0 & 0 & 0 & 3 \end{bmatrix}; \quad \mathbf{R} = \begin{bmatrix} 2 & 0 \\ 0 & 0.1 \end{bmatrix} \tag{98a}$$

$$\mathbf{F} = \begin{bmatrix} 0 & -1.7174 \\ 14.1987 & 0 \end{bmatrix}; \quad \boldsymbol{F_i} = \begin{bmatrix} 0 & -1.2247 \\ 10 & 0 \end{bmatrix} \tag{98b}$$

Then, in order to find optimal $\boldsymbol{Q}$ and $\boldsymbol{R}$ weight matrices and, in consequence, optimal $\boldsymbol{F}$ and $\boldsymbol{F_i}$ gain matrices, optimization was performed using integral performance indices of control (79a)–(79e).

The process variables $C(t)$ and $V(t)$ were chosen as objective functions for optimization. Weights of matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ ($Q1$, $Q2$, $Q3$, $Q4$, $R1$, $R2$) were chosen as decision variables.

Constraints

$$Q_1 \geq 0; \ Q_2 \geq 0; \ Q_3 \geq 0; \ Q_4 \geq 0; \ R_1 > 0; \ R_2 > 0 \tag{99}$$

The constraints are a consequence of LQR's properties, from which stem the following conditions:

– the $\boldsymbol{Q}$ matrix must be positive semidefinite;
– the $\boldsymbol{R}$ matrix must be positive definite.

DOI 10.2478/ama-2022-0018

Optimization was performed using the method of minimizing the maximum out of a set of objective functions (the same as in Section 3.1.3) [16].

For the $f_{out}(t)$=const system, the lowest integral performance index values for both control variables and the best control quality in terms of settling time and overshoot were achieved for the ITAE index. The optimal values for $\boldsymbol{Q}$, $\boldsymbol{R}$, $\boldsymbol{F}$, and $\boldsymbol{F_i}$ matrices are

$$\boldsymbol{Q} = \begin{bmatrix} 10^{-6} & 0 & 0 & 0 \\ 0 & 0.1309 & 0 & 0 \\ 0 & 0 & 12.04 & 0 \\ 0 & 0 & 0 & 10.88 \end{bmatrix};$$

$$\boldsymbol{R} = \begin{bmatrix} 10^{-6} & 0 \\ 0 & 10^{-6} \end{bmatrix} \tag{100a}$$

$$\boldsymbol{F} = \begin{bmatrix} 0 & 370.7667 \\ 262.4497 & 0 \end{bmatrix};\ \boldsymbol{F_i} = \begin{bmatrix} 0 & 3298.0 \\ 3470.2 & 0 \end{bmatrix} \tag{100b}$$

For the $f_{in}(t)$=const system, analogically, the lowest integral performance index values were achieved for the ISE index. The optimal values for $\boldsymbol{Q}$, $\boldsymbol{R}$, $\boldsymbol{F}$, and $\boldsymbol{F_i}$ matrices are:

$$\boldsymbol{Q} = \begin{bmatrix} 10^{-6} & 0 & 0 & 0 \\ 0 & 0.3995 & 0 & 0 \\ 0 & 0 & 0.1092 & 0 \\ 0 & 0 & 0 & 4.528 \end{bmatrix};$$

$$\boldsymbol{R} = \begin{bmatrix} 0.0113 & 0 \\ 0 & 10^{-6} \end{bmatrix} \tag{101a}$$

$$\boldsymbol{F} = \begin{bmatrix} 0 & -8.6885 \\ 256.1044 & 0 \end{bmatrix};\ \boldsymbol{F_i} = \begin{bmatrix} 0 & -20.0 \\ 3305.0 & 0 \end{bmatrix}. \tag{101b}$$

3.3. Fuzzy control system

A fuzzy logic controller (FLC) is used for calculating control variables. A fuzzy model consists of three blocks: fuzzification, inference (with rule base), and defuzzification (see Fig. 5), where $e_C(t)$, $de_C(t)/dt$, $e_V(t)$, and $de_V(t)/dt$ are control error of $C(t)$, derivative control error of $C(t)$, control error of $V(t)$, and derivative control error of $V(t)$, respectively; $\mu_{NL}(e_C)$, $\mu_{NS}(e_C)$, and $\mu_{NP}(e_V)$ are membership degrees of FLC's input variables; $\mu_{ZO}(f_{in})$, $\mu_{SO}(f_{in})$, and $\mu_L(C_{in})$ are membership degrees of FLC's output variables.

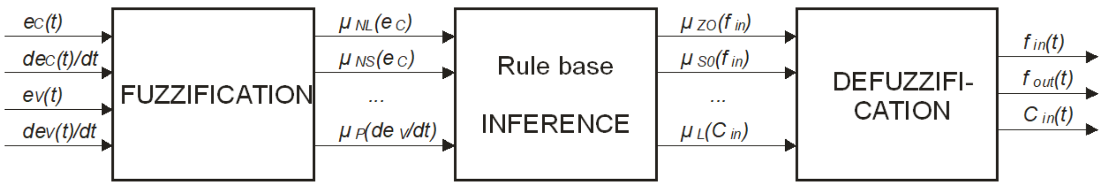


Fig. 5. Block diagram of fuzzy controller

The purpose of fuzzification is to determine a membership degree for individual fuzzy sets of input functions of the system. It is necessary to define the membership function for every input variable. Fuzzy inference comes down to calculating result membership functions of the conclusions (output variables of the system) based on the evaluation of the degree of membership for individual input variables. It is achieved based on a rule base, which is a set of relationships existing between input and output fuzzy sets of the system. The rules, which constitute the rule base, can take the following form:

$$R_1\text{: IF } (x_1 = A_1) \text{ AND } (x_2 = B_1) \text{ THEN } (y = C_1), \tag{102}$$

where x_1, x_2 are input variables of the system, y is output variable of the system, A_1, and B_1 are fuzzy sets of input variables, and C_1 is a fuzzy set of the output variable.

The purpose of defuzzification is to calculate fuzzy and sharp values based on the result membership functions of conclusions, according to a chosen criterion, e.g., the center of gravity (COG) method. The acquired output value of the system is a number, for instance, a specific voltage value on the output of FLC [17, 18].

In practical application of the fuzzy modeling, often, instead of the formal approach, a simplified approach is applied, for example, a simplified Mamdani inference system, which was applied in this research. It consists of the following stages and fuzzy sets operations [19]: intersection—t-norm MIN, union—t-norm MIN, implication—t-norm MIN, aggregation—s-norm MAX, defuzzification— COG.

The following variables were defined as input variables of the fuzzy controller: control error of $C(t)$ is $e_C(t)$, derivative control error of $C(t)$ is $de_C(t)/dt$, control error of $V(t)$ is $e_V(t)$, and derivative control error of $V(t)$ is $de_V(t)/dt$. Derivative control error enables the evaluation of rate and sign of control error change. The control variables $f_{in}(t)$, $f_{out}(t)$, and $C_{in}(t)$ were defined as output variables of the fuzzy controller. Gains K_{CP}, K_{CD}, K_{VP}, and K_{VD} were implemented in the tracks of FLC's input variables. Their purpose is to adjust the input signals to the range defined for individual inputs of the FLC. Chosen gain values were K_{CP} = 14, K_{CD} = 0.025, K_{VP} = 15, and K_{VD} = 0.1. Additionally, the dead zone algorithm was applied after the operation of derivation in order to improve the calculation efficiency. The control system and the fuzzy controller were developed gradually. At first, gains for FLC's inputs $e_V(t)$ and $de_V(t)/dt$ were found and a fuzzy inference system was developed for the process variable $V(t)$, without taking $C(t)$ into consideration. As soon as a satisfactory performance for $V(t)$ was achieved, the same steps were taken in regard to FLC's inputs $e_C(t)$ and $de_C(t)/dt$ and output $C_{in}(t)$. The control system's structure is shown in Fig. 6.

The linguistic values and membership function type of FLC's inputs are presented in Tab. 2. The range of all input variables is [−1, 1].

In Tab. 3, the linguistic values and membership function type of FLC's inputs are presented. The range of the fuzzy variables f_{in} and f_{out} is [0, 1] [m^3/s]. The range of the fuzzy variable C_{in} is [0, 25] [kmol/m^3] because such range was assumed in the process of fuzzy controller design as the achievable range of component concentration. The linguistic values and membership functions for controllers output variable f_{out} are identical to those for f_{in}.

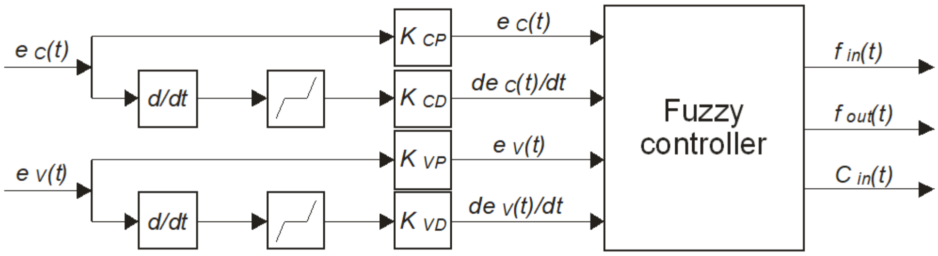


Fig. 6. FLC structure. FLC, fuzzy logic controller

Tab. 2. Linguistic values and membership functions for FLC inputs

Linguistic variable – error of C (eC)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	Negative	N	L-function	–	−0.2	0
2	About zero	ZO	Symmetrical triangular	−0.08	0	0.08
3	Positive	P	R-function	0	0.2	–
Linguistic variable – derivative of error of C (deC/dt)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	Negative	N	L-function	–	−0.2	0
2	About zero	ZO	Symmetrical triangular	−0.2	0	0.2
3	Positive	P	R-function	0	0.2	–
Linguistic variable – error of V (eV)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	Negative large	NL	L-function	–	−0.45	0
2	Negative small	NS	Symmetrical triangular	−0.1	−0.05	0
3	About zero	ZO	Symmetrical triangular	−0.01	0	0.01
4	Positive small	AP	Symmetrical triangular	0	0.05	0.1
5	Positive large	LP	R-function	0	0.45	–
Linguistic variable – derivative of error of V (deV/dt)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	Negative	N	L-function	–	−1	0
2	About zero	ZO	Symmetrical triangular	−0.02	0	0.02
3	Positive	P	R-function	0	1	–

FLC, fuzzy logic controller

Tab. 3. Linguistic values and membership functions for FLC outputs

Linguistic variable – inflow rate (fin)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	About zero	ZO	L-function	–	0.05	0.2
2	Small	S0	Symmetrical triangular	0.14	0.2	0.26
3	Medium	M	Asymmetrical triangular	0.2	0.35	0.6
4	Large medium	LM	Asymmetrical triangular	0.5	0.6	0.8
5	Large large	LL	R-function	0.6	0.8	–
Linguistic variable – concentration (Cin)						
	Linguistic value		**Membership function**	**Characteristic points**		
1	About zero	ZO	L-function	–	0	4.9
2	Small	S0	Symmetrical triangular	4.5	5	5.5
3	Large	L	R-function	5.1	10	–

FLC, fuzzy logic controller

The rule base was based on the tables of connections between FLC's inputs and outputs for the variable *V(t)* (see Tab. 4) and for the variable *C(t)* (see Tab. 5). The tables facilitated taking the connection between all pairs of input values and all output values into consideration while creating the rule base. The linguistic values in Tabs. 4 and 5 are presented in the form f_{in}/f_{out}.

In order to reduce the settling time of *C(t)*, especially in the case of the negative step of the reference value, a modification

DOI 10.2478/ama-2022-0018

was applied to the rule base. Rules which in case of steady volume *V(t)* increase both inflow and outflow of the substance were added. The aim of the modification was to accelerate achieving the reference value of *C(t)* through removing the substance of incorrect *C(t)* value while maintaining a constant value of *V(t)*. *V(t)* in steady state was a premise for these rules, because the increase of $f_{in}(t)$ and $f_{out}(t)$, while the value of *V(t)* was changing, resulted in an extension of the settling time of this variable. The rule base was extended by three additional rules:

R_{25}: IF (e_C = P) AND (e_V = ZO) AND (de_V/dt = ZO) THEN (f_{in} = LL)(f_{out} = LL) (103a)

R_{26}: IF (e_C = N) AND (e_V = ZO) AND (de_V/dt = ZO) THEN (f_{in} = LL)(f_{out} = LL) (103b)

R_{27}: IF (e_C = ZO) AND (e_V = ZO) AND (de_V/dt = ZO) THEN (f_{in} = M)(f_{out} = M) (103c)

Tab. 4. Rules of the fuzzy controller regarding *C(t)*

Inputs	Linguistic variable	deV/dt		
	Linguistic value	N	ZO	P
eV	NL	ZO/LL	ZO/LM	ZO/LM
	NS	S/LM	S/M	S/M
	ZO	S/S	S/S	S/S
	PS	M/S	M/S	LM/S
	PL	LM/ZO	LM/ZO	LL/ZO

Tab. 5. Rules of the fuzzy controller regarding *V(t)*

Inputs	Linguistic variable	deC/dt		
	Linguistic value	N	ZO	P
eC	N	ZO	ZO	S0
	ZO	S0	S0	S0
	P	S0	L	L

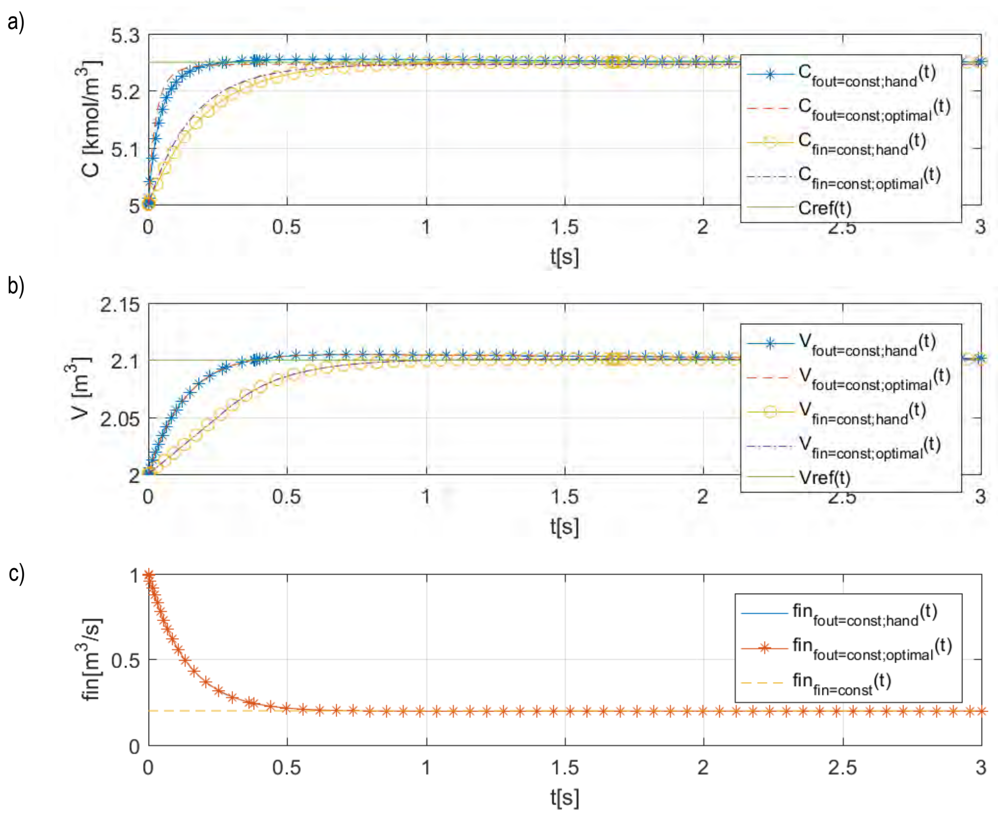
4. CONTROL RESULTS AND DISCUSSION

Best values of integral indices for the best controller type (PI) with optimal constants in both configurations of the system were presented in Tab. 6. For the $f_{out}(t)$ =const system, as well as the $f_{in}(t)$=const system, the lowest index values, and the most advantageous graphs were achieved for the IE criterion.

Tab. 6. Integral performance index values for control using PI controllers

Performance index	Control value	$f_{out}(t)$=const	$f_{in}(t)$=const
IAE	C	1.242E-02	2.258E-02
	V	1.304E-02	5.000E-02
ISE	C	1.170E-03	1.027E-03
	V	5.379E-04	1.670E-03
ITAE	C	1.997E-03	1.410E-03
	V	1.997E-03	2.505E-02
IE	C	2.799E-06	2.517E-06
	V	-2.561E-15	-2.465E-12

PI, proportional-integral



DOI 10.2478/ama-2022-0018 acta mechanica et automatica, vol.16 no.2 (2022)

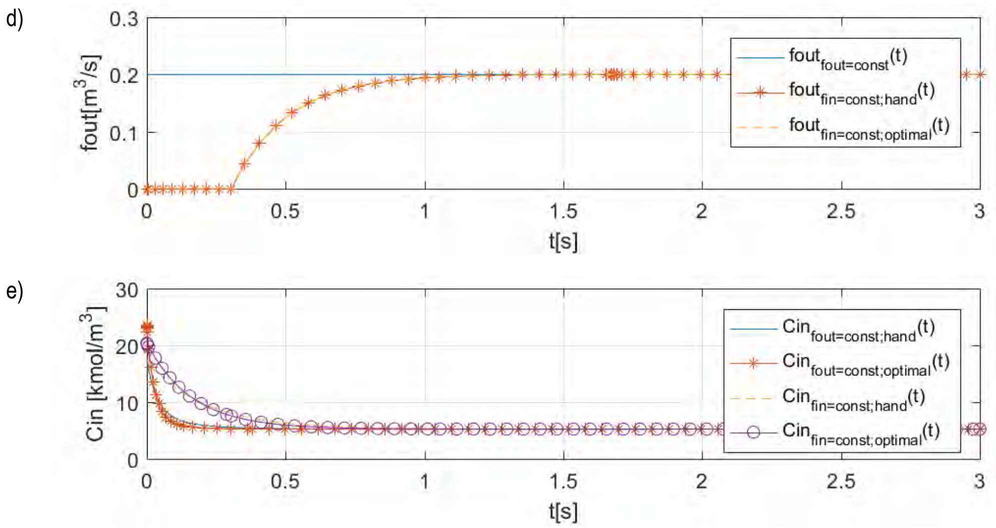


Fig. 7. Control results (multivariable PID) for $f_{out}(t)$=const system and for $f_{in}(t)$=const system

The control results of multivariable PID control with constants chosen by hand and by optimization were presented in Fig. 7. The optimal control with constant $f_{out}(t)$ ensured shorter settling and rise time and allowed to avoid control signals' wind-up (Figs. 7a, b). However, it also caused the process value $V(t)$ to overshoot (see Fig. 7b). Moreover, for the $f_{in}(t)$ = const system, the control signal $f_{out}(t)$ entered a long state of saturation (see Fig. 7d), which is an undesirable effect. For the $f_{out}(t)$ = const system, the control signal $f_{in}(t)$ came close to the end of its range, although it did not reach it (see Fig. 7c). The optimization enabled to achieve shorter settling and rise time of $C(t)$ (Fig. 7a).

The graphs of process and control values for the best values of ***Q*** and ***R*** matrices, which had been selected, tuned manually (98a), (99a) and achieved in the process of optimization (100a), (101a) for the $f_{out}(t)$=const system and for the $f_{in}(t)$=const system were compared. By far, the best results were achieved for the version with constant substance outflow—settling time of variables $C(t)$ (Fig. 8a) and $V(t)$ (Fig. 8b) was significantly shorter and the control signal $C_{in}(t)$ (Fig. 8e) is significantly less demanding compared with the $f_{in}(t)$=const configuration. The reason for this is that in the nonlinear model of the averaging tank, the variable $f_{in}(t)$ influences the rate of change of concentration $C(t)$ both directly and through the change of volume $V(t)$, while the variable $f_{out}(t)$ influences only the change of $V(t)$, which then influences $C(t)$. The advantage is prominent, especially in Fig. 8b, which presents settling time approx. 3 times shorter than in the case of $f_{in}(t)$=const. The optimization improved the control system's performance for both configurations. It is indicated by the values of integral performance indices as well (see Tab. 7). The indices are in all instances lower for the $f_{out}(t)$=const system than for the $f_{in}(t)$=const system and they are lower for optimal ***Q*** and ***R*** matrices compared to the results of manual tuning (with the exception of ISEG). The only reservation is that the maximum values of control variables—$f_{in}(t)$ in the $f_{out}(t)$=const system (Fig. 8c), $f_{out}(t)$ in the $f_{in}(t)$=const system (Fig. 8d), and $C_{in}(t)$ in the $f_{out}(t)$=const system (Fig. 8e) have increased visibly.

Tab. 7. Integral performance index values for $f_{out}(t)$=const system and for $f_{in}(t)$=const system (IC control system)

Performance index	Control value	$f_{out}(t)$=const; hand	$f_{out}(t)$=const; optimal	$f_{in}(t)$=const; hand	$f_{in}(t)$=const; optimal
IAE	C	0.398	0.019	0.415	0.022
	V	0.098	0.011	0.149	0.043
ISE	C	0.064	0.003	0.068	0.004
	V	0.006	0.001	0.010	0.003
ITAE	C	0.481	0.001	0.510	0.002
	V	0.067	0.001	0.161	0.014
IE	C	0.372	0.019	0.380	0.019
	V	0.097	0.011	0.140	0.043
ISEG	C	156.073	156.200	156.078	156.202
	V	24.969	24.981	24.972	24.968

IC, integral control

The control results for the fuzzy control system were shown in Fig. 9. The settling time and the rise time (Fig. 9a, b) were short and overshoot did not occur. The graph of C(t) (Fig. 9a), in particular, achieves the reference value faster than for other control system structures. The control values (Fig. 9c and d) did not reach their limit values. Moreover, the margin separating the control signals from saturating was sufficient, which was not true for PID and IC control. The only reservation in the case of fuzzy control is that the change rates of control signals were very high, which may not be possible to achieve for a real-life control plant. The values of integral performance indices for fuzzy control were presented in Tab. 8. The values of the indices were smaller than for IC, which confirms the superiority of fuzzy control in this instance.

DOI 10.2478/ama-2022-0018

Tab. 8. Integral performance index values (fuzzy control system)

Performance index	Control value	Index value
IAE	C	8.80E-02
	V	8.73E-03
ISE	C	1.42E-03
	V	5.75E-04
ITAE	C	2.22E-04
	V	5.15E-04
IE	C	8.82E-03
	V	8.73E-03

According to Figs. 7–9 and the values of performance indices in Tabs. 6–8, fuzzy control ensured the best performance of the system, with the most advantageous graphs of control variables. The settling time for both process variables was significantly shorter than for IC, which also allowed for good results. Multivariable PID control gave the worst results from among the designed algorithms. It is in line with expectations because a PID controller is dedicated to linear SISO systems and the considered system is nonlinear and MIMO. A fuzzy control system is a control method dedicated, among others, for nonlinear systems. It may be the cause of the advantage over IC, which in deviation from the adopted operating point did not give as good a result.

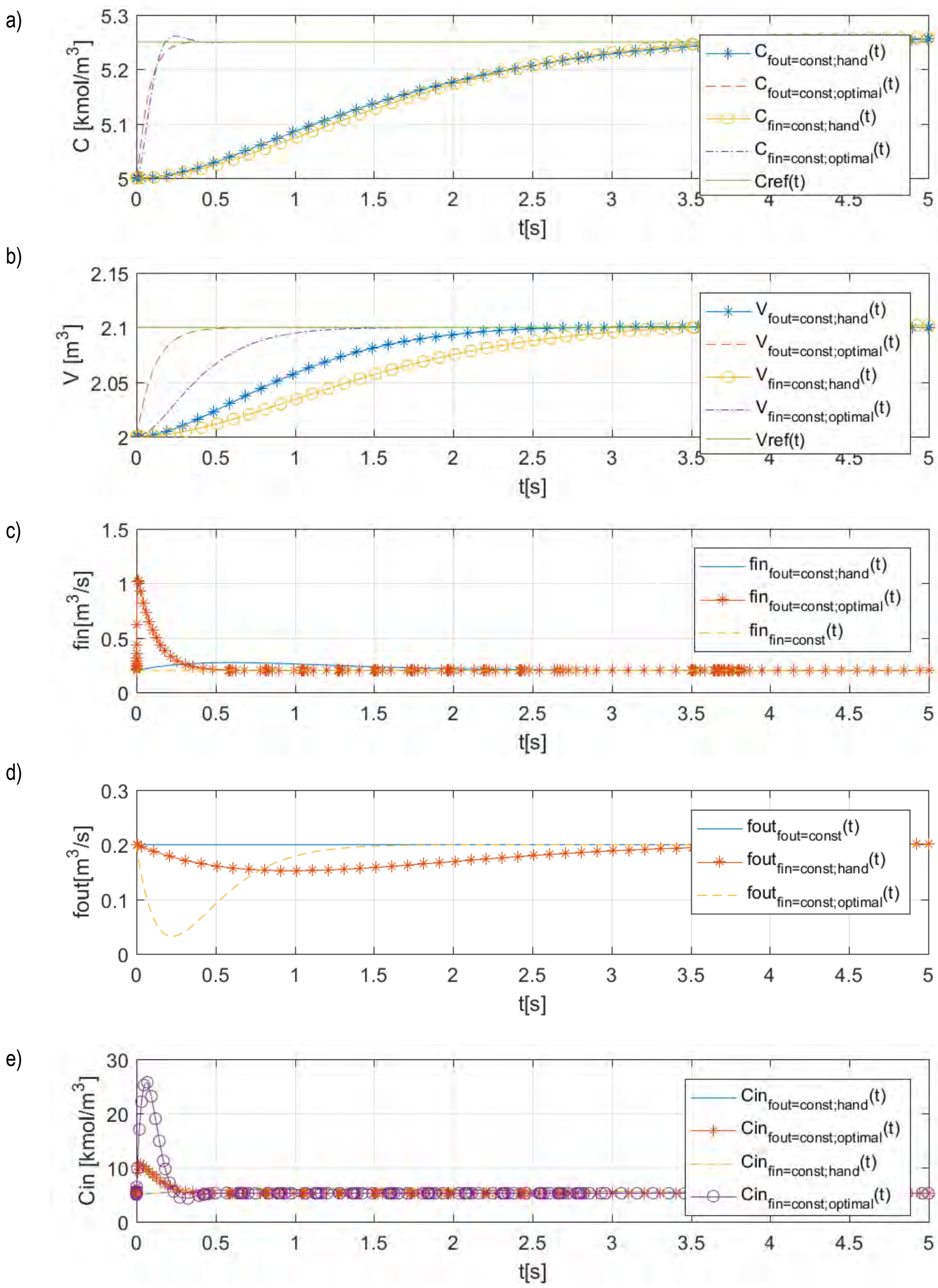


Fig. 8. Control results (IC control system) for $f_{out}(t)$=const system and for $f_{in}(t)$=const system

DOI 10.2478/ama-2022-0018 acta mechanica et automatica, vol.16 no.2 (2022)

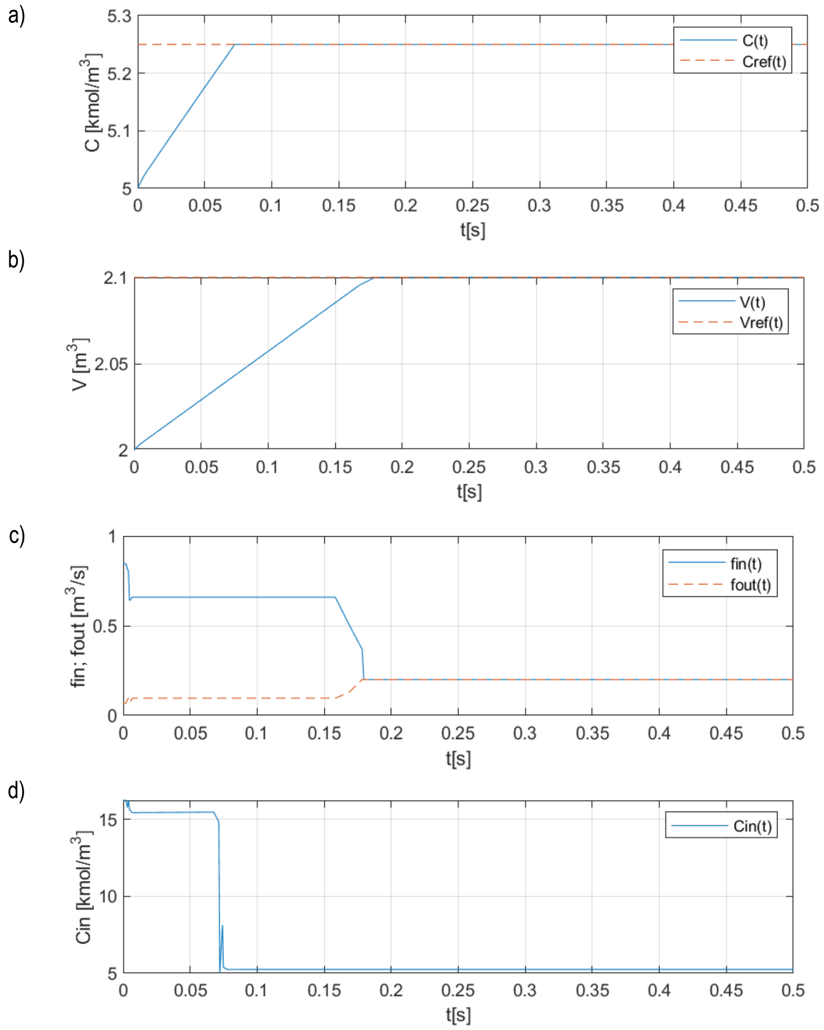


Fig. 9. Control results for fuzzy control system

5. CONCLUSIONS

This paper presented the modeling and control of an averaging tank with variable filling. The aim of the designed control system was to ensure zero steady-state error and quick response with minimal overshoot. Three control systems were designed. Multivariable PID control was developed using optimization for tuning the two PID controllers. The IC structure was designed using LQR for calculating gain matrices and optimization for calculating LQR weight matrices. Both multivariable PID and IC were developed in two versions: with $f_{out}(t)$=const and with $f_{in}(t)$=const. The control structure with FLC is using a Mamdani inference system. A series of simulation tests of the control system (in two versions of multivariable PID and IC) allowed to assess the performance of the system. Despite the fact that the control systems using PID and IC were designed based on the linearized model and tested on the nonlinear model of the control system, the result performance was good in the case of IC and decent for PID. However, due to the nonlinearity of the control system, as the reference values deviated significantly from the assumed equilibrium point, the systems' performance declined. Because fuzzy control is an adequate control method for MIMO and nonlinear control systems, it assured the best performance amongst the three designed algorithms.

The results presented in this paper may be used to synthesize a control system for a physical nonlinear MIMO system. In order to do that, a modification of the equilibrium point assumed in the stage of linearization would be necessary to ensure its consistency with the physical properties of a control system. Additionally, adequate actuators and transport delays should be considered in the model. The design methodology of control algorithms presented in the paper allows performing calculation necessary to acquire parameters of these algorithms for the modified control system model.

An average Programmable Logic Controller (PLC) would be sufficient hardware to implement multivariable PID control and IC. Although not crucial, it would be helpful if the programming environment of the PLC handled matrix operations. When it comes to FLC, it is also possible to implement it in a PLC, although the program may exceed the controller's capabilities, causing a necessity to extend the execution period of the code. However, it is recommended that FLC is implemented in a PLC dedicated to fuzzy control, in an industrial PC (IPC) or a microcomputer dedicated for industrial applications. Before implementing a physical control system, the control algorithms should be tested and adjusted in hardware in the loop (HIL) simulation.

The issue of decline in the performance of multivariable PID and IC, which is caused by deviation from the assumed equilibrium point, may be resolved by determining control parameters for multiple equilibrium points and using them in the gain scheduling method. Another method that would resolve this issue is the cyclic calculation of control parameters for a linearized control system

DOI 10.2478/ama-2022-0018

model whose parameters vary according to the current equilibrium point (nonstationary model).

REFERENCES

1. Spellman FR. Handbook of Water and Wastewater Treatment Plant Operations. CRC Press LLC, Boca Raton, FL. 2003.
2. Astrom KJ, Hagglund T. PID Controllers: Theory, Design and Tuning. 2nd ed. Instrument Society of America, Research Triangle, NC. 1995.
3. Rojas–Moreno A, Parra–Quispe A. Design and Implementation of a Water Tank Control System Employing a MIMO PID Controller. Faculty of Electrical and Electronic Engineering, National University of Engineering Lima. 2008.
4. Johansson KJ. The Quadruple-Tank Process: A Multivariable Laboratory Process with an Adjustable Zero. IEEE Transactions On Control Systems Technology 2000;8(3): 456-465.
5. Saeed Q, Uddin V, Katebi R. Multivariable Predictive PID Control for Quadruple Tank, World Academy of Science, Engineering and Technology. 2010;12: 861-866.
6. Meenatchi Sundaram S, Venkateswaran PR. Smith Predictor Implementation of a High Dead Time Interacting Tank Process. International Conference on Recent Innovations in Electrical, Electronics & Communication Engineering (ICRIEECE), 27-28 July, Bhubaneswar, India. 2018.
7. Janani S. Design of Integral Constant State Feedback Controller Using Ackermann's Function, IOSR Journal of Electronics and Communication Engineering (IOSR-JECE). 2014;9(1): 58-63.
8. Bojan-Dragos C, Hedrea E, Precup R, Szedlak-Stinean A, Roman R. MIMO Fuzzy Control Solutions for the Level Control of Vertical Two Tank Systems. Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019), SCITEPRESS, 29-31 July, Prague, Czech Republic. 2019: 810-817.
9. Berk P, Stajnko D, Vindis P, Mursec B, Lakota M. Synthesis water level control by fuzzy logic. Journal of Achievements in Materials and Manufacturing Engineering. 2011; 45(2): 204-210.
10. Kolankowski M, Piotrowski R. Synthesis of a state feedback controller for an averaging tank with variable filling. XXVI Scientific and Technical Conference Automation - News and Perspectives – AUTOMATION 2022, 25-27 May, Warsaw, Poland (in print). 2022.
11. Close CM, Frederick DK, Newell JC. Modeling and analisys of dynamic systems. 3rd ed. John Wiley & Sons, Hoboken, NJ. 2017.
12. Skogestad S, Postlethwaite I. Multivariable Feedback Control: Analysis and design. 2nd edition. John Wiley & Sons, Hoboken, NJ. 2005.
13. fminmax, https://se.mathworks.com/help/optim/ug/fminimax.html, (access: 18.10.2021).
14. Naidu DS. Optimal Control Systems. CRC Press, Boca Raton, FL. 2003.
15. Kwakernaak H, Sivan R. Linear Optimal Control Systems. John Wiley & Sons, Inc., Hoboken, NJ. 1972.
16. Murray RM. Optimization-Based Control. California Institute of Technology, Pasadena, CA. 2008.
17. Passino KM, Yurkovich S. Fuzzy Control. Addison Wesley Longman, Inc, Menlo Park, CA. 1998.
18. Jantzen J. Foundations of Fuzzy Control. John Wiley & Sons, Inc., Hoboken, NJ. 2007.
19. Mamdani EH. Application of fuzzy algorithms for control of simple dynamic plant. Proceedings of the Institution of Electrical Engineers. 1974;121(12): 1585–1588.

Michał Kolankowski: https://orcid.org/0000-0003-4528-3449

Robert Piotrowski: https://orcid.org/0000-0002-8660-300X

sciendo

DOI 10.2478/ama-2022-0019

acta mechanica et automatica, vol.16 no.2 (2022)
Special Issue *"Modeling of materials with nano/microstructure"*

ON THE NONLOCAL INTERACTION RANGE FOR STABILITY OF NANOBEAMS WITH NONLINEAR DISTRIBUTION OF MATERIAL PROPERTIES

Piotr JANKOWSKI*

*Faculty of Mechanical Engineering, Bialystok University of Technology, ul. Wiejska 45C, 15-351 Bialystok, Poland

p.jankowski@doktoranci.pb.edu.pl

received 17 January 2022, revised 10 February 2022, accepted 20 February 2022

Abstract: The present study analyses the range of nonlocal parameters' interaction on the buckling behaviour of nanobeam. The intelligent nonhomogeneous nanobeam is modelled as a symmetric functionally graded (FG) core with porosity cause nonlinear distribution of material parameters. The orthotropic face-sheets are made of piezoelectric materials. These kinds of structures are widely used in nanoelectromechanical systems (NEMS). The nanostructure model satisfies the assumptions of Reddy third-order beam theory and higher-order nonlocal elasticity and strain gradient theory. This approach allows to predict appropriate mechanical response of the nanobeam regardless of thin or thick structure, in addition to including nano-sized effects as hardening and softening. The analysis provided in the present study focuses on differences in results for nanobeam stability obtained based on classical and nonlocal theories. The study includes the effect of diverse size-dependent parameters, nanobeams' length-to-thickness ratio and distributions of porosity and material properties through the core thickness as well as external electro-mechanical loading. The results show a dependence of nonlocal interaction range on geometrical and material parameters of nanobeam. The investigation undertaken in the present study provides an interpretation for this phenomenon, and thus aids in increasing awareness of nanoscale structures' mechanical behaviour.

Key words: nanobeam, FGM, nonlocal strain gradient theory, buckling, piezoelectric effect

1. INTRODUCTION

Nanoelectromechanical systems (NEMS) generally refer to ultra-small structures that combine mechanical and electrical components in one device. These devices may find applications as force and displacement sensors [1], mass sensors [2], energy harvesters [3] electromechanical actuators [4]. Further, buckling behaviour plays a crucial role in the operation of these devices [5].

Small-scale effects should be considered in the prediction of the appropriate mechanical behaviour of micro- and nanostructures [6]. Therefore, nonlocal theories have modified classical continuum theories in three different ways. The strain gradient ones, which include the couple stress theory [7-9], the modified couple stress theory [10] and Mindlin's strain gradient theories [11,12], as well as the modified strain gradient theory [13], are based on assumption that the mechanical response of small-scaled structures depends on strains and gradients of strains. This approach is correlated with stiffness hardening effects. On the other hand, stress-gradient–based theories were established on the assumption that ultra-small–scale structures' mechanical behaviour depends on stress and gradients of stress. This approach is related to constitutive equations expressed by integral form [14-17] and differential form [18], as well as integrodifferential form [19]. These procedures capture mainly stiffness softening phenomena but hardening effect can be found in special cases. Nevertheless, in nanostructures' experiments, hardening-softening stiffness phenomena are observed [20,21]. A subsequent method that combines advantages of strain gradient and nonlocal elasticity theories was proposed as a higher-order nonlocal elasticity and strain gradient theory [22,23]. The aforementioned theory, known as the nonlocal strain gradient theory, employs nonlocal coefficient and length scale parameters to include both stiffness softening and hardening effects. Another advantage is that by reducing the related parameters, it becomes possible to easily obtain the required results through the strain gradient approach, nonlocal elasticity, or even classical continuum theory [6].

Increasing range of small-scaled devices' applications in various technology fields resulted in researchers' interest in modelling diverse structures' size-dependent mechanical response using various variational and nonlocal approaches. Among these, one can include ultra-small structures' dynamic response including torsional and longitudinal vibrations [24-26], vibrations under buckyballs [27], and considering thermal [28], magnetic [29], magneto-electro-elastic [30], and flexoelectric [31] effects Deflection analysis of micro/nano structures involves influence of diverse load distributions [32-35], magnetic [36] and magneto-electric [37] phenomena, hydrothermal environment [38,39], or novel nonlocal approaches [40] and nonlocal boundary conditions [41].

In order to maintain a subject matter similar to the study, the literature review is focused on a buckling response of nano-sized beam structures. Reddy [42] compared the influence of nonlocal parameter on bending, buckling and vibration for various beam theories based on analytical solution for simply supported nanostructures. Aydogdu [43] compared exponential beam theory with other models in analysing bending, buckling and vibration of Eringen's nanobeam. Roque et al. [44] proposed a meshless method to study bending, buckling and vibrations of nanobeam based on Eringen's nonlocal elasticity, Timoshenko beam model and two different collocation techniques. Thai [45] and Thai and

Vo [46] presented an analytical solution for free vibration, bending and buckling of nanobeam based on newly proposed refined higher-order shear deformation and sinusoidal shear deformation beam models using assumptions of Eringen's nonlocal elasticity. Ghannadpour et al. [47] studied the bending, buckling and vibration characteristics of nonlocal Euler-Bernoulli beam model for diverse boundary conditions (BCs) utilising the Ritz technique to solve defined problems. Şimşek and Yurtcu [48] compared bending and buckling of Euler-Bernoulli and Timoshenko functionally graded (FG) nanobeams based on Eringen's nonlocal theory for simply supported boundary conditions (BCs) using Navier solution procedure. Rahmani and Jandaghian [49] employed Eringen's nonlocal theory and Reddy displacement field to investigate buckling of FG nanobeam for various BCs using Rayleigh–Ritz method. Chaht [50] investigated the effect of Eringen's nonlocal parameter on bending and buckling of simply supported FG nanobeam using sinusoidal shear deformation theory with thickness stretching. Yu et al. [51] analysed the effect of heat conduction on buckling response of nanobeam for diverse BCs based on Eringen's nonlocal theory and Euler-Bernoulli beam model. Nejad et al. [52] conducted a study on buckling behaviour of Euler-Bernoulli two-dimensional FG nanobeam for diverse BCs based on nonlocal elasticity and generalised differential quadrature method (GDQM). Li et al. [53] employed nonlocal strain gradient theory and Euler-Bernoulli beam model to study bending, buckling and vibration of axially FG nanobeam by using GDQM approach. Tuna and Kirca [54] proposed finite element model formulation to investigate bending, buckling and free vibration problem of Eringen's nanobeam with various BCs. Mirjavadi et al. [55] analysed buckling and vibration of FG Euler-Bernoulli nanobeam including von-Karman strains and various BCs, using nonlocal elasticity and GDQM. Khaniki et al. [56] studied buckling in nonuniform Euler-Bernoulli nanobeam in the context of nonlocal strain gradient theory and numerical meshless approach. Alibeigi et al. [57,58] studied the nonlinear bending and buckling behaviours of piezoelectric and piezomagnetic nanobeam in thermal, electrical and mechanical environments in the context of modified couple stress theory and Galerkin approach. Xiao et al. [59] discussed nonlinear thermal buckling of FG porous nanobeam, including magneto-electro-elastic coupling effects using Eringen's nonlocal theory and perturbation solution method. Hashemian et al. [60] utilised Navier solution technique and nonlocal strain gradient theory to compare the influence of small-scale effect on bending and buckling response on nanobeams modelled by various beam theories. Jankowski et al. [61] conducted a study on buckling and vibration of FG porous nanobeam based on nonlocal strain gradient theory together with Reddy third-order and Timoshenko beam theories. Civalek et al. [62,63] proposed finite element formulation to analyse buckling of nanobeam-based structures for different BCs using Euler-Bernoulli assumptions and Eringen's nonlocal elasticity as well as modified couple stress theory.

1.1. Novelty of the paper

Given the constant development of nanoscience and nanoengineering, modelling of ultra-small structures deserves to be recognised as an important issue that scientists in the current era need to deal with. Nevertheless, based on the provided literature survey, it can be gauged that the range of nonlocal interaction for nanostructures has not been investigated yet. There is still a lack of adequate knowledge of nonlocal and length scale parameters' impact on nanostructures' mechanical behaviour, considering the effect of its geometrical as well as material properties or external loads. The present study supplements the research gap by providing a size-dependent buckling analysis of three-layered functionally graded material (FGM) piezoelectric nanobeam. Utilized Reddy third-order shear deformation theory may to an analysis of thick as well as thin structures. Additionally, the beam model does not require a shear correction factor, which should be determined separately for different material parameters' variations, geometric ratios and loadings. Therefore, the theory employed in the present study ensures high accuracy and makes it possible to omit the drawbacks of other theories. In comparison to previous studies [61,64] the present investigation ensures another perspective to size-dependent buckling response of nanobeam. The paper is focused on a range of nonlocal effects on nanostructure mechanical response in view of material and geometrical properties as well as electromechanical loads. The current investigation presents differences, regarding classical continuum theory, of mechanical response of nanostructures taking into account diverse relations of small-scale coefficients together with length to thickness ratios, influence of mechanical forces and electric field, properties of FG material gradation along with distribution and volume of porosity. The present study will widen the understanding of the size-dependent responses of nanostructures and increase awareness about the necessity to use, or alternatively the unfavourable consequences of omitting, nonlocal theories in the modelling of nanoscale smart structures, which is a key value in optimisation and control of NEMS devices.

2. CONSTITUTIVE RELATIONS

One of the favoured theories including size-dependent phenomena is nonlocal elasticity proposed by [15,16,18,65], which postulates that stress at a point in continuum depends on the strain at that point as well at points in the whole domain. On the other hand, theories employing strain-gradient–based assumptions [7-9,11,12] state that materials should not be treated as a collection of points, but rather as atoms with higher-order deformation mechanism at micro- and nanoscales. Askes and Aifantis [22] and Lim et al. [23] found it necessary to bring together two entirely different physical phenomena into a single theory to describe a more real structural response at the nanoscale. The higher-order nonlocal elasticity and strain gradient theory combines nonlocal assumptions of strain gradient and stress gradient. The theory captures both nonlocal phenomena and consequently enables a more effective prediction of the size-dependent mechanical response. Based on the refined theory assumptions, the total stress tensor of the nonlocal strain gradient theory is defined as

$$\boldsymbol{\sigma} = \overline{\boldsymbol{\sigma}} - \nabla \overline{\boldsymbol{\sigma}}^{(1)} \tag{1}$$

where $\nabla = \boldsymbol{e}_i \frac{d}{dx_i}$ is the vector differential operator, in which $\boldsymbol{e}_i$ is the unit vector and x_i is considered the direction of a nonlocal effect in a structure. Then, $\overline{\boldsymbol{\sigma}}$ and $\overline{\boldsymbol{\sigma}}^{(1)}$ are classical and higher-order stress tensor, respectively. The three-dimensional components are

$$\overline{\boldsymbol{\sigma}} = \int_V \alpha_0(\boldsymbol{x}', \boldsymbol{x}, e_0 a)\, \boldsymbol{C} : \boldsymbol{\varepsilon}' dV' \tag{2a}$$

sciendo

$$\bar{\boldsymbol{\sigma}}^{(1)} = \ell^2 \int_V \alpha_1(\boldsymbol{x}', \boldsymbol{x}, e_1 a)\boldsymbol{C}: \nabla\boldsymbol{\varepsilon}' dV' \tag{2b}$$

where $\alpha_0(\boldsymbol{x}', \boldsymbol{x}, e_0 a)$ is the principal attenuation kernel function related to the nonlocality in terms of Euclidean distance between the point $\boldsymbol{x}$ and points $\boldsymbol{x}'$ in the domain V, and $\alpha_1(\boldsymbol{x}', \boldsymbol{x}, e_1 a)$ is the additional attenuation kernel function introduced to describe the nonlocal effect of first-order strain gradient field. Subsequently, ℓ is the material length scale coefficient that describes the higher-order strain gradient field, a is internal characteristic length (e.g., interatomic distance) and e_0 and e_1 are the nonlocal material constants. $\boldsymbol{\varepsilon}$, $\boldsymbol{\varepsilon}'$ and $\boldsymbol{C}$ represent the strain tensor, the strain gradient tensor and the fourth-order elastic modulus tensor, respectively. Applying the linear differential operators in form

$$\boldsymbol{L}_i = 1 - (e_i a)^2\nabla^2, i = 0{,}1 \tag{3}$$

on both sides of Eq. (1) leads to a general differential form of the constitutive equations based on the higher-order nonlocal strain gradient theory, including three length scale parameters: two of them for nonlocal stress and the third for strain gradients

$$[1-(e_1 a)^2\nabla^2][1-(e_0 a)^2\nabla^2]\boldsymbol{\sigma} = [1-(e_1 a)^2\nabla^2]\boldsymbol{C}: \boldsymbol{\varepsilon} - \ell^2[1-(e_0 a)^2\nabla^2]\nabla\boldsymbol{C}: \nabla\boldsymbol{\varepsilon} \tag{4}$$

Assuming that $e_1 = e_0 = e$, one can obtain a simplified model in the form

$$[1-(ea)^2\nabla^2]\boldsymbol{\sigma} = [1-\nabla^2\ell^2]\boldsymbol{C}: \boldsymbol{\varepsilon} \tag{5}$$

The presented model may be easily reduced to stress-based Eringen's nonlocal elasticity taking $\ell = 0$:

$$[1-(ea)^2\nabla^2]\boldsymbol{\sigma} = \boldsymbol{C}: \boldsymbol{\varepsilon} \tag{6}$$

Alternatively, supposing $ea = 0$ indicates the form of Mindlin's strain gradient theory

$$\boldsymbol{\sigma} = [1-\nabla^2\ell^2]\boldsymbol{C}: \boldsymbol{\varepsilon} \tag{7}$$

Considering the above, constitutive relations for present investigation are

$$(1-\mathfrak{B}\nabla^2)\sigma_{xx} = (1-\ell^2\nabla^2)C_{xx}\varepsilon_{xx} \tag{8a}$$

$$(1-\mathfrak{B}\nabla^2)\sigma_{xz} = (1-\ell^2\nabla^2)2C_{xz}\varepsilon_{xz} \tag{8b}$$

where $\mathfrak{B} = (ea)^2$ and differential operator takes the unidirectional form $\nabla^2 = \frac{\partial^2}{\partial x^2}$.

3. DISPLACEMENT FIELD AND MATERIAL PROPERTIES

Consider a symmetric FG nanobeam layered with piezoelectric face-sheets characterised by length L, width b and total thickness $H = h + 2h_p$ that consists of a core made of porous FGM with thickness h, and two piezoelectric layers thicknesses h_p. An ideal mechanical contact is assumed between the FGM core and other layers, which is more justified for structures at nanoscales, owing to possible strong bonds, e.g., chemical bonds. The nanobeam is under electric field induced by external initial electric voltage ϕ_0 and under mechanical load presented as axial in-plane forces $\widehat{N}_{xx}$. The nanostructure coordinate system and its cross-section are illustrated in Fig. 1.

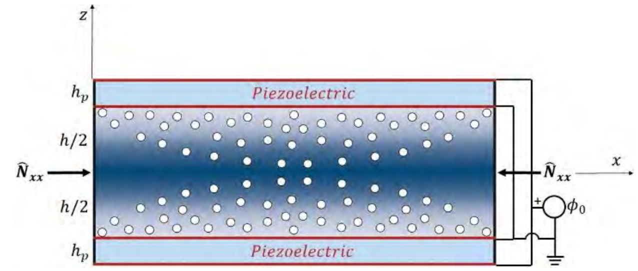


Fig. 1. Model of three-layered porous FGM nanobeam subjected to external mechanical loads and electric field

3.1. Displacement field

Based on assumptions of Reddy third-order shear deformation beam theory [66], the displacement field takes the form

$$u_x(x,z,t) = u_0(x,t) + z\varphi_x(x,t) - \mathrm{c}_1 z^3\left(\varphi_x(x,t) + \frac{\partial w_0(x,t)}{\partial x}\right) \tag{9a}$$

$$u_z(x,t) = w_0(x,t) \tag{9b}$$

where u_x and u_z are the total displacement vector components, and particular elements u_0, w_0 are displacements in axial and transverse directions of a material point in the midplane in the undeformed configuration at any time t. φ_x represents the rotation of the point on the centroidal axis x of the beam, and $c_1 = 4/(3H^2)$.

Assuming linear and infinitesimal Green–Lagrange strain tensor $\boldsymbol{\varepsilon}$, strain–displacement relations are presented as

$$\varepsilon_{xx} = \varepsilon_{xx}^{(0)} + z\varepsilon_{xx}^{(1)} + z^3\varepsilon_{xx}^{(3)} \tag{10a}$$

$$2\varepsilon_{xz} = \gamma_{xz}^{(0)} + z^2\gamma_{xz}^{(2)} \tag{10b}$$

with particular components

$$\left\{\varepsilon_{xx}^{(0)}, \varepsilon_{xx}^{(1)}, \varepsilon_{xx}^{(3)}\right\} = \left\{\frac{\partial u_0}{\partial x}, \frac{\partial \varphi_x}{\partial x}, -\mathrm{c}_1\left(\frac{\partial \varphi_x}{\partial x} + \frac{\partial^2 w_0}{\partial x^2}\right)\right\} \tag{11a}$$

$$\left\{\gamma_{xz}^{(0)}, \gamma_{xz}^{(2)}\right\} = \left\{\varphi_x + \frac{\partial w_0}{\partial x}, -\mathrm{c}_2\left(\varphi_x + \frac{\partial w_0}{\partial x}\right)\right\} \tag{11b}$$

where $\mathrm{c}_2 = 3\mathrm{c}_1$.

3.2. Material properties for FGM core

The presented nanobeam consists of orthotropic piezoelectric face-sheets and FG porous core with a presumption of perfect mechanical connection between each layer. The stiffness coefficients of the porous FGM nanobeam core are described as:

$$C_{xx} = E(z,\Upsilon) \tag{12a}$$

$$C_{xz} = \frac{E(z,\Upsilon)}{2(1+\nu)} \tag{12b}$$

where ν is Poisson's ratio that is assumed to be constant due to low volatility of the coefficient, and consequently, with minor influence on mechanical response. $E(z,\Upsilon)$ is Young's modulus that varies through the nanobeam core thickness h for different porosity distribution Υ in the volume. Material properties of symmetric (with respect to the midplane $z = 0$) FG structure with porosity are described based on modified Voigt's rule of mixtures as follows [64]:

DOI 10.2478/ama-2022-0019

$$E^{(n)}(z) = \left[(E_c - E_m)V^{(n)}(z,g) + E_m\right]\left[1 - \Upsilon(z,\vartheta)\right] \quad (13)$$

where E_c and E_m refer to Young's modulus of ceramic and metal in the upper and lower surfaces that consist of the mixture inside the structure. The volume fraction function $V^{(n)}$ of the nth layer is defined as:

$$V^{(1)} = \left(\frac{z+\frac{h}{2}}{\frac{h}{2}}\right)^{g} \ \wedge \ z \in \langle -\tfrac{h}{2}, 0\rangle \quad (14a)$$

$$V^{(2)} = \left(\frac{z-\frac{h}{2}}{-\frac{h}{2}}\right)^{g} \ \wedge \ z \in \langle 0, \tfrac{h}{2}\rangle \quad (14b)$$

The power-law index g controls the share of constituents in the resultant structure. Diverse porosity distributions are modelled using function $\Upsilon(z,\vartheta)$. In the present study, three diverse porosity distributions [64] are examined, as under:

$$\text{Type 1: } \Upsilon(z,\vartheta) = \vartheta cos\left(\frac{\pi z}{h}\right) \ \wedge \ z \in \langle -\tfrac{h}{2}, \tfrac{h}{2}\rangle \quad (15a)$$

$$\text{Type 2: } \Upsilon(z,\vartheta) = \begin{cases} \vartheta cos\left[\pi\left(\frac{2z}{h}+\frac{1}{2}\right)\right] \ \wedge \ z \in \langle -\frac{h}{2}, 0\rangle \\ \vartheta cos\left[\pi\left(\frac{2z}{h}-\frac{1}{2}\right)\right] \ \wedge \ z \in \langle 0, \frac{h}{2}\rangle \end{cases} \quad (15b)$$

$$\text{Type 3: } \Upsilon(z,\vartheta) = \begin{cases} \vartheta cos\left[\pi\left(\frac{z}{h}+\frac{1}{2}\right)\right] \ \wedge \ z \in \langle -\frac{h}{2}, 0\rangle \\ \vartheta cos\left[\pi\left(\frac{z}{h}-\frac{1}{2}\right)\right] \ \wedge \ z \in \langle 0, \frac{h}{2}\rangle \end{cases} \quad (15c)$$

where ϑ is porosity coefficient that determines maximum volume of voids. Material properties, as well as distributions of porosity through the nanobeam core thickness, are presented in Fig. 2.

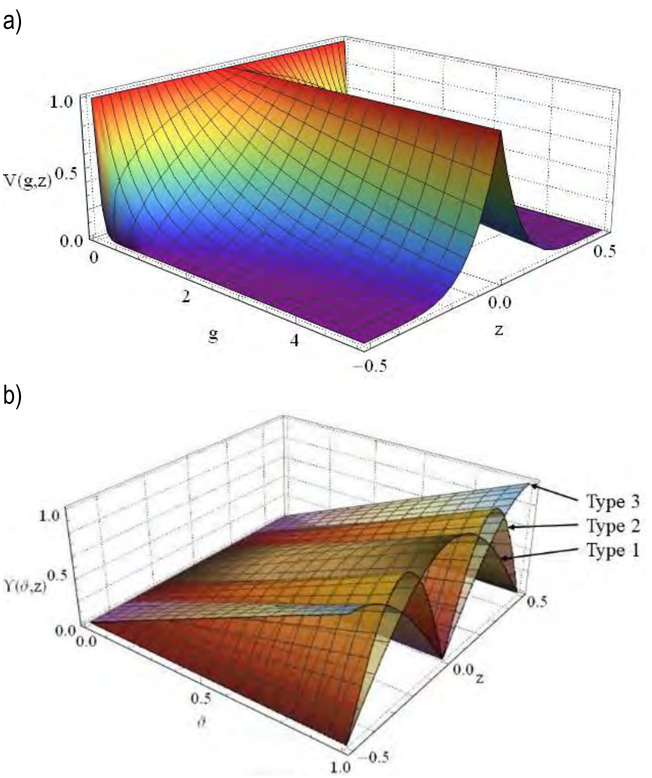


Fig. 2. Properties of nanobeams FGM core: (a) volume fraction function; (b) distribution of porosity

3.3. Piezoelectric characteristics

The electric field is described based on a combination of the half-cosine and linear variations of the electric potential $\breve{\Phi}$ [67]:

$$\breve{\Phi}(x,z,t) = -cos\left(\frac{\pi z_p}{h_p}\right)\Phi(x,t) + \frac{2z_p}{h_p}\phi_0 \quad (16)$$

where $\Phi(x,t)$ refers to plane time-dependent distribution of electric potential and ϕ_0 is an initial external voltage. A new variable z_p is measured from the geometrical centre of the piezoelectric layers, and thus $z_1 = z - h/2 - h_p/2$ and $z_2 = z + h/2 + h_p/2$ are defined for the top and bottom layers, respectively. Based on the electric potential function, the components of electric field E_i are obtained, using a derivative with respect to the appropriate coordinate, as the following:

$$E_x = -\frac{\partial\breve{\Phi}}{\partial x} = cos\left(\frac{\pi z_p}{h_p}\right)\frac{\partial\Phi}{\partial x} \quad (17a)$$

$$E_z = -\frac{\partial\breve{\Phi}}{\partial z} = -\frac{\pi}{h_p}sin\left(\frac{\pi z_p}{h_p}\right)\Phi - \frac{2}{h_p}\phi_0 \quad (17b)$$

According to the nonlocal strain gradient theory, as well as obtained material properties together with electric field contribution, the final form of constitutive relations [64] is presented as follows:

$$(1-\mathfrak{B}\nabla^2)\sigma_{xx}^{p} = (1-\ell^2\nabla^2)(D_{xx}\varepsilon_{xx} - e_x E_z) \quad (18a)$$

$$(1-\mathfrak{B}\nabla^2)\sigma_{xz}^{p} = (1-\ell^2\nabla^2)(2D_{xz}\varepsilon_{xz} - e_z E_x) \quad (18b)$$

$$(1-\mathfrak{B}\nabla^2)D_{x}^{p} = (1-\ell^2\nabla^2)(2e_z\varepsilon_{xz} + \epsilon_x E_x) \quad (18c)$$

$$(1-\mathfrak{B}\nabla^2)D_{z}^{p} = (1-\ell^2\nabla^2)(e_x\varepsilon_{xx} + \epsilon_z E_z) \quad (18d)$$

$$(1-\mathfrak{B}\nabla^2)\sigma_{xx}^{c} = (1-\ell^2\nabla^2)C_{xx}\varepsilon_{xx} \quad (19a)$$

$$(1-\mathfrak{B}\nabla^2)\sigma_{xz}^{c} = (1-\ell^2\nabla^2)2C_{xz}\varepsilon_{xz} \quad (19b)$$

where superscripts p and c refer to the piezoelectric layers and FGM core, respectively, and D_i^p stands for electric displacement components. Additionally, D_{ij} represent the piezoelectric layers' stiffness coefficient and e_i and ϵ_i are, respectively, piezoelectric and dielectric permittivity constants.

4. EQUATIONS OF MOTION AND SOLUTION

4.1. Equations of motion

The equations of motion of the studied nanobeam are derived based on the modified Hamilton's variational principle, in the form

$$\int_0^T(\delta\mathcal{U} - \delta\mathcal{E} + \delta\mathcal{V})dt = 0 \quad (20)$$

where $\delta\mathcal{U}$, $\delta\mathcal{E}$ and $\delta\mathcal{V}$ stands for variations of virtual strain energy, virtual contribution of electric field and virtual work done by external forces, respectively. The procedure that leads to deriving equations of motion expressed by generalised displacements, including nonlocal interaction, is comprehensively described in a previous paper [64]. In the current study, we used previously derived equations of motion, which can be expressed as the following:

$$A_{xx}^{(0)}\frac{\partial^2 u_0}{\partial x^2}+A_{xx}^{(1)}\frac{\partial^2\varphi_x}{\partial x^2}-\mathrm{c}_1A_{xx}^{(3)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)+B_x^{(0)}\frac{\partial\Phi}{\partial x}-\ell^2\left[A_{xx}^{(0)}\frac{\partial^4 u_0}{\partial x^4}+A_{xx}^{(1)}\frac{\partial^4\varphi_x}{\partial x^4}-\mathrm{c}_1A_{xx}^{(3)}\left(\frac{\partial^4\varphi_x}{\partial x^2}+\frac{\partial^5 w_0}{\partial x^3}\right)+B_x^{(0)}\frac{\partial^3\Phi}{\partial x^3}\right]=0 \quad (21a)$$

$$-A_{xz}^{(0)}\left(\varphi_x+\frac{\partial w_0}{\partial x}\right)+2\mathrm{c}_2A_{xz}^{(2)}\left(\varphi_x+\frac{\partial w_0}{\partial x}\right)-\mathrm{c}_2^2A_{xz}^{(4)}\left(\varphi_x+\frac{\partial w_0}{\partial x}\right)+A_{xx}^{(1)}\frac{\partial^2 u_0}{\partial x^2}-\mathrm{c}_1A_{xx}^{(3)}\frac{\partial^2 u_0}{\partial x^2}+A_{xx}^{(2)}\frac{\partial^2\varphi_x}{\partial x^2}-\mathrm{c}_1A_{xx}^{(4)}\frac{\partial^2\varphi_x}{\partial x^2}-\mathrm{c}_1A_{xx}^{(4)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)+\mathrm{c}_1^2A_{xx}^{(6)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)+B_z^{(0)}\frac{\partial\Phi}{\partial x}+B_x^{(1)}\frac{\partial\Phi}{\partial x}-c_1B_x^{(3)}\frac{\partial\Phi}{\partial x}-c_2B_z^{(2)}\frac{\partial\Phi}{\partial x}-\ell^2\left[-A_{xz}^{(0)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)+2\mathrm{c}_2A_{xz}^{(2)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)-\mathrm{c}_2^2A_{xz}^{(4)}\left(\frac{\partial^2\varphi_x}{\partial x^2}+\frac{\partial^3 w_0}{\partial x^3}\right)+A_{xx}^{(1)}\frac{\partial^4 u_0}{\partial x^4}-\mathrm{c}_1A_{xx}^{(3)}\frac{\partial^4 u_0}{\partial x^4}+A_{xx}^{(2)}\frac{\partial^4\varphi_x}{\partial x^4}-\mathrm{c}_1A_{xx}^{(4)}\frac{\partial^4\varphi_x}{\partial x^4}-\mathrm{c}_1A_{xx}^{(4)}\left(\frac{\partial^4\varphi_x}{\partial x^4}+\frac{\partial^5 w_0}{\partial x^5}\right)+\mathrm{c}_1^2A_{xx}^{(6)}\left(\frac{\partial^4\varphi_x}{\partial x^4}+\frac{\partial^5 w_0}{\partial x^5}\right)+B_x^{(1)}\frac{\partial^3\Phi}{\partial x^3}+B_z^{(0)}\frac{\partial^3\Phi}{\partial x^3}-c_1B_x^{(3)}\frac{\partial^3\Phi}{\partial x^3}-c_2B_z^{(2)}\frac{\partial^3\Phi}{\partial x^3}\right]=0 \quad (21b)$$

$$c_1A_{xx}^{(3)}\frac{\partial^3 u_0}{\partial x^3}+A_{xz}^{(0)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)-2c_2A_{xz}^{(2)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)+c_2^2A_{xz}^{(4)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)+c_1A_{xx}^{(4)}\frac{\partial^3\varphi_x}{\partial x^3}-c_1^2A_{xx}^{(6)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)-B_z^{(0)}\frac{\partial^2\Phi}{\partial x^2}+c_2B_z^{(2)}\frac{\partial^2\Phi}{\partial x^2}+c_1B_x^{(3)}\frac{\partial^2\Phi}{\partial x^2}-\ell^2\left[c_1A_{xx}^{(3)}\frac{\partial^5 u_0}{\partial x^5}+A_{xz}^{(0)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)-2c_2A_{xz}^{(2)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)+c_2^2A_{xz}^{(4)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)+c_1A_{xx}^{(4)}\frac{\partial^5\varphi_x}{\partial x^5}-c_1^2A_{xx}^{(6)}\left(\frac{\partial^5\varphi_x}{\partial x^5}+\frac{\partial^6 w_0}{\partial x^6}\right)-B_z^{(0)}\frac{\partial^4\Phi}{\partial x^4}+c_1B_x^{(3)}\frac{\partial^4\Phi}{\partial x^4}+c_2B_z^{(2)}\frac{\partial^4\Phi}{\partial x^4}\right]=\widehat{N}_{\mathcal{E}}\frac{\partial^2 w_0}{\partial x^2}+\widehat{N}_{xx}\frac{\partial^2 w_0}{\partial x^2}-\mathfrak{B}\left[\widehat{N}_{\mathcal{E}}\frac{\partial^4 w_0}{\partial x^4}+\widehat{N}_{xx}\frac{\partial^4 w_0}{\partial x^4}\right] \quad (21c)$$

$$B_x^{(0)}\frac{\partial u_0}{\partial x}+B_x^{(1)}\frac{\partial\varphi_x}{\partial x}+B_z^{(0)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)-c_1B_x^{(3)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)-c_2B_z^{(2)}\left(\frac{\partial\varphi_x}{\partial x}+\frac{\partial^2 w_0}{\partial x^2}\right)-\mathcal{C}_z\Phi+\mathcal{C}_x\frac{\partial^2\Phi}{\partial x^2}-\mathcal{C}_{z\phi}-\ell^2\left[B_x^{(0)}\frac{\partial^3 u_0}{\partial x^3}+B_x^{(1)}\frac{\partial^3\varphi_x}{\partial x^3}+B_z^{(0)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)-c_1B_x^{(3)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)-c_2B_z^{(2)}\left(\frac{\partial^3\varphi_x}{\partial x^3}+\frac{\partial^4 w_0}{\partial x^4}\right)-\mathcal{C}_z\frac{\partial^2\Phi}{\partial x^2}+\mathcal{C}_x\frac{\partial^4\Phi}{\partial x^4}\right]=0 \quad (21d)$$

where $A_{ij}^{(m)}$, $B_i^{(m)}$, $\mathcal{C}_i$ and $\mathcal{C}_{i\Phi}$ refer to resultant stiffness, piezoelectric and dielectric coefficients, and are further described in the Appendix.

4.2. Solution procedure and verification

Navier solution technique is employed for a simply supported nanobeam under axial loading in the following manner:

$$\begin{pmatrix}u_0\\ \varphi_x\\ w_0\\ \Phi\end{pmatrix}=\sum_{n=1}^{\infty}\begin{pmatrix}\bar{u}_0\cos(\beta_n x)\\ \bar{\varphi}_x\cos(\beta_n x)\\ \bar{w}_0\sin(\beta_n x)\\ \bar{\Phi}\sin(\beta_n x)\end{pmatrix}\wedge\ \beta_n=\frac{n\pi}{L} \quad (22)$$

where $\bar{u}_0$, $\bar{\varphi}_x$, $\bar{w}_0$, $\bar{\Phi}$ express maximum amplitudes of displacements and electric potential.
The general governing equations are presented in matrix form as

$$[K][\bar{u}_0\quad \bar{\varphi}_x\quad \bar{w}_0\quad \bar{\Phi}]^T=0 \quad (23)$$

where $[K]$ is symmetric 4 x 4 stiffness matrix including in-plane electromechanical forces induced by external loads. Elements of stiffness matrix are presented in [64].

The presented solution scheme, as well as the obtained numerical results, were extensively compared with results from the literature, and verification may be found in [61,64]. Selected numerical results are presented in Tables 1 and 2, showing excellent accordance with results from the literature.

Tab. 1. Dimensionless buckling load $\bar{N}_{cr}=\widehat{N}_{xx}\frac{L^2}{EI}$ of simply supported homogeneous nanobeam without piezoelectric layers, assumed properties: $L=10\ \mathrm{nm},\ E=30\ \mathrm{MPa},\ \upsilon=0.3$, $I=\int_{-h/2}^{h/2}z^2dz$

L/H	$\mathfrak{B}$	Beam model assumptions				Present
		Euler-Bernoulli	Timoshenko	Reddy	Thai	
		[45]				
5	0	9.8696	8.9509	8.9519	8.9519	8.9519
	2	8.2426	7.4753	7.4761	7.4761	7.4761
	4	7.0761	6.4174	6.4181	6.4181	6.4181
20	0	9.8696	9.8067	9.8067	9.8067	9.8067
	2	8.2426	8.1900	8.1900	8.1900	8.1901
	4	7.0761	7.0310	7.0310	7.0310	7.0310

Tab. 2. Dimensionless buckling load $\bar{N}_{cr}=\widehat{N}_{xx}\frac{12L^2}{E_mh^3}$ of simply supported FG beam without piezoelectric layers, assumed properties: $L=10\ \mathrm{m}, E_c=380\ \mathrm{GPa}, E_m=70\ \mathrm{GPa}, \upsilon=0.3$

L/H	g	[68]	[69]	Present
5	0	49.5970	48.5959	48.5959
	1	20.0899	19.6525	19.6525
	5	10.3708	10.1460	10.1460
20	0	53.3175	53.2364	53.2364
	1	20.7541	20.7212	20.7212
	5	10.6341	10.6171	10.6171

5. PARAMETRIC STUDY

The present section contains novel numerical examples of a differences rate that occurs when neither Eringen's nonlocal elasticity nor the strain gradient theory is used in stability analysis of nanostructures. The nonlocal interaction range coefficient is defined as:

$$\delta=\left|\frac{\bar{N}_{cr}^{l}-\bar{N}_{cr}^{nl}}{\bar{N}_{cr}^{l}}\right|*100\% \quad (24)$$

where $\bar{N}_{cr}^{l}$ and $\bar{N}_{cr}^{nl}$ stand for dimensionless critical buckling load obtained based on the classical (local) theory of elasticity and the nonlocal strain gradient theory. The dimensionless critical load is obtained for both using $\bar{N}_{cr}^{i}=\widehat{N}_{xx}^{i}\frac{12L^2}{E_cH^3}$, $i=l,nl$, where $\widehat{N}_{xx}^{i}$ is the dimension load value. The studied nanobeam is characterised by unit width and total thickness $H=10\ \mathrm{nm}$ that consists of FGM porous core thickness $h=7\ nm$, and piezoelectric layers thickness $h_p=1.5\ nm$. Length of the nanobeam is assumed to be variable while studying the effect of length-to-thickness ratio. Material properties of the nanostructure are taken as the following: $E_c=380\ GPa$ and $E_m=70\ GPa$ as Young's moduli of FGM core; and $D_{xx}=226\ GPa$ and $D_{xz}=44.2\ GPa$ as elastic

DOI 10.2478/ama-2022-0019

coefficients of piezoelectric layers. $e_x = -2.2\ C/m^2$ and $e_z = 5.8\ C/m^2$ are values of piezoelectric permittivity constants, whereas $\epsilon_x = 5.64 \cdot 10^{-9}\ C/Vm$ and $\epsilon_z = 6.35 \cdot 10^{-9}\ C/Vm$ are dielectric permittivity constants. Poisson's ratio is assumed to be $\upsilon = 0.3$ for both piezoelectric face-sheets and FGM core.

First, Fig. 3 presents differences in dimensionless buckling load caused by in-plane mechanical force and external voltage for diverse nonlocal and length scale parameters for diverse length-to-thickness ratios $L\ /\ H$. In this case, we assumed a homogeneous nanobeam $(g = 0)$. Nonlocal to length scale parameters ratio is introduced as $\psi = \mathfrak{B}\ /\ \ell$ and taken for this study as $\psi = 2\ [nm]$ for Fig. 3(a). Then, for Fig. 3(b), the constant value of Eringen's nonlocal coefficient is assumed as $\mathfrak{B} = 0$. Finally, for Fig. 3(c), there is analysis for the pseudorandom ratio of $\mathfrak{B}\ /\ \ell$. Values for pseudorandom ratio were chosen in manner that enables clearly presenting the impact of size-dependent parameters. It is necessary to analyse diverse relationships between the nonlocal parameter of Eringen and the length scale coefficient because these depend on initial stress, rotary inertia, geometrical aspects ratio and BCs, as well as material properties [70-73]. Table 3 presents selected numerical results, which were further used in Fig. 3(b).

Tab. 3. Differences in dimensionless buckling load for diverse length scale parameter assuming $\mathfrak{B} = 0$

ℓ	0	1.5	3	4.5	6
		$L = 50\ nm$			
$\bar{N}^i_{cr}$	6.693	6.753	6.931	7.228	7.645
$\delta(\%)$	0.000	0.888	3.553	7.994	14.212
		$L = 100\ nm$			
$\bar{N}^i_{cr}$	7.096	7.112	7.159	7.238	7.349
$\delta(\%)$	0.000	0.222	0.888	1.998	3.553
		$L = 150\ nm$			
$\bar{N}^i_{cr}$	7.176	7.184	7.205	7.240	7.290
$\delta(\%)$	0.000	0.099	0.395	0.888	1.579

At this stage, it should be stated that the nonlocal interaction range obtained from mechanical critical buckling load and that from critical voltage (external voltage causing nanobeams' buckling response) are equivalent. This is caused by the fact that electric field contribution is represented in equations of motion as the in-plane electric force acting at the same point as mechanical loads. It should be noted, based on results that have been derived but are not presented here, that for all $L\ /\ H$ ratios, material gradations and porosity types along with nanoscale coefficients, the ratio of mechanical buckling load to critical voltage is constant. Therefore, the influence of mechanical loads and applied voltage may be adjusted, based on their relationship to one another, in complexly loaded structures, to omit their buckling response.
From Fig. 3, it may be generally observed that, for increasing value of $L\ /\ H$, the impact of size-dependent coefficients decreases. It should be cleared up, that by increasing the length-to-thickness ratio, the nanostructure is lengthened because the constant thickness is assumed. Both the nonlocal coefficient and the length scale parameter have values in nanoscale. Therefore, it may be concluded that nanoscale effects disappear when nanobeam is lengthened to an extent greater than $150\ nm$. From the other perspective, the change of thickness of the nanostructure should be examined. Nonetheless, the differential operator in constitutive relations acts in the x-direction, and thus that change does not influence mechanical response. On that account, enhancement of the constitutive relations of the theory may be an important step towards analysing nanoscale effects in diverse directions through nanostructures. The increasing value of the length scale parameter, assuming a constant ratio ψ, generates an initial increase in the δ parameter, then its decrease when ℓ is tending to $\ell = 2\ nm$ and finally an increase with further increasing of the length scale parameter. Initial increment of small-scale parameters indicates primary stress and strain gradients from the nonlocal theory.

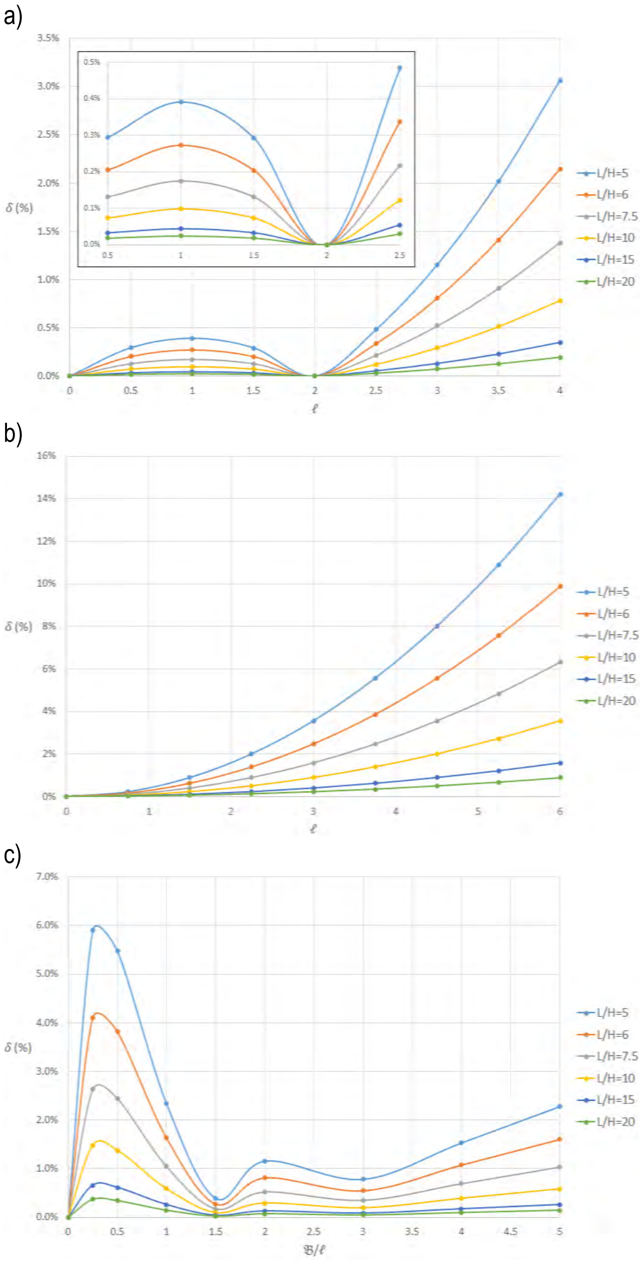


Fig. 3. Effect of length-to-thickness ratio on nonlocal interaction range parameter for dimensionless buckling load as well as critical voltage: (a) constant ratio $\psi = 2\ [nm]$; (b) nonlocal parameter $\mathfrak{B} = 0$; (c) pseudorandom ratio of $\mathfrak{B}/\ell$

DOI 10.2478/ama-2022-0019

Based on the figure, it may be concluded that this observation gleaned from the present study confirms those of previous studies [74,75] that softening and hardening nonlocal effects disappear when $\ell^2 = \mathfrak{B}$. In that case, size-dependent parameters cancel each other out and both constitutive relations and equations of motion are reduced to classical continuum theory. Another conclusion of this observation is the information that increasing the length scale parameter ℓ generates higher differences between classical and nonlocal buckling responses. The length scale coefficient is coupled with a greater number of derivatives in equations of motion than Eringen's nonlocal parameter. Therefore, increasing the value of ℓ increases the number of additional gradients of strains (hardening effect), and then increase in the differences between classical and nonlocal approaches. Furthermore, as can be seen from Fig. 3(c), appropriate identification of ψ parameter has a key role in obtaining the accurate mechanical response of nanostructures.

Dependence of the nonlocal interaction range parameter for buckling behaviour of homogeneous $(g = 0)$ nanobeam on external voltage as well as diverse small-scale parameters is studied in Fig. 4. For this study, it is assumed that $\psi = 1.2\,nm$ and $L = 100\,nm$. The positive/negative values of ϕ_0 refer to compression/extension, and in consequence, lead to small shortening/lengthening of the nanostructure. Applying additional electrical loads generate strains and stresses in the structure. Nonlocal gradients of strains and stresses overlap with electrically generated ones, and consequently, their influence on the buckling behaviour is increased. Increasing positive values of critical voltage cause an increase in the nonlocal interaction range, and, contrariwise, a negative value of external voltage decreases the interaction range parameter values, because positive and negative voltage values induce compression and tension of the structure. It should be also observed that compressively acting external voltage has a higher impact on the difference in the obtained results. This can be explained based on the position that compressive force generates higher induced stresses, a beam is shortened and bent, and consequently the nonlocal interaction is stronger in comparison to an unloaded structure.

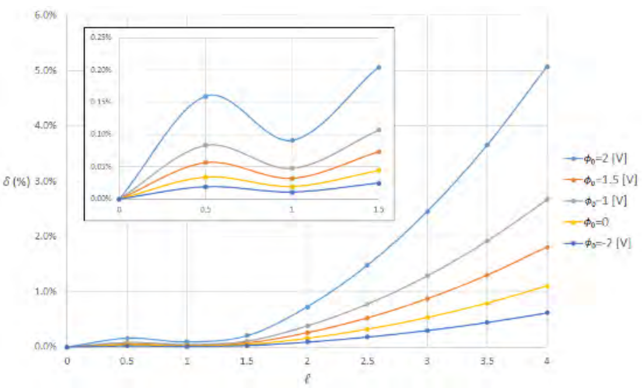


Fig. 4. Effect of external electric voltage and size-dependent parameters on nonlocal interaction range parameter for dimensionless buckling response of the nanobeam

The impact of a material gradation through the nanobeams' core thickness in conjunction with pseudorandom distribution of small-scale coefficient on the nonlocal interaction range parameter for buckling response is studied for diverse aspect ratios and displayed in Fig. 5. Selected results, for nanobeam with $L = 150\,nm$, are presented in Tab. 4. In the current investigation, the nanobeam is subjected to in-plane electric forces induced by external voltage $\phi_0 = 0.65\,V$. The difference between the results obtained from classical and nonlocal based approaches is notably higher for a higher ratio of $L\,/\,H$. It is worth noting that for homogeneous $(g = 0)$ structure, the nonlocal interaction range parameter is lower for a higher length-to-thickness ratio, which is consistent with results presented previously. The influence of size-dependent coefficients is similar to that observed in previous cases. Nevertheless, for nanostructure with a higher $L\,/\,H$ ratio, the difference increases remarkably with increasing the power-law index g. The higher the power-law index value, the softer the structure becomes, and thus the lower is the force needed to buckle the structure. Considering applied electrical force, increasing g index causes that structure to be near to the buckling. From this, we infer that the effect of small-scale parameters is considerably greater when the structure is under loads that are near to critical value. Even if we use nonlocal differential operator for abscissa, gradation of the material parameters through beam thickness has an indirect impact on the behaviour of structure (stress concentration, bending). Thus, shortening and extension are significantly dependent on this material distribution as well as a direct effect of nonlocal parameters.

Tab. 4. Differences in dimensionless buckling load for diverse nonlocal coefficients and the power law index for nanobeam with $L = 150\,nm$

$\mathfrak{B}\,/\,\ell$	$\delta(\%)$					
	g					
	0	1	2	3	4	5
0	0.000	0.000	0.000	0.000	0.000	0.000
0.25	1.516	3.179	4.805	5.978	6.746	7.250
0.5	1.415	2.965	4.483	5.577	6.294	6.763
1	0.606	1.271	1.920	2.389	2.696	2.897
1.5	0.101	0.211	0.320	0.398	0.449	0.483
2	0.302	0.634	0.959	1.193	1.346	1.447
3	0.202	0.424	0.640	0.796	0.899	0.966
5	0.604	1.267	1.915	2.382	2.688	2.888

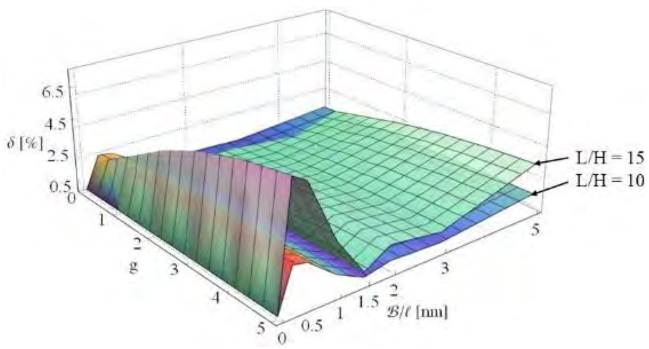


Fig. 5. Effect of material gradation, size-dependent parameters and length-to-thickness ratio on nonlocal interaction range parameter for dimensionless buckling response of the nanobeam

Figure 6 shows an effect of diverse porosity distributions together with a volume of voids on nonlocal interaction range parameter for buckling behaviour of nanobeam $(g = 0)$ with a different length-to-thickness ratio that is subjected to external

DOI 10.2478/ama-2022-0019

electric voltage $\phi_0 = 1.5\,V$. For this study, size-dependent parameters are fixed as $\ell = 4\,nm$ and $\mathfrak{B} = 1\,nm^2$. Analogous to the previous study, as a result of increasing the length of the nanobeam, the value of the nonlocal interaction range parameter decreases. Further, as a result of enlargement in nanobeam length, nano-scaled size-dependent coefficients have less influence on the mechanical response of the nanostructure. On the other hand, increasing volume of voids (ϑ) magnifies differences between classical and nonlocal buckling responses, regardless of porosity distribution. Differences in obtained results for diverse porosity accumulation follow from different impacts on material resultant stiffness. With the higher $L\,/\,H$ ratio and porosity parameter ϑ, the applied voltage is closer to the critical value. It is clearly observed that as the buckling point approaches, the impact of nano-scale coefficients increases.

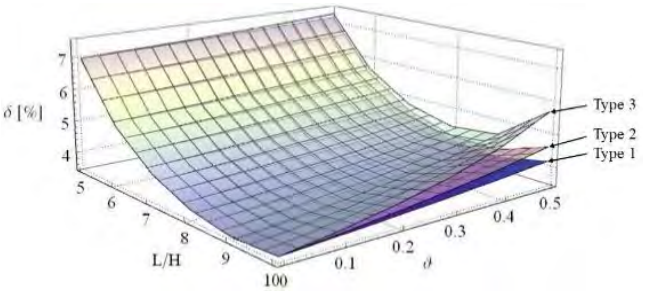


Fig. 6. Effect of length-to-thickness ratio and porosity distribution on nonlocal interaction range parameter for dimensionless buckling response of the nanobeam

Next, the difference between classical and nonlocal based critical porosity is provided. Critical porosity is defined as the volume of voids in the structure, under constant compressive forces, that leads to its buckling. The nonlocal interaction range parameter is obtained in a way that is parallel to the procedure used to derive the buckling-load–based parameters. The impact of diverse porosity distribution in conjunction with nonlocal parameters is displayed in Tab. 5 and Fig. 7. In this case it is considered homogeneous $(g = 0)$, but porous nanostructure under in-plane compressive forces $\widehat{N}_{xx} = 35\,N$. Nonlocal to length scale parameters ratio is assumed $\psi = 2\,[nm]$, and length of the structure as $L = 75\,nm$. Analogous to critical loads, increasing size-dependent parameters increases differences between critical porosity obtained based on classical and nonlocal theories because it strengthens the influence on higher derivatives in equations of motion. Further, increasing nano-scale coefficients' values leads to widening of the differences between Type 1 porosity and Types 2 and 3. The effects of Type 2 and Type 3 porosity distribution are very similar because both are characterised by zero volume of voids in the mid-plane where compressive force is acting; and consequently, they have a similar impact on nanostructure characteristics.

Continuing the previous investigation, the effect of the diverse ratios of nonlocal to length scale coefficients together with the power-law index (g) on nonlocal interaction range parameter for critical porosity is presented in Fig. 8. In this study, it is assumed $L = 100\,nm$ and in-plane mechanical force $\widehat{N}_{xx} = 13\,N$. Likewise, in the previous study, the values of the nonlocal interaction range parameter for Types 2 and 3 are close, and for Type 1 the value is significantly lower, regardless of the power-law index. Further, the impact of nonlocal parameters increases resultant to increasing the power-law index. The higher the value of the index, the lower the resultant stiffness, and consequently, the applied compressive force is closer to the critical one. The study also supplements the statement that appropriate identification of Eringen's nonlocal to length scale parameter ratio is crucial in properly anticipating the mechanical response of nanostructures regardless of their homogeneity or inhomogeneity.

Tab. 5. Differences in critical porosity for diverse nonlocal coefficients ratio for nanobeam with $L = 75\,nm$

	ℓ				
	0	1	2	3	4
TYPE 1					
ϑ^i_{cr}	0.563	0.555	0.563	0.584	0.620
$\delta(\%)$	0.000	1.317	0.000	3.871	10.120
TYPE 2					
ϑ^i_{cr}	0.396	0.390	0.396	0.412	0.439
$\delta(\%)$	0.000	1.400	0.000	4.139	10.894
TYPE 3					
ϑ^i_{cr}	0.266	0.262	0.266	0.277	0.295
$\delta(\%)$	0.000	1.414	0.000	4.182	11.020

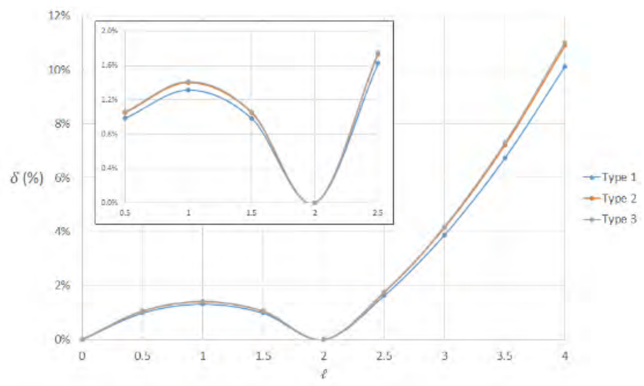


Fig. 7. Effect of porosity distribution and size-dependent parameters on nonlocal interaction range parameter for critical porosity of the nanobeam

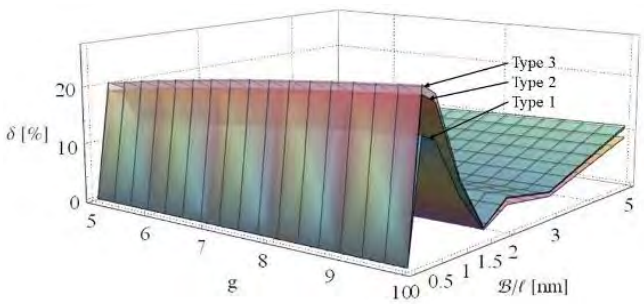


Fig. 8. Effect of porosity distribution and size-dependent parameters on nonlocal interaction range parameter for critical porosity of the nanobeam

6. CONCLUSIONS

The current investigation presents an analysis of differences in the mechanical stability of nanostructures obtained based on classical and nonlocal theories. The smart nanostructure is considered as FG porous nanobeam with piezoelectric layers. FGM parameters are achieved by using modified Voigt's rule of mixtures and assessing the contribution of an electric field based on a combination of half-cosine and linear variations of the electric potential. Employed equations of motion include both Eringen's nonlocal coefficient and length scale parameter based on the nonlocal strain gradient theory. In the modelling of the operation environment in NEMS devices, the nanostructure is subjected to mechanical in-plane forces and electric field. The utilised refined Reddy third-order shear deformation theory allows adopting the beam model to a wide range of structures without committing errors arising from other beam models' assumptions. The presented results include effects of diverse Eringen's to length scale parameters, nanobeams length-to-thickness ratio, material gradation and porosity distributions, as well as external loadings on the stability of nano-scaled smart beam, supplemented with interpretation and discussion.

In the study, we defined the nonlocal interaction range parameter and checked the boundaries between an application of classical and nonlocal elasticity theories for intelligent nanobeam. The investigation presents a comprehensive model of nano-scaled structure that complies with a previously presented study claiming that size-dependent parameters depend on initial stress and geometrical aspects, as well as material properties. The consideration demonstrates that the influence of size-dependent coefficients decreases resultant to increase in geometrical parameters of the structure. On the other hand, the analysis shows that applied force (especially near to the critical value), as well as material properties, have a significant influence on the importance of nonlocal parameters in the modelling of nanostructures. The conducted study furthers the knowledge and understanding of nonlocal parameters' effects on the stability of smart nanostructures that are an important component of NEMS. Consequently, the results of the study positively impact the possibility of optimisation and control of NEMS devices. Nonetheless, simplified constitutive equations [see Eq. (5)] may not be sufficient in specific conditions; therefore, future studies should focus on influence of the non-equal nonlocal coefficients e_1 and e_0. Further, improving constitutive relations, including multidimensional differential operator, may be an important improvement of the theory, especially for nonhomogeneous nanostructures.

REFERENCES

1. Bhushan B. (Ed.) Springer Handbook of Nanotechnology. Springer-Verlag; 2004.
2. Zhao Q, Gan Z, Zhuang Q. Electrochemical sensors based on carbon nanotubes. Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis. 2002;14: 1609–1613.
3. Briscoe J, Dunn S. Piezoelectric nanogenerators – a review of nanostructured piezoelectric energy harvesters. Nano Energy. 2015;14: 15–29.
4. Fennimore A, Yuzvinsky T, Han W-Q, Fuhrer M, Cumings J, Zettl A. Rotational actuators based on carbon nanotubes. Nature. 2003;424: 408-410.
5. Ghayesh MH, Farajpour A. A review on the mechanics of functionally graded nanoscale and microscale structures. International Journal of Engineering Science. 2019;137: 8-36.
6. Lu L, Guo X, Zhao J. A unified nonlocal strain gradient model for nanobeams and the importance of higher order terms. International Journal of Engineering Science. 2017;119, 265-277.
7. Toupin RA. Elastic materials with couple-stresses. Archive for Rational Mechanics and Analysis. 1962;11: 385-414.
8. Mindlin RD, Tiersten HF. Effects of couple-stresses in linear elasticity. Archive for Rational Mechanics and Analysis. 1962;11: 415-448.
9. Koiter WT. Couple stresses in the theory of elasticity. I and II. Nederl Akad Wetensch Proc Ser B. 1964;67: 17-44.
10. Yang F, Chong ACM, Lam DCC, Tong P. Couple stress based strain gradient theory for elasticity. International Journal of Solids and Structures. 2002;39: 2731-2743.
11. Mindlin RD. Micro-structure in linear elasticity. Archive for Rational Mechanics and Analysis. 1964;16: 51-78.
12. Mindlin RD. Second gradient of strain and surface-tension in linear elasticity. International Journal of Solids and Structures. 1965;1: 417-438.
13. Lam DCC, Yang F, Chong ACM, Wang J, Tong P. Experiments and theory in strain gradient elasticity. Journal of the Mechanics and Physics of Solids. 2003;51: 1477-1508.
14. Kroner E. Elasticity theory of materials with long range cohesive forces. International Journal of Solids and Structures. 1967;3: 731-742.
15. Eringen AC. Nonlocal polar elastic continua. International Journal of Engineering Science. 1972;10: 1-16.
16. Eringen AC. Linear theory of nonlocal elasticity and dispersion of plane waves. International Journal of Engineering Science. 1972;10: 425-435.
17. Eringen AC, Edelen DGB. On nonlocal elasticity. International Journal of Engineering Science. 1972;10: 233-248.
18. Eringen AC. On differential equations of nonlocal elasticity and solutions of screw dislocation and surface wave. Journal of Applied Physics. 1983;54: 4703-4710.
19. Romano G, Barretta R. Nonlocal elasticity in nanobeams: the stress-driven integral model. International Journal of Engineering Science. 2017;115: 14-27.
20. Yang B, Vehoff H. Dependence of nanohardness upon indentation size and grain size – A local examination of the interaction between dislocations and grain boundaries. Acta Materialia. 2007;55: 849-856.
21. Voyiadjis GZ, Peters R. Size effects in nanoindentation: an experimental and analytical study. Acta Mechanica. 2010;211: 131-153.
22. Askes H, Aifantis EC. Gradient elasticity and flexural wave dispersion in carbon nanotubes. Physical Review B. 2009;80: 195412.
23. Lim CW, Zhang G, Reddy JN. A higher-order nonlocal elasticity and strain gradient theory and its applications in wave propagation. Journal of the Mechanics and Physics of Solids. 2015;78, 298-313.
24. Demir Ç, Civalek Ö. Torsional and longitudinal frequency and wave response of microtubules based on the nonlocal continuum and nonlocal discrete models. Applied Mathematical Modelling. 2013;37: 9355-9367.
25. Akgöz B, Civalek Ö. Longitudinal vibration analysis for microbars based on strain gradient elasticity theory. Journal of Vibration and Control. 2014;20: 606-616.
26. Cornacchia F, Fabbrocino F, Fantuzzi N, Luciano R, Pena R. Analytical solution of cross- and angle-ply nano plates with strain gradient theory for linear vibrations and buckling. Mechanics of Advanced Materials and Structures. 2021;28: 1201-1215.
27. Żur KK, Farajpour A, Lim CW, Jankowski P. On the nonlinear dynamics of porous composite nanobeams connected with fullerenes. Composite Structures. 2021;274: 114356.
28. Monaco GT, Fantuzzi N, Fabbrocino F, Luciano R. Critical Temperatures for Vibrations and Buckling of Magneto-Electro-Elastic Nonlocal Strain Gradient Plates. Nanomaterials. 2021;11: 87.
29. Jalaei MH, Ghorbanpour Arani A, Nguyen-Xuan H. Investigation of thermal and magnetic field effects on the dynamic instability of FG Timoshenko nanobeam employing nonlocal strain gradient theory. International Journal of Mechanical Sciences. 2019;161-162: 105043.

DOI 10.2478/ama-2022-0019

30. Żur KK, Arefi M, Kim J, Reddy JN. Free vibration and buckling analyses of magneto-electro-elastic FGM nanoplates based on nonlocal modified higher-order sinusoidal shear deformation theory. Composites Part B: Engineering. 2020;182: 107601.
31. Zhao X, Zheng S, Li Z. Effects of porosity and flexoelectricity on static bending and free vibration of AFG piezoelectric nanobeams. Thin-Walled Structures. 2020;151: 106754.
32. Faghidian SA, Żur KK, Pan E, Kim J. On the analytical and meshless numerical approaches to mixture stress gradient functionally graded nano-bar in tension. Engineering Analysis with Boundary Elements. 2022;134: 571-580.
33. Barretta R, Caporale A, Faghidian SA, Luciano R, Marotti de Sciarra F, Medaglia CM. A stress-driven local-nonlocal mixture model for Timoshenko nano-beams. Composites Part B: Engineering. 2019;164: 590-598.
34. Barretta R, Fazelzadeh SA, Feo L, Ghavanloo E, Luciano R. Nonlocal inflected nano-beams: A stress-driven approach of bi-Helmholtz type. Composite Structures. 2018;200: 239-245.
35. Barretta R, Faghidian SA, Luciano R, Medaglia CM, Penna R. Stress-driven two-phase integral elasticity for torsion of nano-beams. Composites Part B: Engineering. 2018;145: 62-69.
36. Jalaei MH, Thai HT, Civalek Ö. On viscoelastic transient response of magnetically imperfect functionally graded nanobeams. International Journal of Engineering Science. 2022;172: 103629.
37. Monaco GT, Fantuzzi N, Fabbrocino F, Luciano R. Trigonometric Solution for the Bending Analysis of Magneto-Electro-Elastic Strain Gradient Nonlocal Nanoplates in Hygro-Thermal Environment. Mathematics. 2021;9: 567.
38. Penna R, Feo L, Lovisi G., Hygro-thermal bending behavior of porous FG nano-beams via local/nonlocal strain and stress gradient theories of elasticity. Composite Structures. 2021;263: 113627.
39. Monaco GT, Fantuzzi N, Fabbrocino F, Luciano R. Semi-analytical static analysis of nonlocal strain gradient laminated composite nanoplates in hygrothermal environment. Journal of the Brazilian Society of Mechanical Sciences and Engineering. 2021;43: 274.
40. Faghidian SA, Żur KK, Reddy JN. A mixed variational framework for higher-order unified gradient elasticity. International Journal of Engineering Science. 2022;170: 103603.
41. Apuzzo A, Barretta R, Faghidian SA, Luciano R, Marotti de Sciarra F. Nonlocal strain gradient exact solutions for functionally graded inflected nano-beams. Composites Part B: Engineering. 2019;164: 667-674.
42. Reddy JN. Nonlocal theories for bending, buckling and vibration of beams. International Journal of Engineering Science. 2007;45: 288-307.
43. Aydogdu M. A general nonlocal beam theory: Its application to nanobeam bending, buckling and vibration. Physica E: Low-dimensional Systems and Nanostructures. 2009;41: 1651–1655.
44. Roque CMC, Ferreira AJM, Reddy JN. Analysis of Timoshenko nanobeams with a nonlocal formulation and meshless method. International Journal of Engineering Science. 2011;49: 976-984.
45. Thai HT. A nonlocal beam theory for bending, buckling, and vibration of nanobeams. International Journal of Engineering Science. 2012;52: 56–64.
46. Thai HT, Vo TP. A nonlocal sinusoidal shear deformation beam theory with application to bending, buckling, and vibration of nanobeams. International Journal of Engineering Science. 2012;54: 58–66.
47. Ghannadpour SAM, Mohammadi B, Fazilati J. Bending, buckling and vibration problems of nonlocal Euler beams using Ritz method. Composite Structures. 2013;96: 584–589.
48. Şimşek M, Yurtcu HH. Analytical solutions for bending and buckling of functionally graded nanobeams based on the nonlocal Timoshenko beam theory. Composite Structures. 2013;97: 378-386.
49. Rahmani O, Jandaghian AA. Buckling analysis of functionally graded nanobeams based on a nonlocal third-order shear deformation theory. Applied Physics A. 2015;119: 1019–1032.
50. Chaht FL, Kaci A, Houari MSA, Tounsi A, Bég OA, Mahmoud SR. Bending and buckling analyses of functionally graded material (FGM) size-dependent nanoscale beams including the thickness stretching effect. Steel and Composite Structures. 2015;18: 425-442.
51. Yu YJ, Xue Z-N, Li C-L, Tian X-G. Buckling of nanobeams under nonuniform temperature based on nonlocal thermoelasticity. Composite Structures. 2016;146: 108–113.
52. Nejad MZ, Hadi A, Rastgoo A. Buckling analysis of arbitrary two-directional functionally graded Euler–Bernoulli nano-beams based on nonlocal elasticity theory. International Journal of Engineering Science. 2016;103: 1–10.
53. Li X, Li L, Hu Y, Ding Z, Deng W. Bending, buckling and vibration of axially functionally graded beams based on nonlocal strain gradient theory. Composite Structures. 2017;165: 250–265.
54. Tuna M, Kirca M. Bending, buckling and free vibration analysis of Euler-Bernoulli nanobeams using Eringen's nonlocal integral model via finite element method. Composite Structures. 2017;179: 269–284.
55. Mirjavadi SS, Afshari MB, Khezek M, Shafiei N, Rabby S, Kordnejad M. Nonlinear vibration and buckling of functionally graded porous nanoscaled beams. Journal of Brazilian Society of Mechanical Sciences and Engineering. 2018;40: 352.
56. Khaniki HB, Hosseini-Hashemi Sh, Nezamabadi A. Buckling analysis of nonuniform nonlocal strain gradient beams using generalized differential quadrature method. Alexandria Engineering Journal. 2018;57: 1361–1368.
57. Alibeigi B, Tadi Beni Y. On the size-dependent magneto/electromechanical buckling of nanobeams. European Physical Journal Plus. 2018;133: 398.
58. Alibeigi B, Tadi Beni Y, Mehralian F. On the thermal buckling of magneto-electroelastic piezoelectric nanobeams. European Physical Journal Plus. 2018;133: 133.
59. Xiao W, Gao Y, Zhu H. Buckling and post-buckling of magneto-electro-thermo-elastic functionally graded porous nanobeams. Microsystem Technologies. 2019;25: 2451-2470.
60. Hashemian M, Foroutan S, Toghraie D. Comprehensive beam models for buckling and bending behavior of simple nanobeam based on nonlocal strain gradient theory and surface effects. Mechanics of Materials. 2019;139: 103209.
61. Jankowski P, Żur KK, Kim J, Reddy JN. On the bifurcation buckling and vibration of porous nanobeams. Composite Structures. 2020;250: 112632.
62. Civalek Ö, Uzun B, Yayli MÖ. Stability analysis of nanobeams placed in electromagnetic field using a finite element method. Arabian Journal of Geoscience. 2020;13: 1165.
63. Civalek Ö, Uzun B, Yayli MÖ. Finite element formulation for nanoscaled beam elements, ZAMM – Journal of Applied Mathematics and Mechanics. 2021;e202000377.
64. Jankowski P, Żur KK, Kim J, Lim CW, Reddy JN. On the piezoelectric effect on stability of symmetric FGM porous nanobeams. Composite Structures. 2021;267: 113880.
65. Eringen AC. Nonlocal Continuum Field Theories, Springer;2002.
66. Reddy JN. Energy principles and variational methods in applied mechanics. John Wiley & Sons;2017.
67. Wang Q. On buckling of column structures with a pair of piezoelectric layers. Engineering Structures. 2002;24(2): 199–205.
68. Nguyen T-K, Vo TP, Nguyen B-D, Lee J. An analytical solution for buckling and vibration analysis of functionally graded sandwich beams using a quasi-3D shear deformation theory. Composite Structures. 2016;156: 238-252.
69. Vo TP, Thai HT, Nguyen T-K, Maheri A, Lee J. Finite element model for vibration and buckling of functionally graded sandwich beams based on a refined shear deformation theory. Engineering Structures. 2014;64: 12–22.
70. Ghavanloo E, Fazelzadeh SA. Evaluation of nonlocal parameter for single-walled carbon nanotubes with arbitrary chirality. Meccanica. 2016;51: 41-54.

71. Thai HT, Vo TP, Nguyen TK, Kim SE. A review of continuum mechanics models for size-dependent analysis of beams and plates. Composite Structures. 2017;177: 196-219.
72. Zhang Z, Wang CM, Challamel N. Eringen's length scale coefficient for buckling of nonlocal rectangular plates from microstructured beam-grid model. International Journal of Solids and Structures. 2014;51: 4307-4315.
73. Mehralian F, Tadi Beni Y, Zeverdejani MK. Calibration of nonlocal strain gradient shell model for buckling analysis of nanotubes using molecular dynamics simulations. Physica B: Condensed Matter. 2017;521: 102-111.
74. Zeighampour H, Tadi Beni Y. Size dependent analysis of wave propagation in functionally graded composite cylindrical microshell reinforced by carbon nanotube. Composite Structures. 2017;179: 124-131.
75. Lu L, Guo X, Zhao J. Size-dependent vibration analysis of nanobeams based on the nonlocal strain gradient theory. International Journal of Engineering Science. 2017;116: 12–24.

The work has been conducted within WZ/WM-IIM/3/2020 project and was financed by the funds of the Ministry of Science and Higher Education, Poland

Piotr Jankowski: https://orcid.org/0000-0003-4137-5899

Appendix

$$\left\{A_{xx}^{(0)},A_{xx}^{(1)},A_{xx}^{(2)},A_{xx}^{(3)},A_{xx}^{(4)},A_{xx}^{(6)}\right\}=$$
$$b\int_{-h/2-h_p}^{h/2} D_{xx}\{1,z,z^2,z^3,z^4,z^6\}dz + b\int_{-h/2}^{0} C_{xx}\{1,z,z^2,z^3,z^4,z^6\}dz + b\int_{0}^{h/2} C_{xx}\{1,z,z^2,z^3,z^4,z^6\}dz +$$
$$b\int_{h/2}^{h/2+h_p} D_{xx}\{1,z,z^2,z^3,z^4,z^6\}dz, \tag{A.1}$$

$$\left\{A_{xz}^{(0)},A_{xz}^{(2)},A_{xz}^{(4)}\right\}=$$
$$b\int_{-h/2-h_p}^{h/2} D_{xz}\{1,z^2,z^4\}dz + b\int_{-h/2}^{0} C_{xz}\{1,z^2,z^4\}dz + b\int_{0}^{h/2} C_{55}\{1,z^2,z^4\}dz + b\int_{h/2}^{h/2+h_p} D_{xz}\{1,z^2,z^4\}dz, \tag{A.2}$$

$$\left\{B_{x}^{(0)},B_{x}^{(1)},B_{x}^{(3)}\right\} = b\int_{-h/2-h_p}^{h/2} e_x\frac{\pi}{h_p}sin\left(\frac{\pi z_2}{h_p}\right)\{1,z,z^3\}dz + b\int_{h/2}^{h/2+h_p} e_x\frac{\pi}{h_p}sin\left(\frac{\pi z_1}{h_p}\right)\{1,z,z^3\}dz, \tag{A.3}$$

$$\left\{B_{z}^{(0)},B_{z}^{(2)}\right\} = b\int_{-h/2-h_p}^{h/2} e_z cos\left(\frac{\pi z_2}{h_p}\right)\{1,z^2\}dz + b\int_{h/2}^{h/2+h_p} e_z cos\left(\frac{\pi z_1}{h_p}\right)\{1,z^2\}dz, \tag{A.4}$$

$$\left\{B_{x\phi}^{(0)},B_{x\phi}^{(1)},B_{x\phi}^{(3)}\right\} = b\int_{-h/2-h_p}^{h/2} e_x\frac{2}{h_p}\phi_0\{1,z,z^3\}dz + b\int_{h/2}^{h/2+h_p} e_x\frac{2}{h_p}\phi_0\{1,z,z^3\}dz, \tag{A.5}$$

$$C_x = b\int_{-h/2-h_p}^{h/2}\epsilon_x\left(cos\left(\frac{\pi z_2}{h_p}\right)\right)^2 dz + b\int_{h/2}^{h/2+h_p}\epsilon_x\left(cos\left(\frac{\pi z_1}{h_p}\right)\right)^2 dz, \tag{A.6}$$

$$C_z = b\int_{-h/2-h_p}^{h/2}\epsilon_z\left(\frac{\pi}{h_p}sin\left(\frac{\pi z_2}{h_p}\right)\right)^2 dz + b\int_{h/2}^{h/2+h_p}\epsilon_z\left(\frac{\pi}{h_p}sin\left(\frac{\pi z_1}{h_p}\right)\right)^2 dz, \tag{A.7}$$

$$C_{z\phi} = b\int_{-h/2-h_p}^{h/2}\epsilon_z\frac{2}{h_p}\phi_0\frac{\pi}{h_p}sin\left(\frac{\pi z_2}{h_p}\right)dz + b\int_{h/2}^{h/2+h_p}\epsilon_z\frac{2}{h_p}\phi_0\frac{\pi}{h_p}sin\left(\frac{\pi z_1}{h_p}\right)dz. \tag{A.8}$$

DOI 10.2478/ama-2022-0020

NUMERICAL STUDY OF TRANSIENT ELASTOHYDRODYNAMIC LUBRICATION SUBJECTED TO SINUSOIDAL DYNAMIC LOADS FOR ROUGH CONTACT SURFACES

Mohamed F. ABD ALSAMIEH*

*Mechanical Design & Production Department, Military Technical College, Ismail Al Fangari, El-Qobba Bridge, El Weili, Cairo, Egypt

mohamed.fahmy203@hotmail.com

received 22 August 2021, revised 23 November 2021, accepted 2 December 2021

Abstract: The purpose of this paper is to study the behaviour of transient elastohydrodynamic contacts subjected to forced harmonic vibrations, including the effect of surface waviness for concentrated counterformal point contact under isothermal conditions. Profiles of pressure and film thickness are studied to reveal the combined effects of sinusoidal external load and surface roughness on the lubrication problem. The time-dependent Reynolds' equation is solved using Newton–Raphson technique. The film thickness and pressure distribution are obtained at different snap shots of time by simultaneous solution of the Reynolds' equation and film thickness equation including elastic deformation and surface waviness. It is concluded that the coupling effects of the transient sinusoidal external load and wavy surface would result in increase in modulations of the pressure and film thickness profile in comparison to the case where the smooth contact surfaces are subjected to sinusoidal external load.

Key words: transient elastohydrodynamics, surface roughness, sinusoidal load, wavy surfaces

1. INTRODUCTION

Elastohydrodynamic lubrication is commonly known as a mode of fluid-film lubrication, in which the deformed surface and the pressure-dependent viscosity are taken into account to describe the formation of lubricant film. In fact, 90% of all machines and mechanisms comprise load-bearing and gear-transmitting contacts operating under elastohydrodynamic lubrication. However, it is generally known that machine elements operate under time-dependent conditions; therefore, the time-dependent motion of the elements results in transient lubrication. In such elements, the assumption of steady-state solution is inappropriate to study both film thickness and contact pressure to protect such components from direct contact and damage. In addition, the manufacturing processes of machine components cannot guarantee that the surfaces will be perfectly smooth; there will be imperfections on the surfaces, such as waviness. This imperfect surface may cause changes in the film thickness and pressure. The existence of such imperfections, however, may result in direct contacts of asperities once the film thickness is below a certain limit (see [1-3]).

Effect of transient conditions for elastohydrodynamic lubrication subjected to variation of load has been studied both experimentally and numerically in the past. Experimental investigations using optical interferometry technique reported in [4-10] showed that an oil entrapment is formed during rapid increase of the load, and as the entrainment speed increases, this entrapment of oil is diminished. The behaviour of transient elastohydrodynamic lubrication subjected to vibration has been studied numerically by many researchers such as in [11-14]. Their results revealed the solution of transient contact conditions subjected to variation of load is completely different from steady-state solution, especially at a high value of amplitude and frequency of vibration.

In the past, longitudinal and transverse roughness of ridges was studied both theoretically and experimentally. A numerical solution for the moving of a sinusoidal single transverse ridge through elastohydrodynamic lubrication for line and point contact problem have been investigated by many researchers as shown in [15-16]. Experimental studies using optical interferometry of the moving of a single transverse ridge through circular elastohydrodynamic contact have been presented in [17-18]. Effect of geometrical characteristics of the ridge on the formation of lubricant film thickness for elastohydrodynamic lubrication of point contact problem under rolling/sliding conditions haven been studied both experimentally and numerically in [19-20]. They used a single flat-top transverse ridge and the results showed that the lubricant film thickness was mainly affected by the geometrical characteristics of the ridge. The passage of a single flat-top transverse ridge through elastohydrodynamic lubrication for point contact problem has been studied numerically using multigrid technique in [21]. Analytical solution for different surface features of rectangular, rounded bottom and triangular shape have been provided in [22-23]. The results revealed that, an optimal indentation profile should have a smooth shape with appropriate width and depth. Recently, the behavior of a single ridge passing through elastohydrodynamic lubrication for different ridge shape and size including; flat-top, triangular and cosine wave pattern to get an optimal ridge profile has been studied numerically [24]. The results showed that, the film thickness profile and the pressure distribution through the contact zone were mainly affected by the geometrical characteristics of the ridge.

However, in real applications, elastohydrodynamic contacts are subjected to variation of load, geometry or velocity of the

DOI 10.2478/ama-2022-0020 acta mechanica et automatica, vol.16 no.2 (2022)

contacting surfaces, often with more than one of the parameters varying at the same time, making the prediction of the film thickness a very difficult problem, even if these variations are known, which is not usually the case. The analysis of such behaviour may provide some views of the local force on the surfaces, and thus lead to a more reasonable prediction of the stress fields. Influence of the sinusoidal varying loads due to vibrations on the film thickness and pressure profile of the isothermal, elastohydrodynamically lubricated point contact with a wavy surface have been numerically investigated in [25]. The results revealed that the external sinusoidal dynamic load induces the modulations on the film thickness and pressure profile. The waviness on one contacting surface causes changes in the elastohydrodynamic behaviours to become more pronounced.

In the present paper, a detailed numerical solution and analysis of the elastohydrodynamic contacts subjected to a harmonic variation of load as well as the waviness of surface is considered for concentrated counterformal point contact under isothermal conditions. The behaviour of the transient elastohydrodynamic of point contact problem was evaluated relative to film thickness and pressure distribution, and on that basis, a time-dependent Reynolds' equation is solved using Newton-Raphson technique with Gauss-Seidel iteration method. The current numerical results have been compared with the results of the numerical work reported in [25].

2. BACKGROUND THEORY

The time-dependent Reynolds' equation can be written in a dimensionless form as:

$$\frac{\partial}{\partial X}\left(\frac{\bar{\rho}H^3}{\bar{\eta}}\frac{\partial P}{\partial X}\right)+\frac{\partial}{\partial Y}\left(\frac{\bar{\rho}H^3}{\bar{\eta}}\frac{\partial P}{\partial Y}\right)=\emptyset\left\{\frac{\partial(\bar{\rho}H)}{\partial X}+\frac{bE'}{\eta_0 u}\left(\bar{\rho}\frac{\partial H}{\partial \bar{t}}+H\frac{\partial\bar{\rho}}{\partial \bar{t}}\right)\right\} \quad (1)$$

The following dimensionless variables apply:

$$X=\frac{x}{b}, Y=\frac{y}{b}, \bar{\eta}=\frac{\eta}{\eta_0}, H=\frac{hR}{b^2}, P=\frac{p}{P_{Her}}, \phi=\frac{12R^2\eta_0 u}{b^3 P_{Her}}.$$

The film thickness is described in a dimensionless form as:

$$\Delta P_{k,l}^{n}=\frac{\left(-F_{i,j}-J_{k-1,l}^{i,j}\Delta P_{k-1,l}^{n}-J_{k+1,l}^{i,j}\Delta P_{k+1,l}^{n-1}-J_{k,l-1}^{i,j}\Delta P_{k,l-1}^{n}-J_{k,l+1}^{i,j}\Delta P_{k,l+1}^{n+1}\right)}{J_{k,l}^{i,j}}$$

where: $J_{k,l}^{i,j}=\frac{\partial F_{i,j}(t)}{\partial P_{k,l}}$ (2)

where m and l are the inlet boundaries in both the rolling and lateral directions, respectively, $H_0(T)$ is the dimensionless initial central film thickness, b is the radius of Hertzian contact region and $\delta(X,Y,T)$ is the total elastic deformation (see [1, 14 and 26]) and is given as

$$\delta_{I,J}(x,y)=\frac{2}{\Pi}\frac{P_{Her}}{E'}\sum_{j=1,2}^{ny}\sum_{i=1,2}^{nx}P_{i,j}D_{i^*,j^*}$$

where D is the contact influence coefficient matrix and is given in [27].

The last term in equation (2), $S.R(X,Y,T)$, represents the geometry of the surface roughness and is given as:

$$S.R(X,Y,T)=\begin{cases}0 & X\geq X_d\\ A\sin\left(2\pi\frac{X-X_d}{\lambda}\right) & X<X_d\end{cases}$$

where A is the dimensionless amplitude, λ is the dimensionless wavelength of the ridge and X_d is the dimensionless position of the located ridge through the contact zone and is given as $X_d={x_d}/{b}$, $x_d=x_0+ut$, x_0 is the initial position of the ridge at t = 0.

The variation of density with pressure was shown in [28] as:

$$\bar{\rho}=1+\frac{\varepsilon PP_{Her}}{1+\zeta PP_{Her}} \quad (3)$$

where ε and ζ are constants dependent upon the type of lubricant used.

The relation of viscosity with pressure was given in [29] as:

$$\bar{\eta}=exp[ln\eta_o+9.67]\left[(1+5.1*10^{-9}PP_{Her})^Z-1\right] \quad (4)$$

where z is the viscosity-pressure index and is given as:

The Reynolds' equation (1) can be solved using Newton-Raphson method in the following numerical form:

$$\sum_{l=2}^{N}\sum_{k=2}^{M}\frac{\partial F_{i,j}(t)}{\partial P_{k,l}}\Delta P_{k,l}=-F_{i,j}(t) \quad (5)$$

Using the Gauss–Seidel iteration method, equation (5) can be written as:

$$\Delta P_{k,l}^{n}=\frac{\left(-F_{i,j}-J_{k-1,l}^{i,j}\Delta P_{k-1,l}^{n}-J_{k+1,l}^{i,j}\Delta P_{k+1,l}^{n-1}-J_{k,l-1}^{i,j}\Delta P_{k,l-1}^{n}-J_{k,l+1}^{i,j}\Delta P_{k,l+1}^{n+1}\right)}{J_{k,l}^{i,j}}$$

where $J_{k,l}^{i,j}=\frac{\partial F_{i,j}(t)}{\partial P_{k,l}}$,

where n is the iteration counter.

The pressure can be updated according to the equation:

$$P_{i,j}^{n}=P_{i,j}^{n-1}+\Omega\Delta P_{i,j}^{n} \quad (6)$$

where Ω is an under-relaxation factor, which ranges from 0.05 to 0.1.

The external load variation ($W(t)$) is balanced by the integration of oil pressure distribution. The instantaneous load equation can be written as:

$$\iint_{-\infty}^{\infty}p(x,y,t)dxdy=W(t) \quad (7)$$

The convergence criteria for the pressure and load balance equations are:

$$\left[\frac{\sum_{j=1,2,..}^{N}\sum_{i=1,2,..}^{M}\left(P_{i,j}^{n}-P_{i,j}^{n-1}\right)^2}{MxN}\right]^{0.5}\leq 10^{-3} \quad (8)$$

$$\left|\iint P(X,Y)dXdY-\frac{2}{3}\pi\right|\leq 10^{-3} \quad (9)$$

where M and N are the total nodal points in both the X- and Y-directions, respectively.

3. RESULTS AND DISCUSSION

In this section, the steady and transient contact pressure distributions and film thickness profiles of elastohydrodynamic lubrication for smooth and wavy surface subjected to harmonic forced vibrations are presented to demonstrate the substantial impact of the dynamic response on the elastohydrodynamic behaviour for both wavy and smooth surfaces. The materials, lubricant properties and operating parameters of the numerical solution presented in [25] are listed in Table 1 and are used in the current numerical solution. They developed a simple transient elastohydrodynamic point contact solver to solve the Reynolds' equation using Jacobi relaxation scheme for solving the pressure, and along with the half-space theory for solving the film thickness equation, which has been found to be suitable at low external loads and low speeds of entraining motion. At higher values of load and particularly moderate to high speeds of entraining motion, it is more appropriate to use Newton–Raphson technique with Gauss–Seidel iteration method. In the current analysis, Newton–Raphson technique with Gauss–Seidel iteration method has been used to solve the Reynolds' equation.

Tab. 1. Lubricant, material properties and operating parameters

Viscosity, η_0	0.0835 Pa s	Equivalent Young's modulus, *E′*	116.87×10^{9} Pa
Viscosity coefficient, *α*	1.95×10^{-8} Pa^{-1}	Load	20 N
ε	5.83×10^{-10}	Speed	1 m/s
ξ	1.68×10^{-9}	Ball radius, *R*	0.01 m
Surface roughness amplitude, *A*	0.024 µm	Wavelength of the ridge, *λ*	0.27 mm

Figure 1 shows the steady-state film thickness and pressure distribution for the smooth contacting surfaces along X-direction presented by the current numerical solution. A load of 20 N and rolling speed of 1 m/s result in the maximum Hertzian dry contact pressure of 0.51 GPa with 0.136 mm radius of Hertzian contact region. It can be observed that all familiar features of elastohydrodynamic point contact are displayed, where pressure distribution is almost semi-ellipsoid and the pressure spike occurs at the exit region as well as the well-known horse shoe shape in the film shape appears. The central film thickness is nearly flattened, and the minimum film thickness is found in the side lobes of the horse shoe shape. It is clear that, the pressure and film thickness presented in Figure 1 turn out to be fairly consistent with those presented in [25].

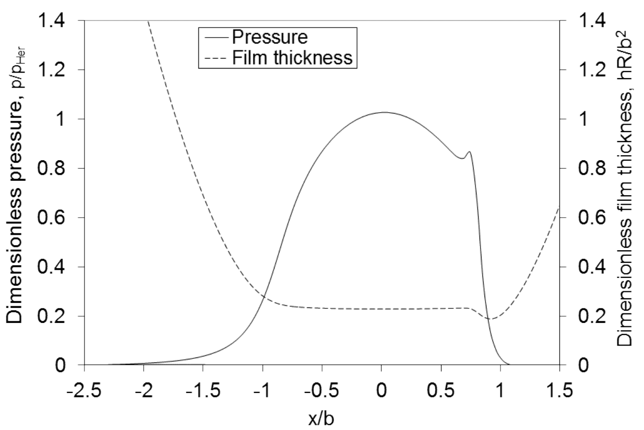


Fig. 1. Pressure and film thickness profile for the steady state of smooth surfaces

In order to investigate the effect of wavy surface for the steady-state elastohydrodynamic behaviors, a sinusoidal wavy surface with 0.024 µm amplitude and 0.27 mm wavelength is considered. Figure 2 shows the influence of waviness on the film thickness and pressure profile. It is clear that the pressure profile is no longer semi-ellipsoid in the contact area due to waviness and the film thickness profile is no longer flattened through the central contact region, in comparison to the case of smooth contact surface shown in Figure 1. The waviness on one contacting surface is almost completely deformed in the contact area, due to the high pressure in this contact area. The same conclusion was given in [25] for the same operating conditions.

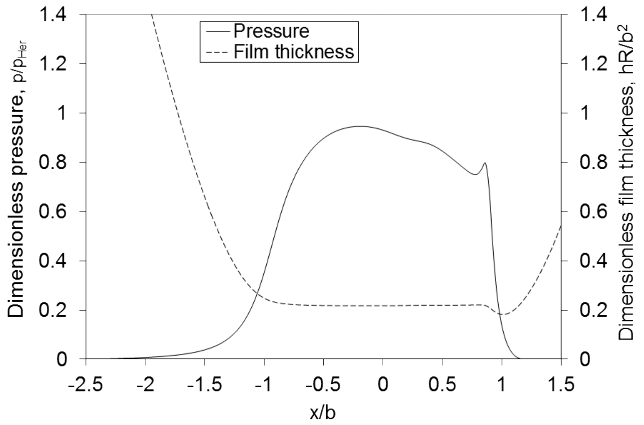


Fig. 2. Pressure and film thickness profile for the steady state of wavy surfaces

The effect of changes in the applied load due to external sinusoidal dynamic load on the pressure and film thickness profile is shown in Figure 3 for the case of smooth contacting surfaces. In fact, in the current analysis, it is assumed that the load varies between zero and a given value of 20 N in a sinusoidal fashion as expressed by:

$$W(t) = W_0 + a \times Sin(\omega t)$$

where, w_0 is the initial load and *a* is the amplitude of the sinusoidal load. The numerical solutions are carried out at *2π* dimensionless frequency excitation.

Figure 3 shows the pressure and film thickness profile in the central line of contact along the rolling direction at different snapshots of time. It is clear that propagation in the pressure and film thickness from the inlet towards the outlet of the contact due to dynamic load is observed, although the contacting surfaces are assumed to be perfectly smooth. From Figure 3, it is clear that the modulation in the pressure and film thickness profile is periodic since the external dynamic load applied to the contact changes sinusoidally. The film thickness and pressure profile at t = 2.5 s (see Figure 3 f) are coincident with those at t = 1.0 s (see Figure 3 c), and the film thickness and pressure profile at t = 3.0 s (see Figure 3 g) are coincident with those at t = 1.5 s (see Figure 3 d). It is obvious that initially, the pressure and film thickness profile exhibit the classic flat film shape through the whole contact zone, which characterises the steady-state film behaviour (see Figure 3 a). The modulation of pressure and film thickness profile induced by the harmonic applied load is shown in Figure 3 b – d. The modulation in film thickness and pressure profile due to the action of harmonic external applied load, which is shown in Figure 3 e – h, is coincident to that shown in Figure 3 a –d since the external dynamic load applied to the contact changes sinusoidally. These results conform well to the results presented by other authors like in [25].

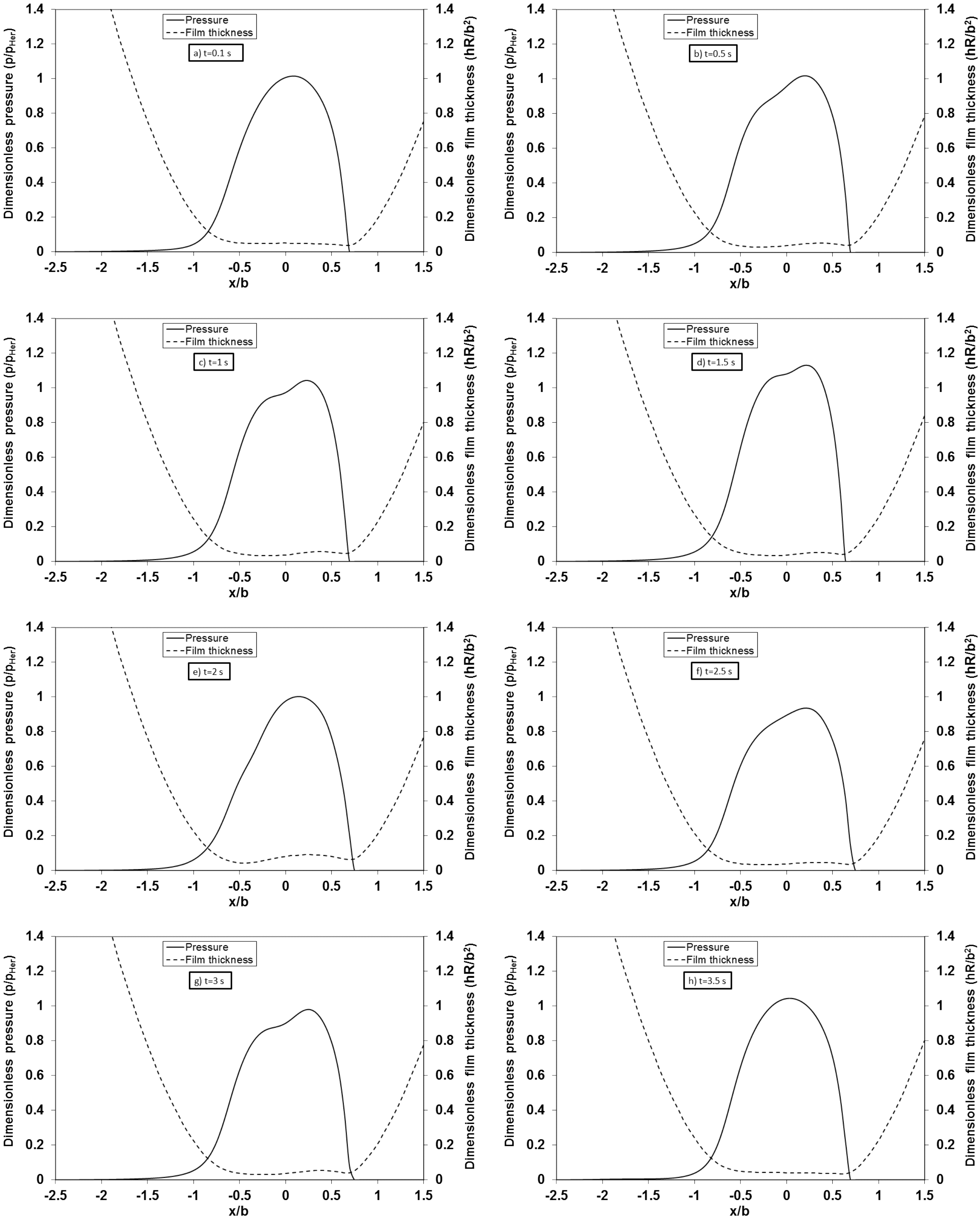


Fig. 3. Film shape and pressure profile in the central line of contact for smooth contacting surfaces

Finally, the combined effects of rough surface elastohydrodynamic lubrication contact subjected to the sinusoidal dynamic load are studied. In this case, a sinusoidal wavy surface and a sinusoidal dynamic load are considered to show their effects on pressure and film thickness profile. Figure 4 shows the changes in the central line pressure and film thickness profile along the X-direction at different snap-shots of time. It is clear that surface waviness causes fluctuations in pressure and film thickness within the nominal contact zone similar to the case of smooth contacting surfaces. Under these conditions, more evident fluctuations in pressure and film thickness profile are observed. However, due to the waviness onto one surface, the modulations of film thickness and pressure profile are no longer periodic as in the case of smooth contact surface. The same feature was found in the numerical solution presented in [25] for elastohydrodynamic lubrication of point contact problem.

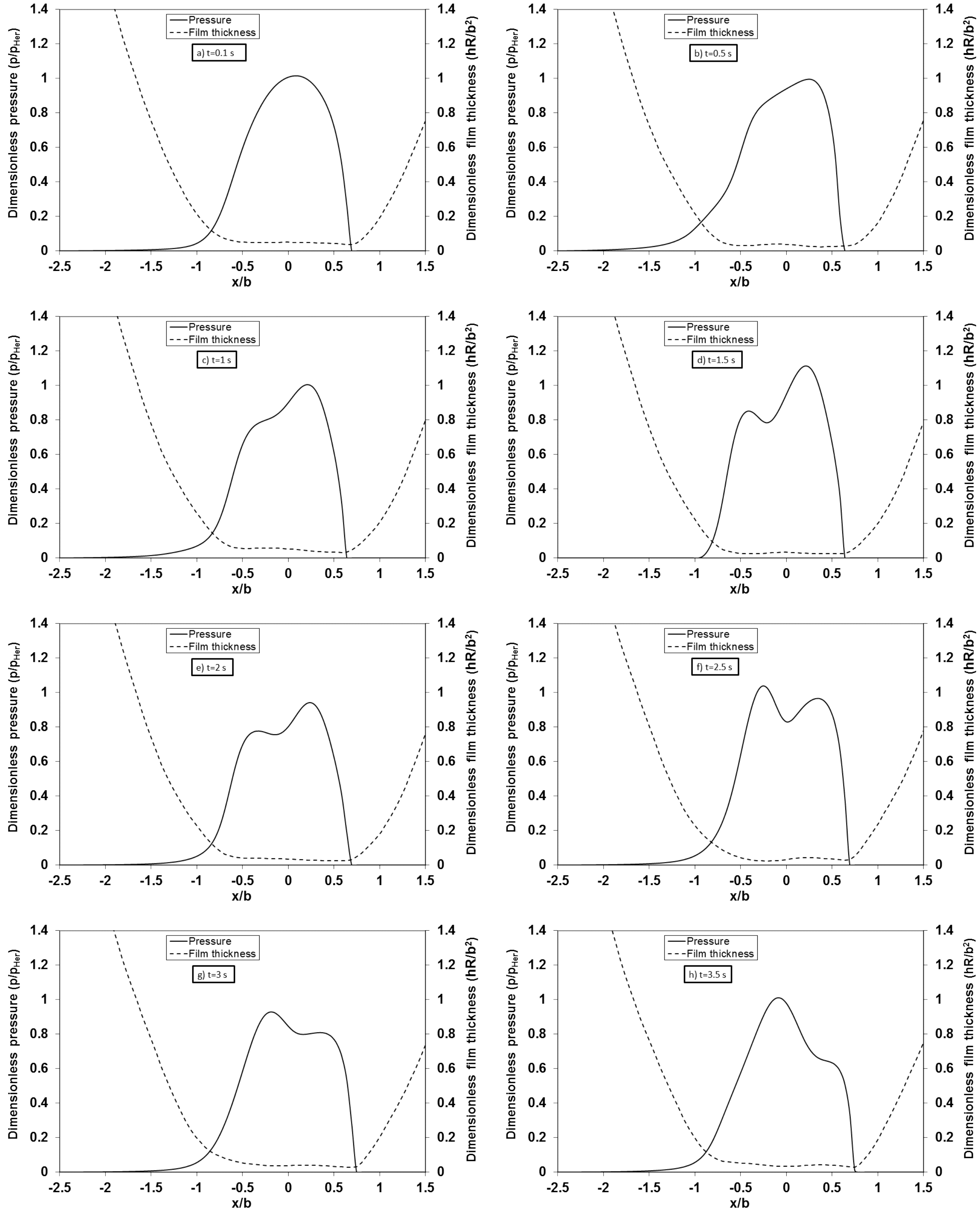


Fig. 4. Film shape and pressure profile in the central line of contact for wavy contacting surfaces

Unfortunately, there is no existing other previous numerical or experimental work in the area of elastohydrodynamic lubrication problem where the wavy surface contacts are subjected to harmonic forced vibrations to compare with. This is an area requiring more research because in real applications, contacts are subjected to variations in load, speed and geometry at the same time, and it is hoped that this investigation and others will trigger further work.

4. CONCLUSIONS

The response of the elastohydrodynamic lubricated contacts subjected to combined effects of harmonic forced vibrations and waviness of contacting surfaces is discussed in this paper using a numerical technique based on Newton–Raphson with Gauss–Seidel iteration method for concentrated counterformal point contact under isothermal conditions. The results of the current numerical solution are compared with the numerical work presented by other researchers for steady and transient problem conditions for smooth and wavy contacting surfaces. The numerical results showed that the external dynamic load induces modulations in the film thickness and pressure profile in the lubricated contact. The combined effects of contacting surface waviness and external dynamic load cause more pronounced modulations in the pressure and film thickness profile in comparison to the case of external dynamic load. This result is supported by the numerical work presented by other researchers which showed a good agreement between both sets of results for all loading and contacting surface conditions.

REFERENCES

1. Gohar R, Rahnejat H. Fundamentals of tribology. Imperial College Press, London; 2008.
2. Simon V. Optimal tooth modifications in face-hobbed spiral bevel gears to reduce the influence of misalignments on elastohydrodynamic lubrication. ASME J Mech. Des. 2014; 136(7):1-9.
3. Simon V. Improvements in the mixed elastohydrodynamic lubrication and in the efficiency of hypoid gears. Proceeding of the IMechE, Part J: Journal of Engineering Tribology. 2019; 324(6):795-810.
4. Wijnant YH, Venner CH, Larsson R, Ericsson P. Effects of structural vibrations on the film thickness in an EHL circular contact. Trans. ASME, J. Trib. 1999; 121(2):259-264.
5. Kilali TEl, Perret-Liaudet J., Mazuyer D. Experimental analysis of a high pressure lubricated contact under dynamic normal excitation force. Trans. Proc. Trib., Proc. Of the 30thLeeds-Lyon Symp. on Tribology. 2004:409-418.
6. Sakamoto M, Nishikawa H, Kaneta M. Behavikaor of Point Contact EHL Films Under Pulsating Loads. Trans. Proc. Trib., Proc. Of the 30thLeeds-Lyon Symp. on Tribology. 2004:391-399.
7. Kalogiannis K, Mares C, Glovnea RP, Ioannides E. Experimental investigation into the Response of Elastohydrodynamic Films to Harmonic Vibrations. International Journal of Mechatronics and Manufacturing Systems. 2011; 4(1):61-73.
8. Zhang X, Glovnea RP. The Behaviour of Lubricated EHD Contacts Subjected to Vibrations. IOP Conf. Ser.; Mater. Sci. Eng. 2017;174.
9. Glovnea RP, Zhang X, Sugimura J. The Effect of Lubricant Supply and Frequency upon the Behaviour of EHD Films Subjected to Vibrations. IOP Conf. Ser.; Mater. Sci. Eng. 2017;174.
10. Glovnea RP, Zhang X. Elastohydrodynamic Films under Periodic Load Variation: An Experimental and Theoretical Approach. Tribology Letters. 2018; 66(3):1-11.
11. Yang P, Cui J, Jin JM, Dowson D. A theoretical study on the response of a point elastohydrodynamic lubrication contact to a normal harmonic vibration under thermal and non-Newtonian conditions. Proc. IMechE, Part C. 2007; 221(9):1089-1110.
12. Morales-Espejel GE. Central film thickness in time-varying normal approach of rolling elastohydrodynamically lubricated contacts. Proc. IMechE, Part C. 2008; 222(7):1271-1280.
13. Felix-Quinonez A, Morales-Espejel GE. Film thickness fluctuations in time-varing normal loading of rolling elastohydrodynamically lubricated contacts. Proc. IMechE, Part C. 2010; 224(12): 2559-2567.
14. Al-Samieh MF. Numerical investigation of Elastohydrodynamic contacts subjected to harmonic load variation. Industrial Lubrication and Tribology. 2019; 71(6):832-841.
15. Venner CH, Lubrecht AA. Numerical simulation of a transverse ridge in a circular EHL contact, under rolling/sliding. Trans. ASME, J.Tribology. 1994; 116(4):751-761.
16. Holmes MJA, Evans HP, Hughes TG, Snidle RW. Transienelastohydrodynamic point contact analysis using a new coupled differential deflection method Part 1: Theory and validation. Proceeding of the IMechE, Part J: Journal of Engineering Tribology. 2003; 217(4): 289-303.
17. Glovnea RP, Choo JW, Olver AV, Spikes HA. Compression of a single transverse ridge in a circular elastohydrodynamic contact. ASME Journal of Tribology. 2003; 125(2):275-282.
18. Armando FQ, Pascal E, Jonathan LS. New experimental results of a single ridge passing through an EHL conjunction. ASME Journal of Tribology. 2003; 125(2):252-259.
19. Felix-Quinonez A, Ehret P, Summers JL. Numerical analysis of experimental observations of a single transverse ridge passing through an elastohydrodynamic lubrication point contact under rolling/sliding conditions. Proceeding of the IMechE, Part J:Journal of Engineering Tribology. 2004; 218(2):109-123.
20. Felix-Quinonez A, Ehret P, Summers JL, Morales-Espejel GE. Fourier analysis of a single transverse ridge passing through an elastohydrodynamically lubricated rolling contact: a comparison with experiment. Proceeding of the IMechE, Part J: Journal of Engineering Tribology. 2004; 218(1):33-43.
21. Ildiko F, Sperka P, Hartl M. Transient calculations in elastohydrodynamically lubricated point contacts. Engineering Mechanics. 2014; 21(5):311–319.
22. Sperka P., Krupka I. and Hartl M. Rapid prediction of roughness effects in sliding EHL contacts. STLE Annual Meeting & Exhibition, Michigan, USA. 2013; May 5-9.
23. Sperka P, Krupka I, Hartl M. Prediction of Shallow Indentation Effects in a Rolling-Sliding EHL Contact Based on Amplitude Attenuation Theory. Japanese Society of Tribologists. 2017; 12(1):1-7.
24. Al-Samieh MF. Effect of changing geometrical characteristics for different shapes of a single ridge passing through elastohydrodynamic of point contacts. Industrial Lubrication and Tribology. 2021; 73(2):283-296.
25. Cupu DRP, Stratmann A, Jacobs G. Analysis of transient elastohydrodynamic lubrication of point contact subjected to sinusoidal dynamic loads. International Conference on Design, Energy, Materials and Manufacture, IOP Conf. Series: Materials Science and Engineering 539. 2019.
26. Al-Samieh MF. Theoretical Investigation of Transient Ultra-Thin Lubricant Film during rapid deceleration. Tribology in industry. 2018; 40(3): 349-357.
27. Johnson KL. Contact mechanics. Cambridge University Press. 1985; Chapter 3: 54.
28. Dowson D, Higginson GR. A numerical solution to the elastohydrodynamic problem. J. Mech. Engng. Sci. 1959; 1(1):6-15.
29. Roelands CJA. Correlation aspects of viscosity-temperature-pressure relationship of lubricating oils. PhD thesis. Delft University of Technology, The Netherlands. 1966.

Mohamed F. Abd Alsamieh: https://orcid.org/0000-0001-9308-8897

DOI 10.2478/ama-2022-0020

Nomenclature

A	Dimensionless surface roughness amplitude
B	Radius of Hertzian contact region
E'	Equivalent Young's modulus
G^*	Materials' parameter, $G^* = E'\alpha$
h	Lubricant film thickness
H	Dimensionless film thickness, $H = hR/b^2$
H_0	Dimensionless central film thickness
p	Pressure
P	Dimensionless pressure, $P = p/P_{Her}$
N, M	Total number of mesh points in X- and Y- directions
R	Radius of contact
W	Normal applied load
X, Y	Dimensionless coordinates, $X = x/b$, $Y = y/b$
u	Rolling speed, $u = (u_A + u_B)/2$
U^*	Speed parameter, $U^* = u\eta_0/E'R_2$
Z	Viscosity-pressure index
α	Viscosity coefficient
δ	Total elastic deformation
η	Lubricant viscosity
η_0	Atmospheric lubricant viscosity
ρ	Lubricant density
ρ_o	Atmospheric lubricant density
$\overline{\rho}$	Dimensionless lubricant density, $\overline{\rho} = \rho / \rho_0$
$\overline{\eta}$	Dimensionless lubricant viscosity, $\overline{\eta} = \eta / \eta_0$
λ	Dimensionless wavelength of the ridge
$S.R$	Surface roughness

Superscripts

i,j	Contravariant influence coefficient indices
n	Iteration index
k,l	Covariant influence coefficient indices