Acta Universitatis Sapientiae

Electrical and Mechanical Engineering

Volume 15, 2023

Sapientia Hungarian University of Transylvania
Scientia Publishing House

Contents

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 1–13

DOI: 10.2478/auseme-2023-0001

IoT Sensor Network Solution for Monitoring Wind Induced Waves in Shallow Water

Áron István JÁSZBERÉNYI

Department of Telecommunications and Media Informatics, Faculty of Electrical Engineering and Informatics, Budapest University of Technology and Economics, Budapest,
e-mail: jaszberenyi.aron@edu.bme.hu

Manuscript received June 7, 2023; revised June 21, 2023

Abstract: This case study presents a possible solution for monitoring wind induced waves in shallow water lakes. Currently widely used instruments for measuring water waves are expensive due to extreme environmental requirements and the technology used. The use of IoT solutions is not common. Water wave measurements require low power consumption and medium to long range communication. Due to the network specifications, it is not obvious that LPWAN technologies are suitable for transmitting the amount of data generated. For the devices developed, LoRaWAN was investigated as the cheapest and 5G as the most reliable network solution, with attention to energy consumption.

Keywords: IoT, WSN, LPWAN, LoRaWAN, 5G, Limnology, Wave monitoring

1. Introduction

Water is an important natural resource. The increasing global population and climate change are putting pressure on the availability and quality of water resources. Field observation is essential for monitoring and evaluating the effectiveness of environmental protection measures, allowing us to identify potential threats and develop strategies to mitigate them. Lake monitoring takes place in extreme environments (e.g. hurricane-force storms and extreme water levels). Information on water quality, temperature and other important factors affecting the ecosystem is crucial for understanding how lakes respond to environmental changes and for developing effective management strategies.

Many of the activities required in lake management can only be accomplished if the lake's current state is monitored on a regular basis. Long-term data can also be utilized to improve forecasts. This is why automated monitoring systems are required to provide real-time data that can be used to make appropriate lake management decisions.

Wind-generated waves have a significant limnological impact. They affect the water quality, sediment transport, and the distribution of aquatic organisms in lakes and reservoirs. The study of wind-generated waves is important for understanding the dynamics of freshwater ecosystems. The monitoring systems must be effective in monitoring and operate for a long period of time without causing harm to the environment or depleting resources. Therefore, it is important to consider the use of renewable energy sources and low-cost materials in the design and implementation of these systems.

Due to the high expense of equipment, it is difficult to conduct simultaneous lake surveys at multiple locations at the same time. This is the reason why there are currently no operational measurements of the waves on Lake Balaton, Hungary's largest lake, only intermittent measurements at numerous sites for research purposes.

2. Wireless Sensor Network and Internet of Things

The extreme environment of the lakes requires on-board power supply and wireless connectivity. The Internet of Things (IoT) provides numerous choices to create low-cost sensor networks, with low energy consumption. Low cost means lower measurement accuracy, but with more measurement nodes the accuracy can be increased. It is better to have several simpler nodes, as a larger area can be monitored with greater overall accuracy, and data from additional devices can be used in the event of a node failure.

Low Power Wide Area Network (LPWAN) technologies provide low-cost but effective long-range connectivity for Wireless Sensor Networks (WSN), making it ideal for IoT devices that need to transmit small amounts of data over long distances. LPWAN networks are designed to be scalable and capable of supporting a large number of devices. The LPWAN technologies sacrifices the data rate and the Quality of Service (QoS) for the low cost and low energy consumption. Communication can take place in the licensed frequency bands or in the unlicensed Industrial Scientific and Medical (ISM) bands. For ISM bands, the duty cycle must also be taken into account. This defines the percentage of time that active communication between devices is enabled [1].

A. ISM based LPWAN technologies

The two most popular LPWAN technologies, that use ISM bands are SigFox and LoRaWAN. In Europe, Sigfox operates on the 868 MHz ISM band, while LoRaWAN operates on the 443 MHz or 868 MHz ISM bands. The applicable duty cycle is 1% [2].

SigFox is available only through one operator. The maximum amount of data that can be sent per day is 140 · 12 bytes. The data rate is 100 bps [3].

LoRaWAN is available through service providers or by self-installation. Service providers often enforce air fairness rules that apply stricter communication rules than the regional standard (e.g. daily message limits). LoRaWAN uses adaptive data rate. The maximal achievable data rate is 5470 bps, which allows 590.76 kB of data to be transmitted per day [4].

B. Cellular LPWAN technologies

In the cellular networks the two available LPWAN solutions are the Long Term Evolution Category M1 (LTE Cat-M1) and NarrowBand IoT (NB-IoT). Both LTE Cat-M1 and NB-IoT offer low power consumption, extended coverage, and support for a large number of connected devices, making them ideal for IoT applications. LTE Cat-M1 has higher data rates and is better suited for applications that require real-time communication, whereas NB-IoT is better suited for applications that require low data rates and long battery life. LTE Cat-M1 and NB-IoT has theoretically no regulations, but service providers may impose their own restrictions or limitations on the use of these technologies. These technologies are available only through operators [5], [6].

C. 5G possibilities

The 5th Generation Cellular Mobile Network, also known as 5G, is the latest and most advanced form of wireless communication technology that promises faster internet speeds, lower latency, and greater connectivity. The design of 5G has been heavily focused on industrial use cases, such as sensor networks and critical IoT applications. LTE Cat-M1 and NB-IoT standards are still available in 5G, as a kind of extension. Offering seamless and future-proof evolution combined with dynamic spectrum sharing. The 3GPP Rel. 16 enables the connection of compliant LTE Cat-M1/NB-IoT devices to the next generation 5G core network. The implementation of the new specifications could be a challenge for the cost-sensitive cellular mobile IoT market, not only due to increased complexity but also due to the fragmented nature of the market [7].

3. Wave measurement

There are several ways to measure water waves in shallow lakes. The most basic measurement technique is the visual observation by an experienced observer. There are so-called in situ mechanical measuring devices, such as the VITUKI type wave tracker, which uses a floating element on the water surface, and there are also water pressure-based measuring devices [8]. The other major

group of measurements is remote sensing. This includes all methods where the measuring device is deployed above the water surface and takes measurements on a fixed or moving platform. With today's technology, visual wave analysis can also record increasingly better results using cameras [9].

In addition to the measurement technique, the placement of the instruments can also vary. The most common method of wave measurement is to measure wave properties at a single point over a long period of time (several weeks of measurements). Multi-point measurements allow analysis of a larger area but are rarely used because of the need for precise time synchronisation between measurement nodes and the investment required for a single instrument.

Due to the extreme environment the instruments must be waterproof and resistant to environmental impacts. For this reason, the instruments on the market are very expensive. Therefore, often few instruments are installed, and sampling is done at a single point. Professional water surface measuring devices currently available on the market cost approximately 25 000 – 30 000 EUR.

4. Hardware architecture

A. Challenges in wave sensing

The tools developed had to meet a number of requirements. They had to remain functional in extreme environmental conditions, while being simple and inexpensive to produce, as they compose an IoT sensor network. Devices should be able to operate for as long as possible without intervention, while running on battery power. Measurement data should be transmitted using LPWAN technology to provide quasi real-time information.

The limonological modelling of lakes has slightly different requirements than the average IoT LPWAN sensor networks in the traditional sense. To process the measurements, the sampling density in the measurement cycles should be around 8 Hz in Hungarian lakes to reconstruct the waveform according to the Nyquist-Shannon sampling theorem. The number of samples collected during the measurement cycle should be ideally a power of two for the fast Fourier transform, but at least 1024 samples, as this provides a sufficient number of waves (at least 100 recommended) for the calculation of statistics. In order to test the lake surface as regularly as possible, the more frequent the measurement cycles, the better. The sampling frequency for this study is 20 minutes per measurement cycle. The amount of data generated by such a large number of samples uses only a negligible fraction of the Internet bandwidth but can be very demanding for a LPWAN network [9], [10].

There are many communication solutions for IoT WSN networks. The water monitoring system in question performs 72 measurement cycles per day,

generating 290 kB of data per day for a minimum of 1024 measurements per cycle. This amount of data is unusually large for a LPWAN network. Most operators have radio usage restrictions to ensure that all user devices can be served by the network (Fair Access Policy). This is stricter than the duty cycle time in general, which makes SigFox, some LoRaWAN and NB-IoT services unsuitable for transmitting measurements from the water monitoring network nodes. Among the LPWAN technologies discussed, LTE Cat-M1 seems to be the only feasible one. However, with data size reduction, private LoRaWAN can offer a cheaper solution in long-term, as only the deployment is costly.

5G LPWAN technologies are not yet available for general use, but the 5G network places a strong emphasis on machine-to-machine communication and quality of service, so although radio modules are currently expensive, their cost could be reduced as the technology becomes more widespread. Reduced Capability (RedCap) modules (non-LPWAN solutions) are already available, where reduced capability also means lower cost [11]. In the future, the LPWAN integrations mentioned in *1/C* may also become available in 5G networks.

B. Hardware architecture

At the heart of the measurement nodes is an ESP32 System on a Chip (SoC) based microcontroller unit (MCU). Various measurement submodules and operational support devices are connected to this: DS18B20 waterproof thermometer module, JSN-SR04T waterproof ultrasonic distance sensor module, DS3231 real time clock (RTC) module, CNC136X110-6 solar panel + TP4056 charge controller, and NCR18650B Lithium-ion battery. The MCU has integrated LoRa and Wi-Fi radio module, but no cellular capabilities. *Table 1.* shows the submodules and their specifications.

The devices are located on a printed circuit board made for the module, in a waterproof box. The system is designed to deliver multi-point remote sensing. The measuring nodes are mounted above the water level and the distance sensor is positioned vertically to the water surface. *Fig. 1* shows a module in the field.

Table 1: Used submodules

Name	Type	Features	Price
MCU with LoRa capability	LILYGO® LORA32 V2.0 868Mhz LoRa ESP32 OLED	ESP chip: ESP32 PICO-D4; Flash: 4 MB; MicroSD Card slot;	~40 EUR
Thermometer (Waterproofed)	DS18B20	Range: -55 °C to +125 °C; Accuracy: ±0.5 °C;	~3 EUR
Ultrasonic Waterproof Range Finder	JSN-SR04T-2.0	Range: 20–600 cm; Angle: 75°; Accuracy: ±1 cm; Resolution: 1 mm;	~6 EUR
Real Time Clock (RTC)	DS3231	Accuracy: ±2 ppm from 0 °C to +40 °C, ±3.5 ppm from -40 °C to +85 °C; Backup-Battery	~4.5 EUR
Solar panel	CNC136X110-6	Size: 136 x 110mm; Work voltage: 6V; Max power: 2W	~12 EUR
Charge controller	TP4056	Input supply voltage: 4.5~6.0 V; Overcurrent Protection	~1 EUR
Battery	NCR18650B	Capacity: 3400 mAh	~9 EUR

Figure 1: Field deployed measurement node

The communication objective of the study is to achieve low energy consumption with Cat-M1 alternatives. The two feasible options are LoRaWAN network and 5G if low power consumption is achievable. In the lack of an easily integratable 5G radio module, a Wistron NeWeb Corporation (WNC) 5G mobile hotspot has been used for communication.

The LoRaWAN network of the study is self-hosted, based on an open-source ChirpStack Network Server [12]. The 5G network was a Non-Standalone

(NSA) test campus network provided by Ericsson Hungary and Budapest University of Technology and Economics [13].

The data is collected in a PostgreSQL database accessible via the Internet [14]. The data is processed using Matlab Online [15].

5. Description of the network operation

A. Data structure

The wave analysis is performed every 20 minutes, with 2100 measurements in one measurement cycle at 8 Hz sampling. During measurement processing, the number of samples should ideally be a power of 2, with a few extra measurements for error filtering, so more than 2048 measurements are performed per cycle. As sampling is very dense, there is no time to send the measurement results until after the measurement cycle has been completed. The measurement nodes transmit the start time of the measurements, the air temperature and the water surface distances in a chain.

B. Node sync

Since sampling takes place simultaneously at several points in the lake, it is important that the measuring nodes take measurements at the same time. Thanks to the RTC module, this synchronization can be achieved.

Node consumption can be minimized by using the sleep mode of ESP32. Using the RTC module, it is possible to dynamically set how long the devices should be in sleep mode after measurement and data transmission. Thanks to the dynamic sleep timing implementation, device idle time is minimized.

C. Measurement accuracy

The measurement accuracy of the distance sensors should be maximized to recover the waveform as accurately as possible. The rangefinder measures the distance travelled by the ultrasound emitted and returned from the water surface as a function of the sound propagation time. For more accurate measurement results, the sound propagation time can be compensated based on the air temperature, taking into account the measurement data of thermometers. The calculation is shown in (1), where T is the air temperature and v is the sound velocity [16]. The accuracy is increased by 1% for a 5.5 °C change.

$$v\left[\frac{m}{s}\right] = 331\left[\frac{m}{s}\right] + 0{,}6\left[\frac{m/s}{^\circ C}\right] \cdot T\,[^\circ C] \tag{1}$$

The temperature sensor cannot be placed in shade completely protected from the sun, and will return a reading of -127 if damaged. In order to safeguard the distance measurements, for thresholds below -20 °C and above +40 °C, the average temperature of 11 °C in Hungary is substituted into (1).

The ESP32 SoC systems have a big advantage over their similar purpose competitors: they have two processor cores in most cases, which can be used for optimization. The real-time operating system of the microcontroller is designed to run on a single core. The ESP32 in use contains a protocol CPU and an application CPU. The two cores have technically the same architecture and use the same memory. This allows to run tasks alternately between the two cores [17]. To refine the measurement sampling frequency, both cores of the microcontroller were used in the implementation as follows. While the application CPU sets the trigger times and waits for the ultrasound to return to the sensor, the protocol CPU calculates when the next sampling can start. *Fig. 2.* shows the usage of the two CPU cores. Measurements were performed in two ways for comparison purposes using single-core and dual-core operations and sampling every second. The results showed that the variance from the exact timing was around 0.300 seconds for the single-core case. For dual-core use, the same variance was only 0.009 seconds. In the dual-core case, the sampling time accuracy of the devices can be increased significantly.

Despite the different measurement principle, the difference between the results of the reference DATAQUA pressure-based water level meter and the developed measurement modules are within the order of magnitude of the accuracy of the height reference measurement [18].

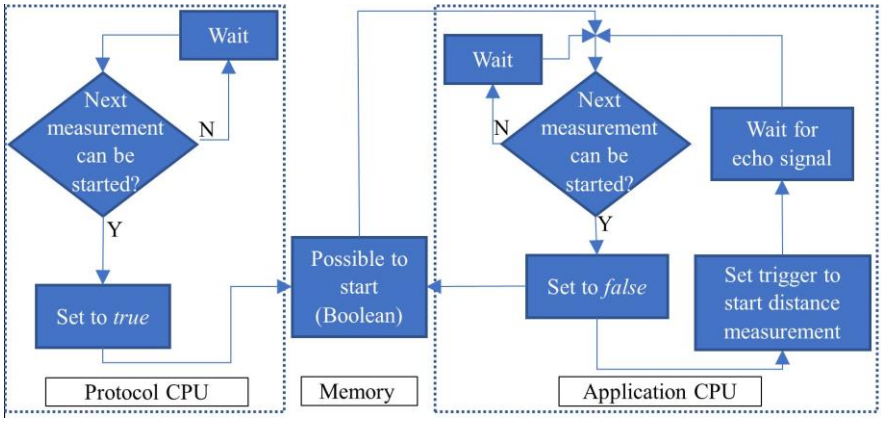


Figure 2: Main tasks during the dual-core measurement

D. Data transmission

The measurement nodes try to transmit the measured data over 2 communication channels: 5G and LoRaWAN. With 5G, data is guaranteed to arrive, while with LoRaWAN the communication distance is greater. The primary communication channel is 5G. If the connection fails, LoRaWAN provides the secondary communication network. In the absence of a 5G radio module that can be easily integrated with the microcontroller, 5G communication was achieved using a mobile hotspot.

The data is stored in a PostgreSQL database running in the cloud. In case of 5G, data can be sent all at once, but for LoRaWAN, measurement results can only be sent in several frames, as the allowed packet size is very small. The spreading factor (SF) is the rate at which the frequency of a signal changes throughout the bandwidth of a channel. Lower LoRa spreading factor provides a higher bit rate for a fixed bandwidth and coding rate. The highest bitrate and packet size is available with SF7. The maximum payload size is 222 bytes. The duty cycle rules must be applied between consecutive packet transmissions [4].

The most significant part of the data to be sent is the distance measurement results. The measured data is stored in approximately 4 bytes, depending on the magnitude of the measured distance. Distance records are concatenated using a separator character, so each measurement represents 5 bytes of data. For 2100 measurements, approximately 10 kB of data are generated. In a measurement cycle, it takes nearly 4.5 minutes to initialise the equipment and perform the measurements. In the remaining time, using LoRa SF7, around 6 kB of data can be sent due to the duty cycle control. Data size can be reduced by converting measurement results to hexadecimal numbers. Since no distance results greater than 4095 mm are expected, 3 B is sufficient to store a single measurement result. If the results are fixed at 3 B, the separator character can be omitted and it is possible to send 2100 measurement results within the measurement cycle.

With Matlab online, data stored in the database can be read and processed automatically using a web browser. The state of the wave in a stationary record can be characterized by the average wave parameters, which can be determined from the recorded surface distance data. The Pierson-Moskowitz or JONSWAP spectrum can be used as a model of the evolved wave. The data processing is beyond the scope of this paper and will not be described in detail.

6. Effectiveness

Two communication solutions were evaluated along the following aspects: network quality and energy consumption.

A. Network

For LoRaWAN, the spreading factor used is SF7 and the frequency range is 868 MHz. The operating class is A, so communication is only one-way, to save energy. Acknowledgement of sent data is disabled (no retransmission if a packet is lost). The coding rate (CR) used for error correction is 4/5. The data rate is 5470 bps. The communication distance is approximately 2 km [4].

For 5G NSA, the N78 frequency band was used. The network was tested using several WNC hotspots. Connected to the outdoor radio units, measurements were made at 3 locations with good (-70 dBm), poor (-106 dBm) and bad (-123 dBm) signal strength. The upload speed decreases steadily as the distance from the radio unit towards the cell boundary increases. The upload speed was 24.95 Mbps at good and 9.42 Mbps at poor signal strength. Response times were measured by sending 150 ping messages. The minimum response time was 15 ms and the average response time was 30 ms for both good and poor signal strength. Under bad signal strength, the response time doubled. In average, the latency variation was only greater than a few milliseconds under poor signal strength. Packet loss did not occur even at bad signal strength.

5G communication provides a reliable data connection, but in safety-critical cases, network problems must be prepared for. In the present study, redundancy is provided by the unified presence of 5G and LoRaWAN networks. The devices will try to send data over the 5G network, but in case of a failed connection, they will send the data as LoRaWAN messages. 5G is the primary channel, as it provides the acknowledgement of incoming packets, so all measurement results are safely stored.

B. Power consumption

The 5G hotspot is a separate device from the monitoring stations. The power consumption of the radio module is recorded separately from the overall energy usage of the device. This gives a more accurate picture of how the consumption of the monitoring stations would be affected if the cellular radio module were integrated. *Table 2*. shows the static power consumption of the MCU and connected modules in different states.

Table 2: Average power consumption

Module	Power consumption
Microcontroller – Hibernation mode	9.5 mA
Microcontroller – Idle state	47.6 mA
Microcontroller – Single core calculation	48.8 mA
Microcontroller – Dual core calculation	52.9 mA
Thermometer	0 mA
RTC module	2.5 mA
Distance sensor – Idle state	3.1 mA

When using 1 processor core, the consumption is stable. When using 2 processor cores, the energy consumption fluctuates, but the average consumption increases only slightly (+8%).

Using the distance sensor, the average consumption of the measuring nodes is 77.7 mA. The peak consumption is 96.4 mA when measurements are taken.

For LoRa communication, radio use statically increases the current consumption by 61.1 mA.

The current consumption of the 5G radio module in the idle state with no connection is 4.4 mA. With active 5G communication, the average increase in consumption was 98 mA for strong and 281 mA for weak signal strength.

The power consumption in a cycle is around 25 mA/20 min in both cases.

B. Energy generation

The measurement modules each have a solar panel connected to a charge control module. Several solar panels were measured in the first week of December 2022. On a lightly cloudy day, when the solar panel was oriented to direct sunlight, the maximum output for the most efficient solar panel was 134.8 mA. On a heavily cloudy day, when sunlight only reached the solar panel through the interconnected clouds, the maximum energy production value for the least efficient solar panel was 17.6 mA. Half an hour before sunset, when the panel was no longer exposed to direct sunlight, the production was 8.4 mA for the most efficient solar panel in the lightly cloudy case and 3.2 mA in the heavily cloudy case. It can be concluded that during periods of high cloudiness the efficiency of photovoltaic charging is drastically reduced.

In the summer months, there can be up to five times the number of hours of sunshine as in December. Summer radiation levels are more than six times higher than in January and December [19]. The operating time of devices can be greatly increased in summer, as most of the time the solar panel produces more than the average consumption of the devices. During the winter months, the batteries of the nodes may need to be replaced and charged from time to time.

7. Conclusion

Operational lake monitoring with multi-point sampling is rare due to the high cost of the equipment. Since it is not possible to prevent the theft of the devices, it is important that they can be replaced as cheaply as possible. The unit price of the developed measuring modules is more than 250 times cheaper than commercially available equipment. Data transmission is key for monitoring. LoRaWAN is the most cost-effective LPWAN solution in this case, as the network can be operated autonomously and uses ISM bands. Given the amount of data that will be generated, it is also worth considering a standard cellular network solution taking into account the consumption. 5G is becoming more widespread, which will also reduce the cost of devices.

Wave measurement has a number of limitations and difficulties. With the self-built IoT measurement system, the desired measurement accuracy was achieved. The data size can be minimized to allow LoRaWAN transmission of the data. Energy consumption was recorded for LoRaWAN and 5G. LPWAN offers a low-cost solution, but in this use case it does not provide energy savings due to the large amount of data to be transported. With the price reduction, 5G could be a good solution.

With wind detection, it may be possible in the future to allow measurement modules to measure only when the wind waves to be monitored are present. This would reduce the amount of non-useful data and reduce the consumption of the measuring instruments. By integrating and properly positioning 5 distance sensors per measurement module, the wave components could be reconstructed even more accurately.

Acknowledgements

This publication is supported by Budapest University of Technology and Economics, and by Ericsson Hungary Ltd. Special thanks to Tibor Cinkler, Tamás Krámer, Csaba Simon and Gergely Seres for their support. The research was supported by the European Union project RRF-2.3.1-21-2022-00004 within the framework of the Artificial Intelligence National Laboratory.

References

[1] Muteba, F., Djouani, K., & Olwal, T., “A comparative survey study on LPWA IOT Technologies: Design, Considerations, challenges and solutions”, *Procedia Computer Science, 155*, 2019, pp. 636–641. https://doi.org/10.1016/j.procs.2019.08.090.

[2] European Telecommunications Standards Institute. (2012), “Final draft ETSI EN 300 220-1 V2.4.1 (2012-01)”.

[3] Sigfox SA. (n.d.). *Sigfox*. Sigfox build. Retrieved May 7, 2023, from https://build.sigfox.com/study.
[4] Semtech Corporation. (2019). (tech.), "LoRa and LoRaWAN: A Technical Overview".
[5] Haltian. (2019, January 30). LTE Cat M1 Overview. Retrieved May 7, 2023, from https://haltian.com/resource/lte-cat-m1-overview/
[6] Haltian. (2019, January 30). NB-IOT Overview. Retrieved May 7, 2023, from https://haltian.com/resource/nb-iot-overview/
[7] Zaidi, A., Bränneby, A., Nazari, A., Marie, M., & Kuhlins, C., „Cellular IoT in the 5G era", Ericsson, 2020, GFMC-20:000025 Uen.
[8] Muszkalay, L., „A Balaton vizének jellemző mozgásai", Budapest: Vízgazdálkodási Tudományos Kutató Intézet, 1973 [The typical movments of the water in Lake Balaton].
[9] Homoródi, K., "Investigation and modelling of wind induced waves in shallow water (dissertation)", 2012.
[10] Holthuijsen, L. H., "Waves in Oceanic and Coastal Waters", *Cambridge University Press*, 2015.
[11] Wang, Y.-P. E., Narayanan, S., Bergman, J., & Höglund, A., "What is reduced capability (RedCap) NR and what will it achieve?", [web log, 2021, May 11], Retrieved May 7, 2023, from https://www.ericsson.com/en/blog/2021/2/reduced-cap-nr.
[12] ChirpStack, open-source lorawan® network server. ChirpStack. (n.d.). Retrieved May 7, 2023, from https://www.chirpstack.io/
[13] „Mindkét Félre Szükség Volt a jövőt Formáló Innovációk megvalósításához", Budapest University of Technology and Economics. (2022, July 14). Retrieved May 7, 2023, from https://www.bme.hu/hirek/20220714/Mindket_felre_szukseg_volt_a_jovot_formalo_innovaciok_megvalositasahoz [Both parties were needed to deliver Future Shaping Innovations]
[14] PostgreSQL. (n.d.). Retrieved May 7, 2023, from https://www.postgresql.org/
[15] MATLAB online. MATLAB & Simulink. (n.d.). Retrieved May 7, 2023, from https://www.mathworks.com/products/matlab-online.html
[16] Shrivastava, A. K., Verma, A., & Singh, S. P., "Distance measurement of an object or obstacleby ultrasound sensors using P89C51RD2", *International Journal of Computer Theory and Engineering*, *2*, pp. 64–68, https://doi.org/10.7763/ijcte.2010.v2.118
[17] FreeRTOS (ESP-IDF). ESP. (n.d.). Retrieved May 7, 2023, from https://docs.espressif.com/projects/esp-idf/en/latest/esp32/api-reference/system/freertos_idf.html
[18] Dataqua Electronics Ltd. (n.d.). "Analogue and digital transmitters, recorders, monitoring", *Dataqua Electronics Ltd.*, Retrieved May 7, 2023, from https://www.dataqua.hu/index.php?lang=en
[19] Hungarian Meteorological Service. (n.d.), "Solar radiation, sunshine duration and cloud cover of Hungary", *Országos Meteorológiai Szolgálat*, Retrieved May 7, 2023, from https://www.met.hu/en/omsz/palyazatok_projektek/hungairy/

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 14–32

DOI: 10.2478/auseme-2023-0002

A Review of Using Visual Odometery Methods in Autonomous UAV Navigation in GPS-Denied Environment

Hussam M. ROSTUM[1], József VÁSÁRHELYI[2]

[1] Institute of Automation and Info-Communication, Faculty of Mechanical engineering and Informatics, University of Miskolc, Miskolc, e-mail: hussam.rostum@uni-miskolc.hu
[2] Institute of Automation and Info-communication, Faculty of Mechanical engineering and Informatics, University of Miskolc, Miskolc, e-mail: jozsef.vasarhelyi@uni-miskolc.hu

Manuscript received November 21, 2023, revised December 10, 2023

Abstract: This review paper centers on strategies employed for location determination in regions lacking GPS signals. It primarily explores a range of vision-based methods tailored for this purpose, categorizing them accordingly. The article delves into the utilization of optical flow for feature extraction-based Visual Odometry (VO) and delves into advanced optical flow estimation methods that hinge on deep learning techniques. It compares the efficacy and practical applications of frequently utilized visual localization methods while also checking the efficiency of previous researches by reapplying the algorithms to new data and comparing the results.

Keywords: Visual odometry, Optical flow, Deep learning Methods, GPS-denied localization.

1. Introduction

According to [1], autonomous navigation systems must possess the capability to estimate, perceive, and comprehend their surroundings in order to accomplish tasks such as path tracking, motion planning, obstacle avoidance, and target detection. GNSS, or Global Navigation Satellite Systems, furnish dependable environmental data and instantaneous positioning for self-governing vehicles such as Autonomous Vehicles (AVs) and Unmanned Aerial Vehicles (UAVs). In scenarios where Unmanned Aerial Vehicles (UAVs) are operating in congested and complex environments, GNSS signals may encounter issues like fading, multipath effects, jamming, and spoofing, which can result in signal loss. Hence, GNSS may not be a reliable solution for UAVs flying at lower altitudes in terrains marked by numerous obstructions, such as forests, cities, and canyons/mountains

[2]. To tackle this problem, over the recent years, scientists have devised various localization approaches specifically for UAVs, including vision-based and LIDAR-based techniques. Incorporating multiple sensors in robotic localization studies can be an expensive undertaking, and it may also add extra weight and power consumption to the device. The challenges mentioned above have impeded the widespread implementation of several robot localization algorithms [3]. As an alternative, visual localization, a computer vision-based technology, has emerged as an appealing option. The operating principle of visual localization is based on capturing images of the surrounding environment using a visual camera, followed by determining the position and orientation of the area around it, thereby producing a map of the unknown territory. One of the major advantages of this technology is that the camera's cost is relatively low, and it has the ability to capture extensive environmental data, including visual aspects such as color and texture. Nonetheless, visual localization demands high computational power, and images require a significant amount of storage space, and software development for visual localization is relatively challenging. Moreover, the visual system is susceptible to lighting conditions, and it may not function effectively in poorly lit environments. At present, there are two primary vision-based localization methods, namely Relative Visual Localization (RVL) and Absolute Visual Localization (AVL). RVL encompasses Visual Odometry (VO) and Visual Simultaneous Localization and Mapping (VSLAM). Visual Odometry (VO) is a technique for calculating the self-motion of a robot, utilizing monocular or binocular cameras. This literature review summarizes the current state of research and challenges facing Visual Odometry localization technologies in environments where GPS is unavailable. The primary contributions are summarized in the following areas:

- The research introduces the working principle of optical flow and its algorithm, as well as the broad applications of optical flow in visual odometry. The main focus of this review is on the introduction of FlowNet and its subsequent improved algorithms, and it includes a comparison to guide the selection of optical flow estimation algorithms based on deep learning.
- The paper summarizes the main challenges in the development of localization technologies under GPS-denied conditions and proposes potential solutions.

The structure of this review paper is as follows. In Section 2, 3, 4 and 5 it introduces optical flow-based techniques used in visual odometry and examines their applications. Section 6 contains the conclusion of the review article.

2. Visual Localization Through optical flow-based visual odometry techniques

Visual odometry is a vision-based navigation method that uses visual features from a sequence of images to estimate the relative pose of a robotic platform. It is a well-established technique for autonomous navigation and localization, in which an agent estimates its location and orientation based on the visual information from its onboard cameras. In the last few years, researchers have put forth several VO techniques, which can be classified into two categories based on what type of camera is employed: monocular camera methods and stereo camera methods. A binocular camera is the most widely employed stereo camera, which can utilize the space between the two cameras to gain depth information. Using RGBD cameras, both image and depth information can be acquired concurrently; however, the scope of the acquired depth data is restricted and dependent on infrared light, in addition to consuming considerable power. The simplicity of structure and affordability that monocular cameras offer have encouraged a plethora of studies to be conducted on them. The utilization of monocular cameras as a vision sensor for the VO method is widespread due to their affordability, compactness, and energy efficiency. This makes them suitable for small platforms, including Unmanned Aerial Vehicles (UAVs) [7], [8]. VO can be categorized into two approaches - the direct method and the feature-based method. The latter is usually regarded as the predominant approach of VO due to its benefits of strong resistance to rotation, fuzziness, and scale transformation. This method functions by estimating the motion pose of the camera via the selection of specific points (like corner points) in the image and the concurrent assessment of the motion conditions of the connected feature points in the two frames preceding and succeeding it. Currently, numerous feature extraction techniques have been developed in the domain of computer vision, for example SURF (Speeded Up Robust Features) [10] (building histogram according to the magnitude of gradient value), SIFT(Scale-Invariant Feature Transform) [9] (generating features by utilizing the histogram of gradient direction and gradient magnitude),and ORB (Oriented FAST and Rotated BRIEF(Binary Robust Independent Elementary Features)) [11] (constructing histogram depending on the pixel value). Chen et al employed the SURF algorithm to identify and correlate feature points. For the purpose of matching, they utilized the Approximate Nearest Neighbor (ANN) algorithm. Initially, the SURF features drawn from each of the template flags are broken down into eight clusters and stored. These clusters are then compared with the databases of the images to be processed and the likely images are identified. Finally, the features are matched with the aid of ANN [12]. *Figure 1.* represents an example of that. Zheng et al. [13] conducted a study in which they applied various algorithms for feature

extraction on images in five different settings. The data in the *Table 1* displays that The ORB was the quickest, whereas SIFT method took the longest time to compute.

Table 1: Point and Time comparison of three feature extraction algorithms (SIFT, SURF, and ORB) [13]

ORB		SIFT		SURF	
Point	**Time(s)**	**Point**	**Time(s)**	**Point**	**Time(s)**
168	0.00282	171	0.01889	86	0.01244
299	0.00411	253	0.01899	254	0.01621
251	0.00470	234	0.01966	187	0.01386
168	0.00221	175	0.02557	183	0.01570
19	0.00175	24	0.01669	20	0.01038

Figure 1: The SURF algorithm extracts feature points and matches them

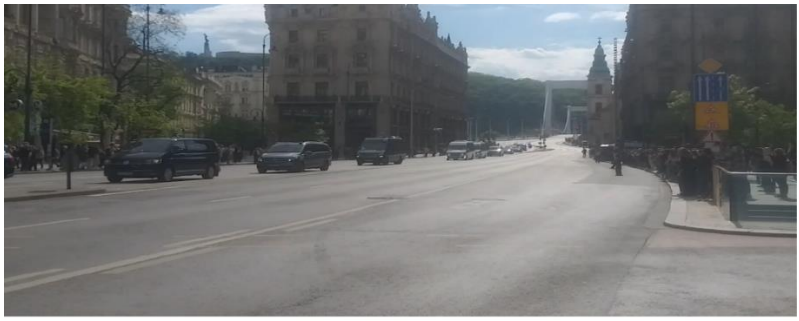
Although the feature-based approach is the most common method for Visual Odometry, it does possess some drawbacks. These include its requirement for extensive computing power and the fact that the feature extraction and descriptor computation processes are lengthy. It has been noted that when using this method to represent image motion, only a few hundred features are extracted, resulting in a significant amount of data loss compared to the hundreds of thousands of pixels that make up the image. This is especially apparent when there is a lack of texture in the image. It has been observed that the features that can be determined from the image are highly limited, making it extremely difficult for the feature-based approach to accurately estimate the camera motion in this circumstance. Consequently, scientists are attempting to utilize optical flow to enhance the feature-based approach. The navigation challenges can be tackled using optical flow techniques, which are based on the same principles of directional sensing

and localization that birds and insects employ while flying [14]. Currently, this technology has become a popular choice for robot positioning and navigation, providing dependable speed and position data. The optical flow method takes advantage of the modifications to pixels in the image sequence in the time domain and the relationship between adjacent frames in order to pinpoint the correlation between the preceding frame and the current frame, thereby computing the movement of elements between adjacent frames. It can be said that the instantaneous change rate of gray level of a certain coordinate point of the two-dimensional image plane is commonly classified as the optical flow vector. Optical flow has a great advantage as it can precisely measure and spot the location of a moving goal without any knowledge of the scene's information. Furthermore, it remains effective even when the camera is in motion. Optical flow not only furnishes insight into the unknown environment, but also assists in figuring out the direction and velocity of the robot and can detect moving subjects without any prior details about the scene. Moreover, because of the relatively advanced development of optical sensors, the price is usually low, and it is quite simple to reduce their size, which can effectively lessen costs and raise portability [4].

3. Traditional Methods of Optical Flow

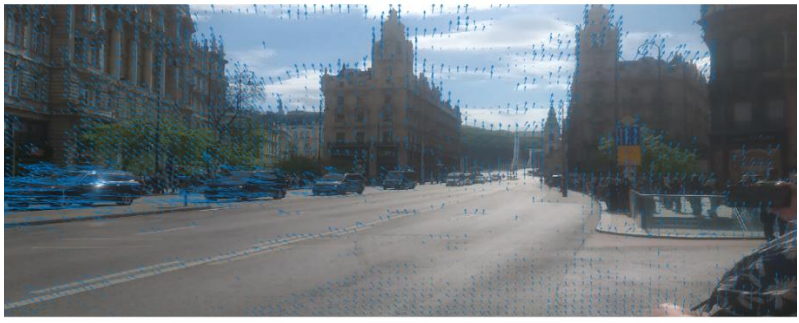
Different methods for computing optical flow have been suggested by researchers, including Lucas-Kanade algorithm [16], HornSchunck algorithm [17], image interpolation algorithm [18], block matching algorithm [19], and feature matching algorithm [9]. Optical flow estimation is typically predicated on the supposition that there is a constancy in brightness and a smoothness present. The constancy of brightness postulates that the luminance of an object does not waver between two successive frames. The idea of smoothness further expresses that the displacement is minimal; thus, the values of pixels in the vicinity are alike [17]. In real-world scenarios, the lighting conditions rarely meet the requirement of having the same intensity in adjacent frames for optical flow methods. Any changes in the lighting conditions will then have an impact on the accuracy of the optical flow measurements. Zhang et al. [20] proposed an optical flow localization technique based on ROF (Rudin-Osher-Fatemi) denoising to tackle the issue of optical flow calculation in non-uniform lighting. The convex optimization theory and duality principle were utilized to decompose the image in changing lighting and to minimize its effect. Boretti et al. [21] applied the Lucas-Kanade method for computing sparse optical flow fields and detected features with an ORB detector to calculate the optical flow of the image and further extract the motion information of the position and attitude of the MAV (Micro Aerial Vehicle) [20]. However, it was difficult to strictly adhere to the constant smoothness assumption in practicality. An image pyramid approach was

taken up to prevent the breakdown of feature detection in times of significant movement [21]. In the same way, Lou et al. [22] used image texture decomposition and image pyramid techniques to improve the Lucas-Kanade optical flow algorithm, diminishing the interference of illumination changes and large displacements on the detection of moving objects. The authors of this paper reproduced the research results of [22] and presented it in *Figure 2*.



(a) Input Image

(b) L-K Optical Flow Algorithm



(c) The Improved Algorithm by Lou etc.

Figure 2: Results of moving object detection

4. Deep Learning Methods of Optical Flow

The performance of deep learning methods for estimating optical flow surpasses that of traditional image-based approaches, and these techniques do not require explicit modeling of the entire problem, making them especially promising for use in optical flow estimation projects [23]. Dosovitskiy et al. [24] were the originators of an optical flow estimation technique based on learning, which they dubbed FlowNet. This supervised learning approach was designed to address the issue of optical flow estimation. In their work, Dosovitskiy et al. [24] introduced the Flying chairs dataset for the purpose of training the FlowNet network. Tests demonstrated that the FlowNet model developed using this synthetic dataset was able to generalize to images of the real world. *Figure 3.* illustrates the structure of two FlowNets. The first, known as FlowNetSimple, consists of a series of networks with only convolutional layers, in which two consecutive frames of input images are superimposed. On the other hand, FlowNetCorr takes two frames of pictures and processes them separately, extracting their respective features through a convolution layer and then performing a matching.

Despite its advantages, one of the major drawbacks of FlowNet is its high prediction error rate, making it unable to correctly process small displacements and real-world data. As a result, FlowNet2 was developed as an improved version.

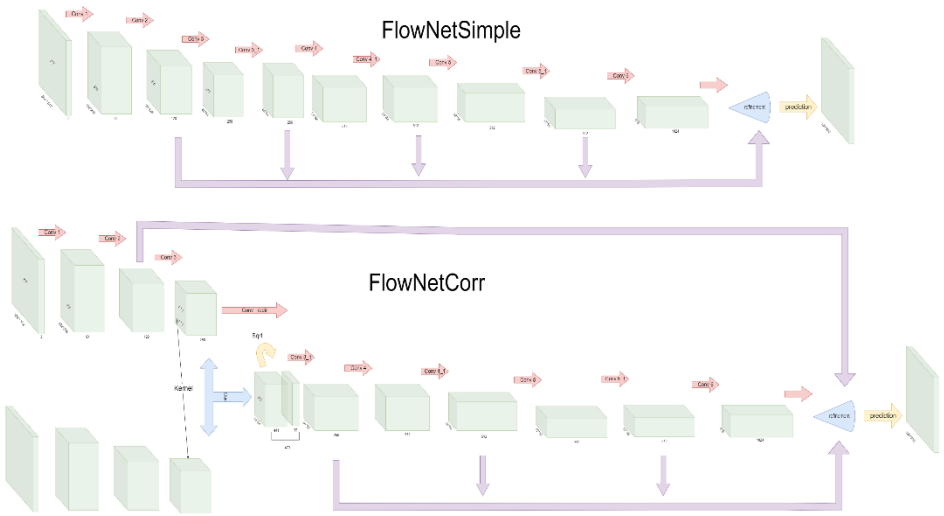


Figure 3: The network structures of FlowNetSimple (top) and FlowNetCorr (bottom)

Ilg et al. [25] demonstrated a substantial decrease in estimation errors when using a serial multiple network architecture in combination with a branch network to address optical flow estimation problems with small object displacement. Sun et al. [26] proposed an enhanced version of PWC-Net to calculate optical flow with high resolution. This network applies convolutional neural networks (CNNs) to capture image features and then employs the pyramid processing principle to predict optical flow at low resolution, while gradually progressing to the desired resolution [26]. PWC-Net has the benefit of being simpler to train than its counterpart, FlowNet2. Zhu et al. [23] proposed EV-FlowNet, based on FlowNet, as a novel self-supervised deep learning channel for estimate optical flow from event-based cameras. To do this, they first presented a novel approach for the depiction of event flow as images. A deep learning network is applied to an image with four channels, it encodes positive events with the first two channels and negative events with the other two channels. EV-FlowNet utilizes a single input of a (256×256×4) image sourced from a specific event stream. Through the use of the estimated traffic from the network, the corresponding gray level image taken from the same camera at the same time as the event is then used as a supervision signal which provides a loss function during the training process. The combination of images and self-monitoring loss is enough to enable the network to accurately predict optical flow solely from events [23].

Figure 4. shows the network structure of EV-FlowNet. The convolutional layer (in green) is responsible for downsampling (encoding), and the convolutional results from each layer are kept and linked to the upsampling (decoding) layer for use as a skip layer. The middle blue part is the residual block, which helps to further extract features.

The last part (in yellow) is the upsampling (decoding) section, which is accomplished by symmetric padding.

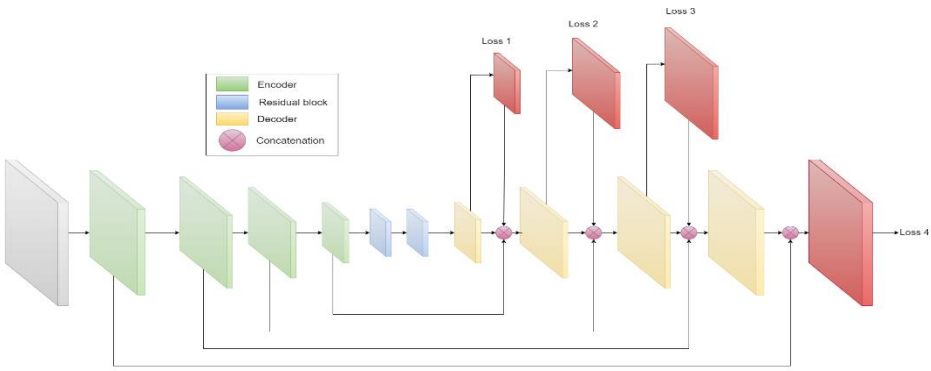


Figure 4: Illustrates the network structure of EV-FlowNet

The decoder activations from each set are passed through a depth wise convolutional layer to generate stream predictions at the resolution. The loss is applied to this stream prediction, and the prediction is also connected to the decoder activation [23]. Shi and Yin [27] introduced GeoNet, a collaborative unsupervised learning system, which was compared to EV-FlowNet for the purpose of estimating monocular depth, optical flow, and ego-motion from videos. The data acquired by the camera is comprised of rigid flow (static characteristics) and non-rigid flow (dynamic features). These are derived from not just the camera's motion, but also the movement of the target object. By taking this into account, the researchers devised a brand-new cascaded architecture with two stages to calculate both the global and fine displacement of the picture, respectively, to be able to adaptively tackle the figure of rigid and non-rigid flow. *Table 2* below presents an evaluation of FlowNet and its improved algorithms. *Figure 5* shows a schematic illustration of the UAV autonomous navigation network designed by Mumuni et al. [33] Ground plane segmentation (G-Seg) maps are also used to calibrate the metric scale. DepthNet estimates depth per frame, EgoMNet estimates relative camera pose, and OFNet predicts optical flow. The network also estimates confidence maps for each task [33].

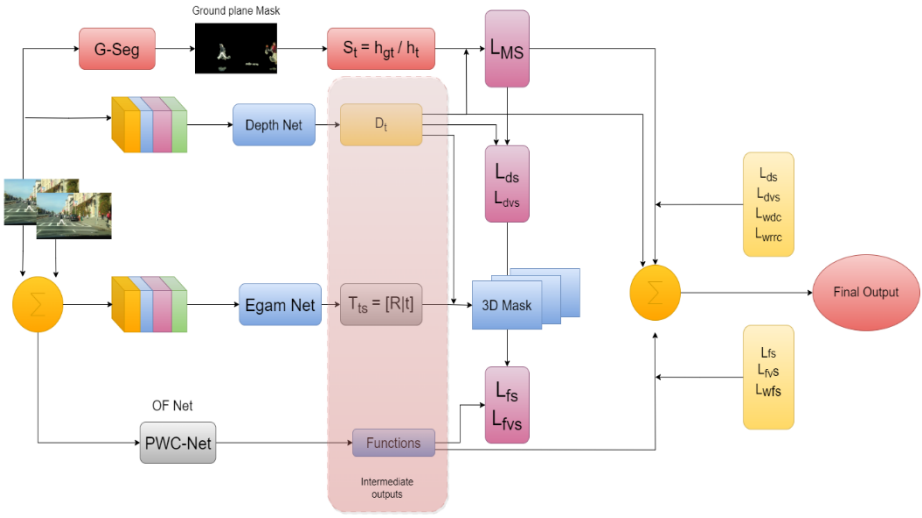


Figure 5: Schematic illustration of the UAV autonomous navigation network

5. Application of Visual odometry

The focus on Visual Odometry (VO) has grown substantially in recent years, as technology has become increasingly employed in the realms of robotics, autonomous driving vehicles, and Augmented Reality (AR). In the absence of GPS, Visual Odometry (VO) has become a popular choice due to its cost effectiveness and ease of access, acting as a supplementary tool to Inertial Navigation Systems (INS) and wheel odometers.

Table 2: A comparison of the benefits and drawbacks of FlowNet and its succeeding modified algorithms

Method	Main Contribution	Main Disadvantages	Dataset
FlowNet [24]	CNN was the first to set the example of being able to forecast OF.	Prediction accuracy not satisfactory, the system not responding to small movements or actual data.	Middlebury [28] + KITTI [29] + Sintel [30] + FlyingChairs [24]
FlowNet2[25]	FlowNet surpassed in terms of accuracy and speed.	The impact of image noise on forecast is still significant.	Middlebury + KITTI + Sintel + Flying Chairs
PWC-Net [26]	Reduce the size of the CNN, process training simpler.	Blurry estimates.	KITTI + Sintel + Flying Chairs
EV-FlowNet [23]	Use event camera data as an input for the network.	The picture capture technique is unusual and causes poor migration.	MVSEC [31]
GeoNet [27]	Evaluation difficulty of rigid and non-rigid flows.	The GeoNet model demonstrates less successful results than the direct unsupervised flow.	KITTI+ Cityscape [32]

VO techniques can be classified into three primary categories: those based on geometry, those based on deep learning, and those that combine the two. Conventional monocular VO algorithms typically involve three primary stages: tracking, optimization, and a closed-loop module. These processes exploit the geometric properties of the scene and often rely on optical flow techniques to identify image features. Despite the fact that methods that have been in use for a long time are usually more dependable and precise when it comes to determining a pose and navigational purposes, they frequently lack the ability to accurately deduce the scale without extra data. Simultaneously, the utilization of deep

learning can address the above issues by training convolutional neural networks with a considerable amount of data. As opposed to manually setting geometric limitations, DL-based approaches can get them by taking advantage of the pre-existing knowledge in the training information. Even when the shift in the apparent position of an object is not significant, it is still possible to calculate a reasonable position and depth. It is evident that online learning can be an effective tool for enhancing efficiency. Despite the advantages of deep learning-based approaches, the traditional methods still have high accuracy in estimating ego-motion [34].

5.1 Obstacle detection

Obstacle detection in VO (Visual Odometry) is a method of analyzing a sequence of images to detect the presence of obstacles or objects in the environment. It is used in navigation and localization systems, as well as in autonomous robots. By analyzing the images, the system can identify objects and obstacles, such as walls, furniture, or other objects that could potentially block the path of the robot or other entities. The optical flow algorithm is advantageous in that it can detect not only the location of a moving object, but also its speed and direction. Furthermore, since it does not need background modelling or updating, it is a very popular choice [22]. Currently, three predominant approaches for detecting obstacles by vision are monocular cues, stereo vision, and motion parallax [53]. The motion parallax method relies heavily on optical flow and uses it to get the movement and shape of both the viewer and the objects in the scene from a set of images. As indicated by Meneses et al. [54], optical flow was adopted to spot obstacles rather than calculating the movement of the robot. According to the definition of optical flow, areas with low optical flow intensity have less relative motion and thus a lower likelihood of containing obstacles. Through robot navigation, the optical flow data is conveyed to a Support Vector Machine (SVM) classifier with a radial basis function (RBF) kernel, which is used to determine if any obstacles are present in the intended path. This allows the robot to modify its course, usually in the direction with a lower intensity of optical flow [54]. Kendoul et al. [53] made use of the dense optical flow approach to gain knowledge about the entire environment and applied the Gunnar-Farneback method to calculate dense optical flow [15]. When the Unmanned Aerial Vehicle (UAV) is in motion, the presence of a multitude of optical flow vectors in its field of view suggests that an obstacle is present ahead. Conversely, if the magnitude of the optical flow vector is less than the set threshold, it implies that there is nothing blocking the UAV's path. The authors tackle the issue of foreground and background being intertwined by utilizing different thresholds to differentiate between regions, based on the magnitude of the optical flow vector. Subsequently, a clustering process is employed to

amalgamate similar regions [53]. Zhang et al. [55] proposed an auto-localization method for scenarios wherein tall buildings stand on both sides and there is no aerial view. To evaluate the rotation of the image captured by the monocular camera, they employed an optimization algorithm with the RANSAC algorithm to fit the points that correspond to the obstructions on both sides into a straight line, thus reducing the variance of errors in the forward-looking points of the robot. Then, they used Kalman filtering algorithm to determine the vanishing point of the straight line, which is the point at which the parallel obstacles on either side converge at an infinite distance. The Ackermann steering model and singular value decomposition are utilized in the algorithm for calculating the trajectory of the car, while wheel encoders are used to figure out the rate of translation. The application of optical flow in CNNs makes it easier to process objects in motion within very active settings. Rashed et al. [56] boosted the results of semantic segmentation by taking advantage of motion and depth information from optical flow when distinguishing between roads, structures, and trees, among others. A CNN architecture that is based on semantic segmentation through the combination of multiple modes of data is mainly used in autonomous driving, where earlier knowledge is utilized to make the segmentation process better [57].

5.2 Multi-sensor fusion

Under conditions where GPS is not available, the standard approaches to localization are usually SLAM (simultaneous localization and mapping), inertial IMU (inertial measurement unit) positioning, and visual localization. SLAM is precise, but the size and price of it are considerable; IMU apparatuses tend to be erratic, causing integral mistakes; and visual localization needs considerable real-time computing power support. Currently, multi-sensor fusion localization has been demonstrated to be successful in lessening localization errors, augmenting system robustness, and decreasing expenses. Consequently, it has become a popular approach among researchers. Shen et al. [5] put forward a multi-sensor fusion localization algorithm leveraging the Extended Kalman Filter (EKF). Visual Odometry (VO) was employed to calculate the velocity and position of the UAV, and a magnetometer was utilized to measure its attitude. Subsequently, the Extended Kalman Filter [5] was employed to adjust any drift observed in the inertial navigation system (INS). Kim et al. [35] suggested a way to enhance Shen's research by devising a Feature Point Threshold Filter (FPTF) algorithm that can enhance the performance of INS + Optical Flow sensor fusion by modifying the threshold according to the elevation and speed of the UAV. Due to the inability of monocular visual odometry to precisely determine depth and distance, a significant issue that it encounters is scale blur. To address this, Yu et al. [36] use prior information from the environment to find a solution. In their

research, a GPE (Ground Point Extraction) algorithm based on Delaunay triangulation [37] and ground points to combine ground points from subsequent frames is adopted, supposing a steady camera height over the ground and fitting the ground plane employing the least squares method. The Point Aggregation (GPA) algorithm determines the true scale by calculating the ground plane based on the collected data. Subsequently, a RANSAC-based optimizer is employed to solve a least squares problem in order to finalize the scale.

The results from experiments conducted on the KITTI dataset demonstrate that the framework put forward by Yu et al. [36] is effective in terms of both translation and rotation errors. Additionally, the framework demonstrates excellent computational efficiency, achieving a performance frequency of 20 Hz on the KITTI dataset. Mostafa et al. [4] suggested a new intelligent hybrid vision-aided inertial navigation system to deal with the issue of scale ambiguity in vehicle motion estimation that optical flow provides. The system is constructed of three distinct components: a Visual Odometry (VO) module, a Gaussian Regression Process (GPR) for predicting the movements of an Inertial Navigation System (INS) and another GPR for forecasting the drift of the VO. The operation can be broken down into three phases: first, when GNSS signals are available, the monocular VO and INS drift estimator are trained; second, the monocular VO drift estimator is trained to model the errors related to the speed estimates of the monocular VO; and finally, if GNSS signals are lost, the monocular VO drift estimator is used for prediction. This scheme has the advantage of being able to build up a representation of any discrepancies caused by the monocular visual odometry drift or the inside navigation system drift when GNSS data is accessible, as well as foreseeing these errors during times when GNSS readings are not available. Previous regression techniques using Gaussian Process Regression [38] and Support Vector Machine (SVM) [39] may yield unreliable results if there are insufficient features available or when there are frequent duplicated patterns. Mostafa et al. [4] put forth a technique that efficiently resolves the major issue of not being able to address missing optical flow vectors in some regions of an image due to conflicting matching. It has been demonstrated through experiments that the algorithm is able to effectively decrease positioning error during a GNSS signal loss. In comparison to the pure VO/INS procedure and VO/INS with GPR rectification, the algorithm reduced positioning error to 47.6% and 76.3% respectively, when a GNSS signal disruption happened for one minute. Xu et al. [40] present a multi-layer methodology for multi-target tracking which combines regular optical flow restrictions with product terms and utilizes a Sequential Convex Programming (SCP) technique to tackle the ensuing nonconvex optimization issue. To enhance the precision and dependability of the autonomous navigation algorithm of the aircraft, they modelled and inspected the mistakes of each sensor in the VO/IMU

integrated navigation system and investigated the Kalman filtering-based free combination filtering equation [40]. Nevertheless, drones and other miniaturized systems have restrictions when it comes to their dimensions, the amount of cargo they can carry, and the power they possess, issues that are regularly experienced in the realms of computer vision and robotics.

In recent times, researchers have been attempting to enhance the visual localization system, due to the challenge of needing high computing power and portability. The authors of the paper [41] by He et al. conducted an examination of the state-of-the-art edge-based visual odometry (EBVO) and developed an optimization framework known as PicoVO. This framework can effectively lower the computation and memory requirements. Santamaria et al. [43] came up with an efficient, cost-effective and high-performance approach to state estimation to allow MAVs to autonomously fly with minimal processing power requirements. This technique incorporates a smart camera with a monocular camera, an ultrasonic distance sensor and a three-axis gyroscope. The camera allows for high-frequency optical flow measurements, range of reflective surfaces, and three-axis angular rate to be taken, thus ruling out the need for the CPU to execute real-time image processing. The inventiveness of their proposed algorithm is that it does not rely on the optical flow information to determine the linear velocity, but rather directly observes the motion state with the initial optical flow information, thus separating the process and measurement noise. Pastor-Moreno et al. [6] created a system, OFLAAM, intended to be used on micro air vehicles (MAVs). The design of the system is composed of a downward-pointing optical flow camera, a forward-facing monocular camera, and an inertial measurement unit (IMU). The use of a localization module enables the mapping of these features into a specific vocabulary. This module applies the DBoW2 algorithm for the purpose of position correction by loopback detection [44] in order to address the problem of optical flow drift. By merging a high-speed optical flow localization with a low-rate positioning algorithm, an autonomous localization of the MAV can be achieved while also reducing the overall computational load [6]. Dong et al. [45] came up with a creative relative localization technique for utilization with unmanned aerial vehicles. They took photos of the ground from a camera attached beneath the drone, then used the SURF algorithm to identify points between two frames and the Fast Approximate Nearest Neighbors (FLANN) algorithm to determine the optical flow. The velocity of UAVs can be determined through the application of the optical flow motion estimation equation and known parameters. The proposed system incorporates measurements from a SINS, electronic compass, optical flow, altimeter system, and laser rangefinder to achieve relatively precise localization data. A summary of the sensor fusion VO algorithms is presented in *Table 3* below, which highlights the main issues and their respective solutions.

Table 3: Comparison of VO Algorithms for Sensor Fusion

Method	Objectives	Solutions
Shen et al. [5] Kim [35]	Implement multi-sensor fusion	- Create an EKF-based fusion algorithm. - Create a fusion algorithm using FPTF
Yu et al. [36] Mostafa [4] Xu et al. [40]	Solving scale ambiguity	- Train scale drift predictors in the presence of GNSS signals - examination of each sensor's error
PicoVO [42] Angel [43] OFLAAM [6] Dong [45]	Use navigation algorithms on low-computing platforms.	- real-time picture processing - Combining low-speed positioning techniques with high-speed OF - Using OF Motion Estimation

5.3 Estimate Speed-Distance-Position

Evaluating the state of motion is a significant area of research when it comes to localization, and the VO technique that is based on optical flow/feature matching can supply an abundance of details about self-motion. Ho et al. [46] applied an extended Kalman filter (EKF) in combination with images from a monocular camera to analyze the divergence of the temporal flow vector during the UAV's vertical touchdown, ascertain the altitude and vertical velocity of the UAV, and govern the UAV's landing [46]. For every taken picture, they used the FAST algorithm to discover corners and followed them in the subsequent frame utilizing a Lucas-Kanade tracker, which enabled them to quantify the contraction and expansion of optical flow and calculate the divergence of optical flow. Mumuni et al. [33] incorporated Structure of Motion (SFM) [47] with Optical Flow and Monocular Depth Estimation (MDE), to augment the precision of depth estimation. Generally, the dense measurements derived from MDEs are uncertain in scale, while the sparse depths derived from SFM models possess a metric scale. MDE furnishes abundant depth measurements, so only some sparse depth information from SFM is required to complete the picture. By combining these two approaches, the precision of the unmanned aerial vehicle's depth estimation can be improved. Recent years have seen impressive results from CNN-based optical flow models [48] [49], however Mumuni et al.'s [50] algorithm is more efficient in terms of memory and computation [50]. To apply optical flow for the purpose of localization on small unmanned aerial vehicles (UAVs), it is necessary to develop a more efficient optical flow algorithm. Because of that McGuire et al. [52] developed the EdgeFlow algorithm, which builds on a feature density distribution-based collision detection algorithm [51] by introducing a variable time horizon for subpixel flow detection and employing spatial edge distribution for the purpose of image motion tracking. Tests have shown that the procedure is

computationally effective enough to be executed at close to the speed of frames on restricted embedded processors, offering reliable speed and distance estimations for Unmanned Aerial Vehicles in unfamiliar environments. Zheng et al. [13] used the Lucas-Kanade sparse optical flow algorithm in their localization approach for indoor UAVs, which requires real-time optical flow calculation, and additionally implemented Forward-Backward bidirectional tracking optimization. They opted for ORB feature extraction for its real-time capabilities, KNN forward-backward bidirectional matching for feature matching, and RANSAC algorithm to filter the matching outcomes. According to the experimental results, this approach has a high degree of accuracy in predicting velocity and location [13]. A novel end-to-end network is suggested in work by Huang et al. [57] for learning optical flow and calculating camera ego motion. They utilized an autoencoder and estimated optical flow using the PWC-Net created by Sun et al [26] (CNN Encoder).

6. Conclusion

This review examines the optical flow-based visual odometry of localization without GPS. This research focused on the traditional approaches (such as feature extraction) and more unconventional methods (like those that involve deep learning). This review outlines the fundamental concepts for each category and, if relevant, illustrates how they are used in practice. Recent research into visual localization techniques has shown an evolving trend towards more cost-effective, compact solutions that offer greater precision. In recent years, deep learning has seen a surge in growth, prompting some scientists to explore using neural networks for visual localization tasks. This has produced noteworthy outcomes. FlowNet has been identified as the most prominent approach for extracting image motion information and providing localization support, due to its utilization of optical flow extraction and its improved algorithms.

At the end, this article checked and regenerated all the results in the previous research and reapplied all the algorithms with other data and compared the results with the old ones.

References

[1] Aguilar, Wilbert G., Verónica P. Casaliglla, and José L. Pólit, "Obstacle avoidance based-visual navigation for micro aerial vehicles", Electronics 6.1 (2017): 10, pp. 1–23.

[2] Chao, H., et al., "A comparative study of optical flow and traditional sensors in UAV navigation", in *Proc. American Control Conference. IEEE*, 2013, pp. 3858–3863.

[3] Mur-Artal, R., Montiel, J. M. M., and Tardos, J. D., "ORB-SLAM: a versatile and accurate monocular SLAM system", *IEEE transactions on robotics*, vol 31, no. 5, pp. 1147–1163, 2015.

[4] Mostafa, M. M., et al., "A smart hybrid vision aided inertial navigation system approach for UAVs in a GNSS denied environment", *Navigation: Journal of The Institute of Navigation* Vol. 65, no. 4, pp. 533–547, 2018.

[5] Shen, C., et al., "Optical flow sensor/INS/magnetometer integrated navigation system for MAV in GPS-denied environment", *Hindawi Publishing Corporation, Journal of Sensors*, 2016, pp. 1–10.

[6] Pastor-Moreno, D., Shin, H. S., and Waldock, A., "Optical flow localisation and appearance mapping (OFLAAM) for long-term navigation.", in *Proc. 2015 International Conference on Unmanned Aircraft Systems (ICUAS), IEEE*, 2015, pp. 980–988.

[7] Wei, W., et al., "A survey of uav visual navigation based on monocular slam", in *Proc. 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC)*, IEEE, 2018, pp. 1849–1853.

[8] Jeon, J., et al., "Run your visual-inertial odometry on NVIDIA Jetson: Benchmark tests on a micro aerial vehicle", *IEEE Robotics and Automation Letters, vol* 6, no. 3, pp. 5332–5339, 2021.

[9] Lowe, D. G., "Distinctive image features from scale-invariant keypoints", *International journal of computer vision*, vol. 60, pp. 91–110, 2004.

[10] Bay, H., Tuytelaars, T., and Van Gool, L., "Surf: Speeded up robust features", *Lecture notes in computer science* 3951, 2006, pp. 404–417.

[11] Rublee, E., et al., "ORB: An efficient alternative to SIFT or SURF", in *Proc. 2011 International conference on computer vision*, IEEE, 2011, pp. 2564–2571.

[12] Chen, L., et al., "Design of a multi-sensor cooperation travel environment perception system for autonomous vehicle", *Sensors 12.9*, 2012, pp. 12386–12404.

[13] Wenxuan, Z, Xiao, J, and Xin, T., "Integrated navigation system with monocular vision and LIDAR for indoor UAVs", in *Proc. 12th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, IEEE, 2017, pp. 924-929.

[14] Srinivasan, M., et al., "Honeybee navigation en route to the goal: visual flight control and odometry", *The Journal of experimental biology*, 199.1, 1996, pp. 237–244.

[15] Farnebäck, G., "Two-frame motion estimation based on polynomial expansion", in *Proc. 13 Image Analysis: 13th Scandinavian Conference, SCIA 2003*, Halmstad, Sweden, June 29–July 2, 2003, Springer Berlin Heidelberg, 2003, pp. 363–370.

[16] Lucas, B. D., and Kanade, T. "An iterative image registration technique with an application to stereo vision", in *Proc. IJCAI'81: 7th international joint conference on Artificial intelligence*, Vol. 2. 1981, pp. 674–679.

[17] Horn, B. K. P., and Schunck, B. G. "Determining optical flow", *Artificial intelligence* 17.1-3, 1981, pp. 185–203.

[18] Srinivasan, M. V., "An image-interpolation technique for the computation of optic flow and egomotion", *Biological cybernetics* 71.5, 1994, pp. 401–415.

[19] Farid, K., Fantoni, I., and Nonami, K. "Optic flow-based vision system for autonomous 3D localization and control of small aerial vehicles", *Robotics and autonomous systems*, 57.6-7, 2009, pp. 591–602.

[20] Zhang, L., Xiong, Z., Lai, J., and Liu, J., "Research of optical flow aided MEMS navigation based on convex optimization and ROF denoising", *Optik*, vol. 158, pp. 1575–1583, 2018.
[21] Boretti, C., et al. "Visual Navigation Using Sparse Optical Flow and Time-to-Transit", in *Proc. 2022 Intern. Conf. on Robotics and Automation (ICRA)*, IEEE, 2022, pp. 9397–9403.
[22] Li, L., Liang, S., and Zhang, Y., "Application research of moving target detection based on optical flow algorithms", *Journal of Physics: Conference Series*, vol. 1237., no. 2., IOP Publishing, 2019, p. 022073.
[23] Zhu, A. Z., et al., "EV-FlowNet: Self-supervised optical flow estimation for event-based cameras", arXiv preprint arXiv: pp. 1802.06898 (2018).
[24] Dosovitskiy, A., et al., "Flownet: Learning optical flow with convolutional networks", in *Proc. of the IEEE international conference on computer vision*, 2015, pp. 2758–2766.
[25] Ilg, E., et al. "Flownet 2.0: Evolution of optical flow estimation with deep networks", in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2462–2470.
[26] Sun, D., et al., "Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume." in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018.
[27] Zhichao, Y., and Shi, J., "Geonet: Unsupervised learning of dense depth, optical flow and camera pose", in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1983–1992.
[28] Scharstein, D., et al., "High-resolution stereo datasets with subpixel-accurate ground truth", in *Proc. Pattern Recognition: 36th German Conference, GCPR 2014*, Münster, Germany, September 2-5, 2014, Proceedings 36, Springer International Publishing, 2014, pp. 31–42.
[29] Geiger, A., et al., "Vision meets robotics: The kitti dataset", *The International Journal of Robotics Research* 32.11, 2013, pp. 1231–1237.
[30] Butler, D. J., et al., "A naturalistic open source movie for optical flow evaluation." in *Proc. Computer Vision–ECCV 2012: 12th European Conference on Computer Vision*, Florence, Italy, October 7-13, 2012, Proceedings, Part VI 12. Springer Berlin Heidelberg, 2012, pp. 611–625.
[31] Zhu, A. Z., et al. "The multivehicle stereo event camera dataset: An event camera dataset for 3D perception", *IEEE Robotics and Automation Letters*, 3.3, 2018, pp. 2032–2039.
[32] Cordts, Marius, et al., "The cityscapes dataset for semantic urban scene understanding" in *Proc. of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 3213-3223.
[33] Mumuni, F., Mumuni, A., and Amuzuvi, C. K., "Deep learning of monocular depth, optical flow and ego-motion with geometric guidance for UAV navigation in dynamic environments", *Machine Learning with Applications*, 10, 2022, p. 100416.
[34] Zhang, J., et al., "Deep online correction for monocular visual odometry", in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 14396–14402.
[35] Taegyun, K., et al., "Improved optical sensor fusion in UAV navigation using feature point threshold filter", *International Journal of Aeronautical and Space Sciences*, 2022, pp. 1-12.
[36] Yu, T., et al., "Accurate and robust stereo direct visual odometry for agricultural environment", in *Proc. 2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 2480–2486.
[37] Pinggera, P., et al., "Know your limits: Accuracy of long range stereoscopic object measurements in practice", in *Proc. Computer Vision–ECCV 2014: 13th European Conference*, Zurich, Switzerland, Sept. 6-12, 2014, Proceedings, Part II 13. Springer International Publishing, 2014, pp. 96–111.
[38] Guizilini, V., and Ramos, F., "Visual odometry learning for unmanned aerial vehicles", in *Proc. 2011 IEEE International Conference on Robotics and Automation*, IEEE, 2011, pp. 6213–6220.

[39] Ciarfuglia, T. A., et al. “Evaluation of non-geometric methods for visual odometry”, *Robotics and Autonomous Systems* 62.12, 2014, pp. 1717–1730.
[40] Xu, Q., et al., “An Optical Flow Based Multi-Object Tracking Approach Using Sequential Convex Programming”, in *Proc. 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, IEEE, 2020, pp. 1216–1221.
[41] Schenk, F., and Fraundorfer, F., “Robust edge-based visual odometry using machine-learned edges”, in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2017, pp. 1297–1304.
[42] He, Y., et al., “Picovo: A lightweight rgb-d visual odometry targeting resource-constrained iot devices”, in *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 5567–5573.
[43] Santamaria-Navarro, A., et al., “Autonomous navigation of micro aerial vehicles using high-rate and low-cost sensors”, *Autonomous robots*, 42, 2018, pp. 1263–1280.
[44] Gálvez-López, D., and Tardos, J. D., “Bags of binary words for fast place recognition in image sequences”, *IEEE Transactions on Robotics*, 28.5, 2012, pp. 1188–1197.
[45] Zhuoning, D., Li, W., and Zhou, Y. “An autonomous navigation scheme for UAV in approach phase”, in *Proc. IEEE Chinese Guidance, Navigation and Control Conference (CGNCC)*, IEEE, 2016, pp. 982–987.
[46] Ho, H. W., de Croon, G., and Chu, Q. P., “Distance and velocity estimation using optical flow from a monocular camera”, *International Journal of Micro Air Vehicles*, 9.3, 2017, pp. 198–208.
[47] Ioannou, P. and Fidan, B., “Adaptive control tutorial”, SIAM, 2007.
[48] Liu, L., et al. “Learning by analogy: Reliable supervision from transformations for unsupervised optical flow estimation.” in *Proc. of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 6489-6498.
[49] Jonschkowski, R., et al., “What matters in unsupervised optical flow”, in Proc. *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, Aug. 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020, pp. 557–572.
[50] Mumuni, F, and Mumuni, A., “Bayesian cue integration of structure from motion and CNN-based monocular depth estimation for autonomous robot navigation”, *International Journal of Intelligent Robotics and Applications*, 6.2, 2022, pp. 191–206.
[51] Lee, D.-J., et al., “See and avoidance behaviors for autonomous navigation”, *Mobile Robots* Xvii, vol. 5609, SPIE, 2004, pp. 23–34.
[52] McGuire, K., et al., “Efficient optical flow and stereo vision for velocity estimation and obstacle avoidance on an autonomous pocket drone”, *IEEE Robotics and Automation Letters*, 2.2, 2017, pp. 1070–1076.
[53] Farid, K., “Survey of advances in guidance, navigation, and control of unmanned rotorcraft systems”, *Journal of Field Robotics*, 29.2, 2012, pp. 315–378.
[54] Meneses, M. C., Matos, L. N., and Prado, B. O., “Low-cost Autonomous Navigation System Based on Optical Flow Classification”, arXiv preprint arXiv:1803.03966 (2018).
[55] Zhang, J., et al. “Monocular visual navigation of an autonomous vehicle in natural scene corridor-like environments”, in *Proc. 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, 2012, pp. 3659–3666.
[56] Rashed, H., et al., “Motion and depth augmented semantic segmentation for autonomous navigation”, in *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 364–370.
[57] Huang, Y., et al., “Learning optical flow with R-CNN for visual odometry”, *IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 14410–1441.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 33–44

DOI: 10.2478/auseme-2023-0003

Performance Analysis of Combining Beamforming and ZF/ MMSE-SIC Equalization Techniques for MIMO DWPT-COFDM Systems

SADOUKI BOUAMAMA Reda[1], MENEZLA Fayssal[2], DEBBAL Mohammed[3]

[1] University of Saida, Algeria, Electromagnetism, Photonics and Optronics Laboratory (LEPO), Sidi Bel Abbés, Algeria, e-mail: bouamama.sadouki@yahoo.fr
[2] University Center Nour El Bachir of El Bayadh, Algeria, Electromagnetism, Photonics and Optronics Laboratory (LEPO), Sidi Bel Abbés, e-mail: menezla@yahoo.fr
[3] University of Belhadj BOUCHAIB, Ain-Temouchent, Algeria, Telecommunication Laboratory (LTT), Tlemcen, Algeria, e-mail: debbal.mohammed@gmail.com

Manuscript received September 29, 2023; revised November 30, 2023

Abstract: This study explores the enhancement of wireless communication through the integration of OFDM with MIMO and advanced techniques like COFDM, adaptive MIMO, and beamforming. It evaluates a 2×2 and 4×4 MIMO-DWPT-COFDM system, considering modulation (BPSK and QPSK) and equalization (ZF-SIC and MMSE-SIC) over a Rayleigh fading channel. MATLAB results reveal that more antennas, BPSK modulation, and MMSE-SIC equalization significantly improve BER performance. Specifically, for a BER of 10^{-4}, the 2×2 MIMO system requires E_b/N_0values of 10.4 (MMSE-SIC) and 11.4 (ZF-SIC), while the 4×4 MIMO system needs 5.5 (MMSE-SIC) and 5.9 (ZF-SIC). These findings emphasize the need for thoughtful system design and parameter optimization in achieving reliable and efficient wireless communication.

Keywords: MIMO-DWPT COFDM, adaptive modulation, equalization techniques, OFDM systems, Rayleigh fading channel.

1. Introduction

The rapid progress in communication is driven by the increasing number of users demanding high-speed data services, necessitating expanded bandwidth [1]. However, wireless communication faces challenges like Inter-Symbol Interference (ISI) and Multi-User Interference (MUI), particularly in high-density user environments [2], [3].

Simultaneously, Orthogonal Frequency Division Multiplexing (OFDM) and its variants like Coded OFDM (COFDM) have gained widespread adoption in wireless networks [4-8]. When coupled with Multiple Input Multiple Output (MIMO) spatial multiplexing, these systems deploy multiple antennas at each end, enhancing capacity and spectral efficiency [9-13]. They rely on the Inverse Fast Fourier Transform (IFFT) and Fast Fourier Transform (FFT) algorithms, which have drawbacks, including side-lobes due to a rectangular window and the need for a Cyclic Prefix (CP) to mitigate ISI, consuming 25% bandwidth [14-17].

An alternative approach utilizes the Inverse Discrete Wavelet Packet Transform (IDWPT) and DWPT algorithms, employing Wavelet Packets (WPs) as subcarriers instead of IFFT/FFT, overcoming issues like high Peak-to-Average Power Ratio (PAPR) and bandwidth wastage [18-20], with added improvements possible through antenna diversity techniques [21-23].

Equalization techniques in MIMO and MIMO-OFDM systems, including Zero-Forcing (ZF), Minimum Mean Square Error (MMSE), ZF-Successive Interference Cancellation (ZF-SIC), and MMSE-SIC, have been studied to mitigate ISI and enhance link reliability [24-31]. Spatial signal processing in multiple-antenna wireless devices, like ZF-Beamforming (ZF-BF) and MMSE-Beamforming (MMSE-BF), further enhances system performance [37].

This study focuses on implementing adaptive beamforming in a MIMO-DWPT-based COFDM system, introducing a novel algorithm to reduce ISI, enhance data reliability, and overall system performance through wireless physical layer network coding and ZF/MMSE-SIC equalization. An integrated approach optimizes beamforming weights and equalization filters to mitigate ISI effects from multipath propagation.

2. Proposed System Model

The general architecture of the proposed MIMO-DWPT-based COFDM system, incorporating an adaptive beamforming scheme at both the transmitter and receiver, is depicted in *Fig. 1*. This system transmits various signal variations (symbols) from N_t transmitters to N_r receiving antennas via distinct MIMO channels. To counter multi-path fading effects, known pilot symbols are periodically inserted among the transmitted encoded symbols. The OFDM system in this study employs two modulation techniques: Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK). COFDM symbols generated represent orthogonal sub-carrier frequencies, enabling simultaneous transmission of information symbols from different sub-carriers through the transmission channel, which can be flat fading or frequency-selective fading. During this journey, the signals encounter Additive White Gaussian Noise (AWGN) and undergo processing using an *n*-point IDWPT (Inverse Discrete

Wavelet Packet Transform) algorithm [22]. Notably, before this transformation, symbols are multiplied by distinct beamforming coefficients in the time domain, a crucial aspect of the adaptive beamforming scheme. The equivalent model of MIMO transmission is defined as:

$$S_m = \mathcal{Y}_m \cdot \mathcal{H}_m + n_m \tag{1}$$

where m denotes the number of carriers, $\mathcal{Y}_m \in \mathrm{C}^{N_t \times 1}$ represents the transmitted vector, $\mathcal{H}_m \in \mathrm{C}^{N_r \times N_t}$ signifies the frequency matrix of the channel, $S_m \in \mathrm{C}^{N_r \times 1}$ the received vector and $n_m \in \mathrm{C}^{N_r \times 1}$ the AWGN noise. Since the weights are specific to each antenna and to each carrier, the transmitted signal can be written as:

$$\mathcal{Y}_m = \sum_{m=1,L} b_m \cdot \mathcal{X}_m \cdot g_m \cdot (2 \cdot n - m) \tag{2}$$

where $b_m \in \mathrm{C}^{N_t \times 1}$ is the broadcast beamforming weight vector and x_m is the complex symbol sent on the sub carrier *m*.

In the proposed system, a configuration of 64 sub-carriers is employed, with Binary Phase Shift Keying (BPSK) or Quadrature Phase Shift Keying (QPSK) modulation applied uniformly across all sub-carriers. This uniformity in constellation choice simplifies signal processing and ensures consistency. The communication channel is modelled as a Rayleigh fading channel, a common representation for wireless environments characterized by multipath propagation and random fading.

Now, let's focus on the receiver side, as depicted in *Fig. 1*. Here, a series of inverse operations is executed to reconstruct the transmitted data. Initially, the received signal is multiplied by the beamforming weights, which were calculated at the transmitter end. Pilot symbols, interspersed within the transmitted signal, are extracted at specific instances to estimate crucial channel parameters. This pilot-based estimation is essential for adapting the equalization and decoding processes to the prevailing channel conditions.

To further enhance signal recovery, the receiver employs an equalization approach, specifically the Minimum Mean Square Error with Successive Interference Cancellation (MMSE-SIC) or Zero-Forcing with Successive Interference Cancellation (ZF-SIC) technique. These methods are vital for mitigating the effects of interference, such as Inter-Symbol Interference (ISI), which is commonly encountered in wireless communication.

Additionally, the received signal undergoes processing using the Inverse Discrete Wavelet Packet Transform (IDWPT) algorithm. This step is instrumental in retrieving the data symbols from the received signal, effectively undoing the wavelet-based modulation applied at the transmitter. The recovered data symbols are subsequently subjected to demodulation and decoding processes to reconstruct the originally transmitted data.

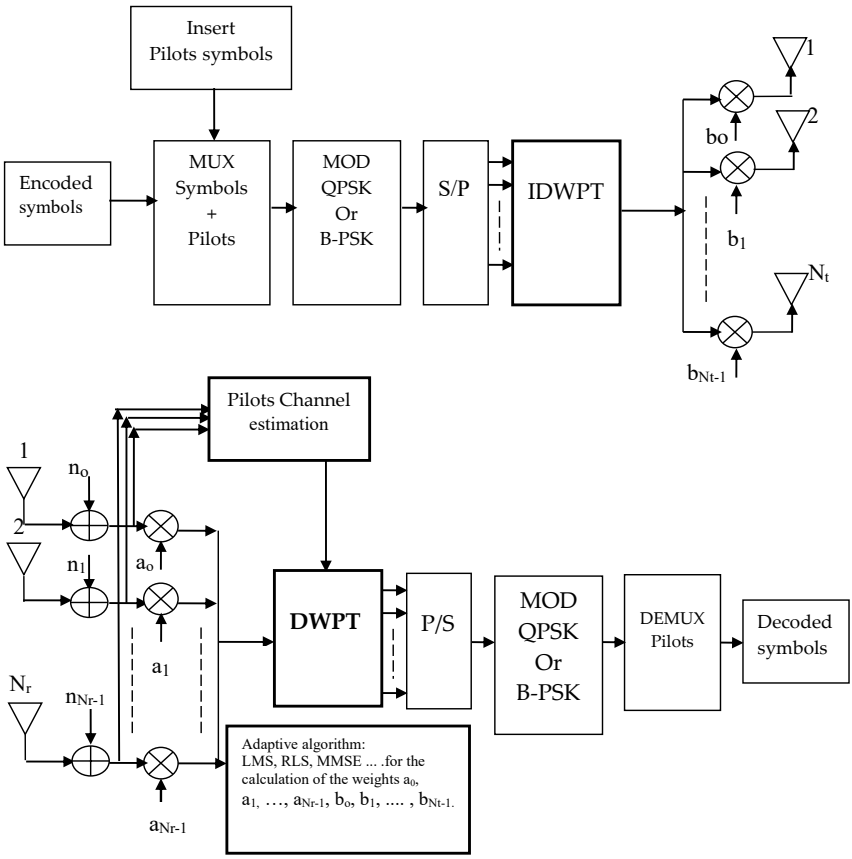


Figure 1: Block diagram of the proposed MIMO-DWPT COFDM system with beamforming and MMSE-SIC/ZF-SIC equalizer

By integrating these sophisticated techniques at the receiver, the proposed system is equipped to handle the challenges posed by the wireless channel, ensuring reliable and efficient communication even in the presence of fading and interference. In the upcoming sections, we will delve into the performance evaluation of this system, shedding light on its effectiveness in real-world scenarios.

For MIMO-DWPT COFDM system with adaptive beamforming and ZF-SIC, equalization:

$$SNR_1 = \frac{E\left[\left(\sum_{m=1}^{L-1} a_m^H \cdot b_m \cdot \mathcal{H}_m \cdot W_{ZF-SIC} \cdot \mathcal{X}_m \cdot g_m(0)\right)^2\right]}{E\left[\left(W_{ZF-SIC} \cdot a^H \cdot n\right)^2\right]} = \frac{E\left\{\left|\mathcal{X} \cdot g(0)\right|^2\right\}}{\sigma_n^2} = \tag{3}$$

$$= \frac{P \cdot \mathcal{X}}{\sigma_n^2} \cdot \left|g(0)\right|^2$$

For MIMO-DWPT COFDM system with adaptive beamforming and MMSE-SIC equalization:

$$SNR_2 = \frac{E\left[\left(\sum_{m=1}^{L-1} a_m^H \cdot b_m \cdot \mathcal{H}_m \cdot W_{MMSE-SIC} \cdot \mathcal{X}_m \cdot g_m(0)\right)^2\right]}{E\left[\left(W_{MMSE-SIC} \cdot a^H \cdot n\right)^2\right]} = \frac{E\left\{\left|\mathcal{X} \cdot g(0)\right|^2\right\}}{\sigma_n^2} = \tag{4}$$

$$= \frac{P \cdot \mathcal{X}}{\sigma_n^2} \cdot \left|g(0)\right|^2$$

3. Results and Discussions

In this section, the authors embark on a comprehensive performance analysis, primarily focusing on the Bit Error Rate (BER), to evaluate the effectiveness of two key equalization techniques: Zero-Forcing with Successive Interference Cancellation (ZF-SIC) and Minimum Mean Square Error with Successive Interference Cancellation (MMSE-SIC). These equalization methods are applied within the context of a MIMO-DWPT COFDM system with beamforming, utilizing Binary Phase Shift Keying (BPSK) and Quadrature Phase Shift Keying (QPSK) modulation schemes. The objective is to assess how these equalization techniques perform under various conditions, including the presence of Additive White Gaussian Noise (AWGN) and a multipath Rayleigh channel, both common challenges in real-world wireless communication scenarios.

To facilitate these performance evaluations, MATLAB simulations are employed, and the key simulation parameters are outlined in *Table 1*. These parameters serve as the foundation for the experiments and analysis, ensuring a consistent and standardized approach to the assessment of system performance.

Table 1: Simulation Parameters

Parameters	Specifications
Channel	Rayleigh
Beamforming Type	Adaptive
Signal Constellation	BPSK-QPSK
MIMO	2×2, 4×4
Convolutional coder	1/2
Number of sub-carriers	64
Wavelet package	Haar

The simulations will provide valuable insights into the robustness and reliability of the proposed MIMO-DWPT COFDM system with adaptive beamforming, shedding light on its ability to maintain low BER rates in the presence of noise and channel impairments. The subsequent analysis of these results will contribute to a deeper understanding of the system's real-world applicability and its potential advantages in various wireless communication scenarios.

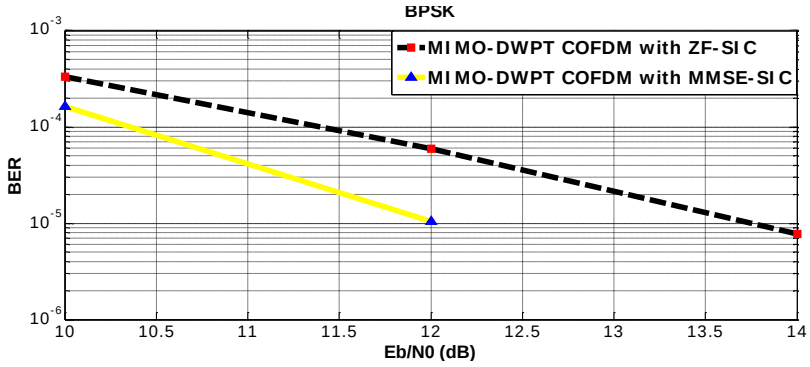


Figure 2: BER Performance comparison of ZF-SIC and MMSE-SIC equalization techniques for 2×2 MIMO-DWPT COFDM system with beamforming using BPSK modulation

Fig. 2 presents the Bit Error Rate (BER) performance of a 2×2 MIMO-DWPT COFDM system with beamforming, utilizing both Zero-Forcing with Successive Interference Cancellation (ZF-SIC) and Minimum Mean Square Error with Successive Interference Cancellation (MMSE-SIC) equalizers. The modulation scheme employed in this evaluation is Binary Phase Shift Keying (BPSK).

The graph in *Fig. 2* offers valuable insights into how the system's BER performance varies under different equalization techniques. It provides a visual representation of how effectively ZF-SIC and MMSE-SIC mitigate interference

and noise in the transmission, ultimately impacting the system's ability to maintain a low BER, which is critical for reliable data communication.

By analyzing the BER curves in *Fig. 2*, it is possible to discern which equalization technique performs better under the specific conditions and configurations of the 2×2 MIMO-DWPT COFDM system with beamforming. This information is instrumental in making informed decisions about the choice of equalization method in practical wireless communication applications, where achieving high data reliability is of utmost importance.

The observations from *Fig. 2* indeed underscore the importance of a higher signal-to-noise ratio (E_b / N_0) in achieving better Bit Error Rate (BER) performance. Additionally, the analysis reveals that the Minimum Mean Square Error with Successive Interference Cancellation (MMSE-SIC) equalizer consistently outperforms the Zero-Forcing with Successive Interference Cancellation (ZF-SIC) equalizer in terms of BER.

For instance, at an E_b / N_0 value of 12 dB, the MMSE-SIC-based system achieves an improved BER of approximately 10^{-5}, while the ZF-SIC-based system lags slightly behind with a BER of approximately 5×10^{-5}. This distinction underscores the superiority of the MMSE-SIC equalization technique, particularly in scenarios where a higher E_b / N_0 is available.

To further illustrate the impact of varying the number of transmitting and receiving antennas on the BER performance of the proposed system using MMSE-SIC and ZF-SIC equalizers, *Fig. 3* is presented. This figure showcases the simulation results obtained as the number of antennas is increased from 2 to 4 at both the transmitter and receiver ends.

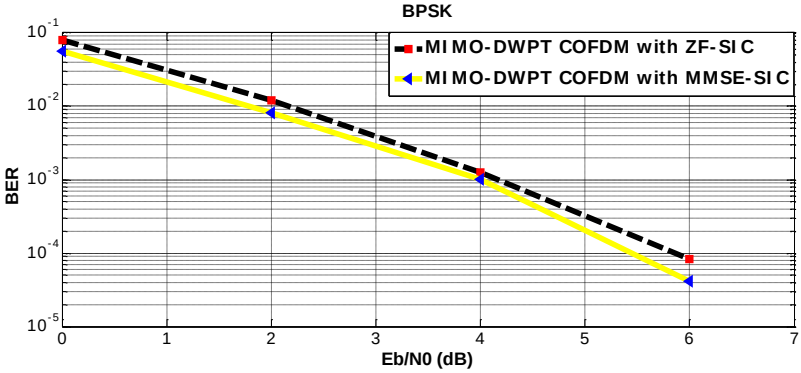


Figure 3: BER Performance comparison of ZF-SIC and MMSE-SIC equalization techniques for 4×4 MIMO-DWPT COFDM system with beamforming using BPSK modulation

Fig. 3 provides a comprehensive view of how the BER performance evolves with changes in antenna configuration. The data gleaned from this analysis will offer valuable insights into the scalability and adaptability of the MIMO-DWPT COFDM system with beamforming when subjected to varying antenna configurations. These insights can guide the design and deployment of such systems in real-world communication scenarios, where the number of antennas may vary based on available resources and specific operational requirements.

The findings from *Fig. 3* affirm that the MMSE-SIC equalization technique outperforms the ZF-SIC equalization technique in terms of BER. Moreover, increasing the number of both transmit and receive antennas has a positive impact on system performance by reducing the bit error rate for both equalization methods. This improvement can be attributed to the ability of a higher number of antennas to mitigate noise, combat fading, reduce interference, and enhance both system throughput and capacity (the number of users the system can support).

Indeed, to achieve a BER of 10^{-4}, the 2×2 MIMO system employing MMSE-SIC requires an E_b / N_0 of 10.4, while the ZF-SIC-based system necessitates an E_b / N_0 of 11.4. On the other hand, for the 4×4 MIMO system, the required E_b / N_0 values are 5.5 and 5.9 for MMSE-SIC and ZF-SIC, respectively. This comparison highlights the advantage of using a larger number of antennas, as it enables the system to achieve the desired BER performance with lower bit energy requirements. It also underscores the importance of efficiency and high throughput in systems with fewer antennas to support high-capacity applications.

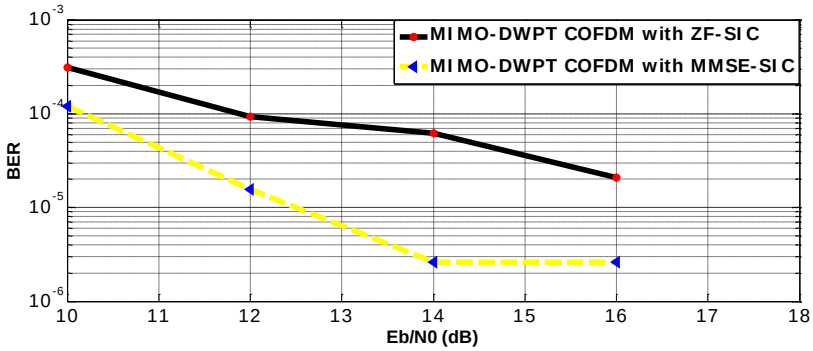


Figure 4I BER Performance comparison of ZF-SIC and MMSE-SIC equalization techniques for 2×2 MIMO-DWPT COFDM system with beamforming using QPSK modulation

To further analyze the BER performance of the proposed system, the authors investigated the impact of the modulation technique on system performance with both equalization techniques. *Fig. 4* presents a comparison of the BER

performance for the 2×2 MIMO-DWPT COFDM system with beamforming, using QPSK modulation and both ZF-SIC and MMSE-SIC equalizers.

Fig. 4 demonstrates that the MMSE-SIC equalization technique consistently outperforms the ZF-SIC equalizer, irrespective of the modulation scheme used. At an E_b / N_0 value of 12 dB, the system employing QPSK modulation achieves a BER of approximately 9×10^{-5} with ZF-SIC, while the MMSE-SIC-based system boasts a significantly improved BER of approximately 1.6×10^{-5}. This outcome underscores the suitability of MMSE-SIC for maintaining low BER rates in wireless communication scenarios, even when more complex modulation schemes like QPSK are employed.

Fig. 5 provides a detailed view of the impact of increasing the number of antennas on the Bit Error Rate (BER) performance of a MIMO-DWPT COFDM system with beamforming, utilizing both Zero-Forcing with Successive Interference Cancellation (ZF-SIC) and Minimum Mean Square Error with Successive Interference Cancellation (MMSE-SIC) equalization techniques. The modulation technique employed in this analysis is Quadrature Phase Shift Keying (QPSK).

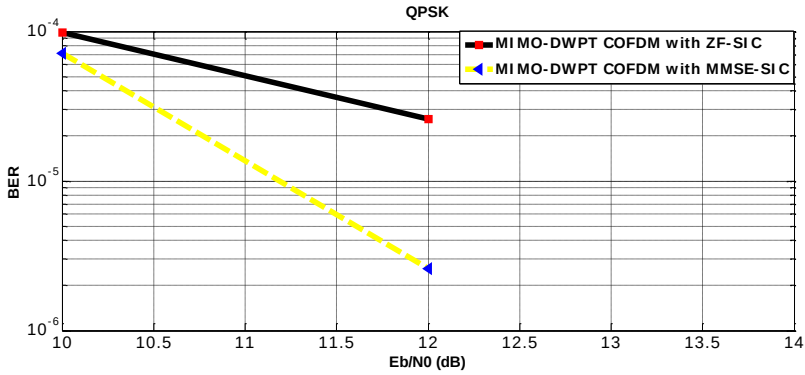


Figure 5: BER Performance comparison of ZF-SIC and MMSE-SIC equalization techniques for 4×4 MIMO-DWPT COFDM system with beamforming using QPSK modulation

The findings from *Fig. 5* highlight several key observations:

1. Antenna Configuration Effect: Increasing the number of antennas, especially in 4×4 MIMO systems, significantly impacts system performance. MMSE-SIC consistently outperforms ZF-SIC, particularly in configurations with more antennas.

2. BER Performance at Low E_b / N_0 : Both ZF-SIC and MMSE-SIC perform similarly at low E_b/N_0 values, with BER curves converging.

3. Increasing Gap at Higher E_b / N_0: As E_b / N_0 increases, MMSE-SIC maintains a substantial performance advantage over ZF-SIC, especially in high signal-to-noise ratio conditions.

At E_b / N_0 of 10 dB, ZF-SIC achieves a BER of approximately 9.8×10^{-5}, while MMSE-SIC achieves a lower BER of about 7×10^{-5}. This performance gap persists at higher E_b / N_0 levels, such as 12 dB, where ZF-SIC reaches a BER of roughly 2.6×10^{-5}, and MMSE-SIC excels with a BER of approximately 2.6×10^{-6}. This indicates a significant 62.4% improvement in BER when transitioning from ZF-SIC to MMSE-SIC.

These results consistently highlight the superiority of MMSE-SIC over ZF-SIC across all scenarios, emphasizing the benefits of higher antenna configurations. Additionally, the study suggests that Binary Phase Shift Keying (BPSK) outperforms Quadrature Phase Shift Keying (QPSK) for both equalization techniques. This insight is valuable for optimizing MIMO-DWPT COFDM systems with beamforming. In summary, these findings offer practical guidance for enhancing wireless communication system performance, considering equalization techniques, antenna configurations, and modulation schemes.

4. Conclusion

This study conducted a thorough performance analysis of a MIMO-DWPT COFDM system with adaptive beamforming, considering various modulation schemes and equalization techniques in a Rayleigh fading channel. Key findings include the importance of increased antennas for improved SNR and a BER of 10^{-4}. BPSK modulation with MMSE-SIC detector proved superior, achieving lower BER at an E_b / N_0 of 12 dB. These results highlight the significance of antenna configuration and modulation scheme selection in designing robust MIMO-DWPT COFDM systems for wireless communication.

References

[1] Kumar, S., & Payal, "Design of multi-beam free space optical communication system for mitigation of atmospheric and geometric nonlinearities", *Journal of Optics*, *50*, pp. 664–670, 2021.

[2] Bindle, A., Gulati, T., & Kumar, N., "Exploring the alternatives to the conventional interference mitigation schemes for 5G wireless cellular communication network", *International Journal of Communication Systems*, *35*(4), e5059, 2022.

[3] Sah, M. B., Bindle, A., & Gulati, T., "Analysis of Different Interference Mitigation Techniques Based on Bit Error Rate (BER) in 5G", in *International Conference on Innovative Computing and Communications*, 2022, Springer, Singapore, pp. 295–311.

[4] Kumar, S., "Enhancing Performance of FSO Communication Link using Coherent Optical OFDM with Cascaded EDFA", in *2020 5th International Conference on Communication and Electronics Systems (ICCES)*, 2020, June, IEEE, pp. 349–355.

[5] Chide, N., Deshmukh, S., Borole, P. B., & Chore, N. "An overview of OFDM variants and their applications", *International Journal of Electronics Communication and Computer Engineering*, *4*(2), pp. 47–51, 2013.

[6] Hamamreh, J. M., Hajar, A., & Abewa, M., "Orthogonal frequency division multiplexing with subcarrier power modulation for doubling the spectral efficiency of 6G and beyond networks", *Transactions on Emerging Telecommunications Technologies*, *31*(4), pp. 1–18, 2020.

[7] Sabna, N., "An improved channel estimation for underwater acoustic OFDM and COFDM systems", *International Journal of Systems, Control and Communications*, *12*(3), pp. 264–278, 2021.

[8] Zhao, W., Liao, Q., Huang, D., & Guo, Y., "Performance analysis of the satellite-to-ground continuous-variable quantum key distribution with orthogonal frequency division multiplexed modulation", *Quantum Information Processing*, *18*(1), pp. 1–22, 2019.

[9] Guerreiro, J., Dinis, R., & Campos, L. "On the achievable capacity of MIMO-OFDM systems in the CathLab environment", *Sensors*, *20*(3), pp. 1–16, 2020.

[10] Kumar, S., "Performance Investigation of MIMO Based CO-OFDM FSO Communication Link for BPSK, QPSK and 16-QAM under the Influence of Reed Solomon Codes", *Journal of Engineering & Technological Sciences*, *53*(5), pp. 1–16, 2021.

[11] Payal, Sharma, D., & Kumar, S., "Performance enhancement of multi-beam FSO communication system with the application of MIMO with CO-OFDM", *Journal of Optics*, *52*(2), pp. 810–817, 2023.

[12] Rawat, A., Kaushik, R., & Tiwari, A., "An Overview of MIMO OFDM System for Wireless Communication", *International Journal of Technical Research & Science*, VI (X), pp. 1-4, 2021.

[13] Liao, Y., Hua, Y., & Cai, Y., "Deep learning based channel estimation algorithm for fast time-varying MIMO-OFDM systems", IEEE Communications Letters, 24(3), pp. 572-576, 2019.

[14] Jallouli, K., Mazouzi, M., Diguet, J. P., Monemi, A., & Hasnaoui, S., "MIMO-OFDM LTE system based on a parallel IFFT/FFT on NoC-based FPGA", *Annals of Telecommunications*, 77, pp. 689–702, 2022.

[15] Locharla, G. R., Mahapatra, K. K., & Ari, S., "Variable length mixed radix MDC FFT/IFFT processor for MIMO-OFDM application" *IET Computers & Digital Techniques*, *12*(1), pp. 9–19, 2018.

[16] Garbo, G., & Mangione, S., "An improved detection technique for cyclic-prefixed OFDM", *Journal of Networks*, *5*(7), pp. 759–765, 2010.

[17] Kaur, J., & Sharma, V. "A-STBC incorporated power-efficient Radio over Fibre system", *Optics Communications*, *441*, pp. 84–89, 2019.

[18] Akho-Zahieh, M. M., & Abdellatif, N., "Performance Analysis of MIMO Wavelet Packet Multicarrier Multicode CDMA System with Antenna Selection", *Advances in Electrical and Electronic Engineering*, *17*(4), pp. 423–435, 2019.

[19] Nguyen, T., Hoang, P., Ngo, Q., Dang, K., & Huu, P. N., "Wavelet packet transform in MIMO-OFDM systems", in *2013 Third World Congress on Information and Communication Technologies (WICT 2013), IEEE*, 2013, December, pp. 98–101.

[20] Arjun, K. R., & Surekha, T. P., "Peak-to-Average Power Ratio reduction in Wavelet based OFDM using Modified Selective Mapping for Cognitive Radio Applications", *Walailak Journal of Science and Technology (WJST)*, *18*(12), 19814–12, 2021.

[21] Kulkarni, V. M., & Bhalchandra, A. S., "Peak-to-average power ratio reduction in wavelet packet modulation Using Binary Phase Sequence", *Indonesian Journal of Electrical Engineering and Computer Science*, *14*(3), pp. 1210–1219, 2019.

[22] Sadouki, B. R., & Djebbouri, M., "Application of wavelet packet in the MIMO-COFDM systems", *IOSR Journal of Electronics and Communication Engineering (IOSR-JECE)*, *9*(2), pp. 60–67, 2014.
[23] Tören, M., & Çiflikli, C., "Peak-to-Average-Power-Ratio (PAPR) reduction methods with wavelet transform in MIMO-OFDM", *IETE Journal of Research*, *64*(3), pp. 415–421, 2018.
[24] Kashyap, R., & Bagga, J., "Equalization techniques for MIMO systems in wireless communication: A review", *International Journal of Engineering and Advanced Technology (IJEAT)*, *3*(5), pp. 260–264, 2014.
[25] Jain, A., Shukla, P., & Tharani, L., "Comparison of various equalization techniques for MIMO system under different fading channels", in *2017 2nd International Conference on Communication and Electronics Systems (ICCES)*, 2017, October, pp. 913–918.
[26] Misra, J., & Mandal, R., "A Review on Various Approaches to Reduce Intersymbol Interference in MIMO–OFDM System", *IJIRST–International Journal for Innovative Research in Science & Technology*, *1*(8), pp. 74-78, 2015.
[27] Prayitno, R. A., Wirastuti, N. M. A. E. D., & Hartawan, I. G. A. K. D. D., "Analisis Unjuk Kerja Zero Forcing Equalizer Pada Sistem OFDM MIMO", *Jurnal SPEKTRUM*, *5*(1), pp. 68–73, 2018.
[28] Girija, S. P., & Rao, R., "Fractional weighted ZF equalizer: A novel approach for channel equalization in MIMO-OFDM system under impulse noise environment", *Communications in Science and Technology*, *6*(1), pp. 1–10, 2021.
[29] Ida, Y., Ahn, C. J., Matsumoto, T., & Matsufuji, S., "Achievement accurate CSI for AF relay MIMO/OFDM based on complex HTRCI pilot signal with enhanced MMSE equalization", *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, *98*(11), pp. 2254–2262, 2015.
[30] Grover, H., & Pathania, R., "Performance Analysis of MMSE Equalizer and Time-Frequency Block Code for MIMO-OFDM Under Single-RF System", in *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, *IEEE*, 2018, April, pp. 1277–1281.
[31] Rahman, M. M., Alam, M. S., Islam, S., Samy, M. A. A., & Al Islam, N., "Cooperative MIMO OFDM system based on Amplify and Forward Relay: Evaluation of ZF-SIC and MMSE-SIC equalization", *Przeglad Elektrotechniczny*, *1*(9), pp. 77–81, 2018.
[32] Shrivastava, N., & Trivedi, A., "Combined beamforming with space-time-frequency coding for MIMO–OFDM systems", *AEU-International Journal of Electronics and Communications*, *69*(6), pp. 878–883, 2015.
[33] Varade, S., & Kulat, K., "Space-time Turbo Codes with Adaptive Beamforming for MIMO-OFDM: Performance Analysis & Evaluation", *International Journal of Computer Science Engineering & Technology*, *1*(6), pp. 278-283, 2011.
[34] Vaigandla, K. K., Siluveru, M., & Bolla, S. R., "Analysis of PAPR and Beamforming for 5G MIMO-OFDM", *International journal of analytical and experimental modal analysis*, XII (X), pp. 483-490, 2020.
[35] Jung, G. W., Kim, H. K., & Lee, Y. H., "Zero-forcing beamforming with user grouping in spatially correlated channel environments", in *2020 International Conference on Computing, Networking and Communications (ICNC)*, *IEEE*, 2020, February, pp. 1–5.
[36] Ammar, H. A., Adve, R., Shahbazpanahi, S., & Boudreau, G., "Statistical analysis of downlink zero-forcing beamforming", *IEEE Wireless Communications Letters*, *9*(11), pp. 1965–1969, 2020.
[37] Kim, C. K., "Performance of MIMO-OFDM systems with multi-beamforming based on MMSE", *Journal of the Institute of Electronics Engineers of Korea TC*, *48*(6), pp. 6–13, 2011.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 45–59

DOI: 10.2478/auseme-2023-0004

Electrical Power Generator Faults Analysis Using Fault Tree and Bayesian Network

Toufik TOUIL[1,2], Abdelaziz LAKEHAL[3]

[1] LGMM laboratory, Department of Mechanical Engineering, August 20, 1955 University, P.O. Box 26, El-Hadiek Road Skikda 21000, Algeria. e-mail: toufik.touil@bakerhughes.com
[2] Baker Hughes, Algerian Engineering Services Company, 148 chemin de wilaya, Blida, Algeria.
[3] Laboratory of Research on Electromechanical and Dependability, University of Souk Ahras, Algeria, e-mail: a.lakehal@univ-soukahras.dz

Manuscript received March 16, 2023; revised October 10, 2023

Abstract: This paper presents a model to predict Electrical Power Generator (EPG) faults. The fault tree (FT) model is developed and used to help maintenance engineers in fault analysis procedure of this rotating machine. By identifying the main, intermediate and basic events it's possible to construct the FT with logical reasoning. The top dreaded event is defined. By using a Bayesian network (BN) as a complementary tool, fault prediction of the EPG becomes possible and easy. By using the developed BN, the probability of occurrence of the top event (EPG failure) is calculated. Also, by this approach, we can process complex information that causes system faults in an easy and simple way. The essential elements to do this analysis are the reliable and good exploitation of the information previously stored in the system. The use of the BN in combination with the FT gives the possibility of qualitative and quantitative analysis, diagnosis, and prediction of faults from the same Bayesian model. The flexibility of the proposed BN model in this paper allows better and precise decision making. Also, priorities regarding maintenance job are defined and resources are a priori prepared.

Keywords: Bayesian network, fault tree, electrical generator, fault prediction, information, maintenance priorities.

1. Introduction

Electric Power Generator (EPG) is used in many industrial and energy sectors. However, it can be prone to breakdowns. Faults of these rotating machines can cause serious damage to the economic sector, and in some cases hinder the development of society. Therefore, it is necessary to understand the root causes of EPG faults and providing an effective reference to prevent them. To do this,

this article presents a solution for identifying faults and taking preventive actions with the objective to prevent and minimize the risks of faults by using the fault tree (FT) and the Bayesian Network (BN). The second objective of the proposed solution is to pre-evaluate generator faults in an accurate and efficient manner. The fault prediction plays an important role in the safety systems of power plants and reduces maintenance costs, electricity production losses and increases the life cycle of EPGs.

Previous studies have conducted and proposed various solutions to improve the efficiency of EPGs, identify risks, and assess their impacts on this equipment.

Firstly, traditional analysis methods are used by researchers, such as Failure Mode and Effect Criticality Analysis (FMECA) [1,2], FT analysis [3], and Ishikawa diagram method [4]. However, these methods remain of limited use because they can be costly, not flexible to accommodate unforeseen changes in systems, also provide a false sense of security by overlooking human errors that can contribute to EPG faults. Other studies proposed artificial intelligence methods for industrial systems faults analysis, such as: artificial neural networks for fault diagnosis [5], fuzzy logic [6], and support vector machine [7]. Other authors have used artificial techniques for optimizing EPG operation and improve its availability such as: particle swarm optimization [8] and genetic algorithms [9]. These artificial intelligence techniques are given a strong contribution in the mastering of the EPG function, and they have improved the fault diagnosis and prediction when combined with traditional methods.

Several previous works in the literature show that hybrid fault diagnosis techniques based on traditional techniques and artificial intelligence methods allow better decision making and high performance when they are used together. Akhtar and Kirmani combine the operational failures with fuzzy logic in a Fuzzy Fault Tree model [10]. The contribution shows that the proposed model allows a good assessment of reliability and provides excellent fault analysis. Two traditional methods: FT and FMECA are combined to another traditional method to determine critical component in a diesel generator [11], they have also been utilized for conducting quantitative analyses of biogas plants by integrating them with recursive operability analysis, with a focus on elucidating the roles of procedural errors and component failures [12]. In other contributions, such as a maintenance plan for a turbine of hydroelectric power plant, Reliability-Centered Maintenance (RCM) was combined with FMECA [13] to better understand all potential failure scenarios that can affect industrial machines, while ensuring their reliability. Also, two artificial intelligence methods: genetic algorithm and fuzzy logic are combined to provide an expert system with the capability of anomaly detection and to allow the system to expose network problems autonomously [14].

In addition, some studies have integrated classical analysis methods and BNs such as: FT analysis and BN [15], Fuzzy Fault Tree with BN [16] and FMECA with BN [17]. In the application side there are few research works that discuss the fault diagnosis of EPG systems with BN. In this article FT is combined with a BN to resolve the problematic of fault prediction of these power systems. A mapping of the FT will be shown and a qualitative analysis will be made by the development of the BN structure and finally we try to give a quantitative analysis by inference in the developed BN.

This article is organized as follows: section 2 presents the EPG protection systems to allow a better understanding of the relationship between the system components. In section 3, a brief review of various methods and tools used for diagnosis and prediction of EPG faults, is presented. An application of our approach on EPG is presented in section 4. Finally, some recommendations and conclusions are provided at the end of this paper.

2. Generator protection systems description

Power generation stations play a vital role in human life due to the importance of energy in industry and daily life. To mitigate recurrent system failures and power supply interruptions that result in significant economic losses, many industrial companies in this sector have resorted to innovation and development of control and protection systems for electrical generators.

These systems are specifically designed to prevent various faults that may occur during operation, utilizing a control program equipped with advanced sensors that facilitate the detection of faults and enable the immediate implementation of appropriate decisions.

Most EPGs are rotating machines created and developed with different technologies over time, used to produce electric energy on the one hand and to operate systems (industrial machines) on the other hand, which may be necessary for their proper functioning. In addition, they have several protection systems (*Fig.1*), including the lubricating oil system, cooling and ventilation system, control system, and it is also equipped with sensors that record various parameters such as temperature, vibration, etc. Due to their working conditions, EPGs are subject to unexpected failures that affect their components such as the rotor, bearings, etc.

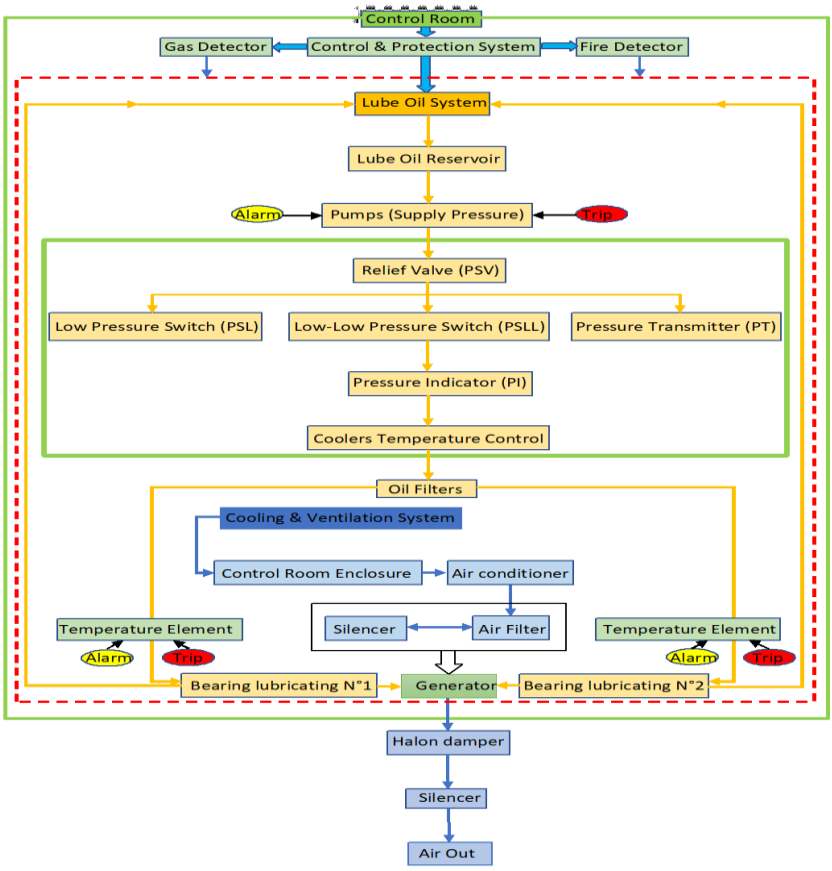


Figure 1: Schematic diagram of a generator protection system.

To ensure providing a reliable supply of clean and fresh lubrication oil to the generator bearings and removing heat generated by friction, a reliable lubrication oil system is essential for the generator operation. Also, during the generation of electrical energy, heat is produced by eddy currents and Joule losses and by aerodynamic and mechanical friction. This heat must be dispersed to maintain the efficiency of the generator. The cooling and ventilation systems of the generator dissipate heat by cooling the generator and ventilating. Ventilation system is intended to pressurize the generator enclosure to prevent the infiltration of any combustible leakage gas, and to provide cooling airflow through the enclosure. In addition to these systems, a control system includes everything related to the control of voltage, temperature, synchronization, protection relays, and vibrations monitoring.

3. Research tools and methodology

There are several commonly used methods for fault and risk analysis, including the following examples: FT analysis, Failure Mode and Effects Analysis (FMEA), Functional Safety Analysis, Quantitative Risk Analysis using BN, etc. The methods used and presented in the rest of this paper are FT and BN.

A. FT Analysis

FT analysis is one of the methods developed for systematic fault assessment. FT is used to represent graphically the events that can lead to a system failure, as it is the main reference in the study of industrial machines faults. The FT is based mainly on defining the set of primary events and basic events sequentially, which in turn lead to the occurrence of the undesirable event specified in this study as the fault of the EPG. The connection between the various specified events is carried out by logical gates of type "AND" and "OR", which are used to represent the dependency relationships between the events.

B. BN analysis

BN is a probabilistic analysis tool or graphical probabilistic model that allows the representation of cause-and-effect relationships between events that lead to a fault of the system. BN is a powerful tool for modeling complex systems that involve multiple random variables and conditional dependencies between them, allowing for the calculation of the posterior probability of an event based on the prior probability of its causes.

In a BN, each node represents an event, and each link represents a cause-and-effect relationship between events. The probabilities of each event are updated based on the available information (previously stored information). The components of a BN include nodes, arcs, conditional probabilities, conditional probability tables.

The Bayes theorem is expressed as a formula in probability theory, utilized for the calculation of conditional probabilities. This method systematically adjusts the probabilities in order to obtain more information that may help us avoid total failure of the systems. The Bayesian approach has been widely applied in all fields. From the Bayesian formula, we have (*Fig. 2*):

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \tag{1}$$

where P(A∩B) represents the probability of the intersection of events A and B, i.e., the probability that events A and B occur simultaneously, then

$$P(A \cap B) = P(A) \cdot P(B/A). \tag{2}$$

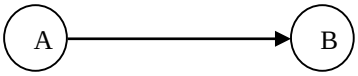


Figure 2: Simple Bayesian network

Substituting (2) into (1), we get:

$$P(A/B) = \frac{P(B/A)\cdot P(A)}{P(B)} \quad (3)$$

$P(A/B)$ is the probability of event A given event B has occurred.
$P(B/A)$ is the probability of event B given event A has occurred.
$P(A)$ is the prior probability of event A.
$P(B)$ is the marginal probability of event B.

Let $A1, A2, \dots, An$ be the possible causes of B (*Fig.3*).

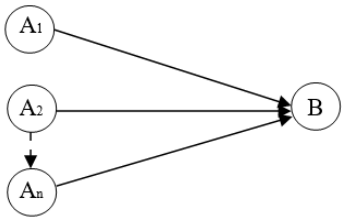


Figure 3: Simple Bayesian network for multiple events and a single effect

Then:

$$P(B) = P(A_1) \cdot P(B/A_1) + P(A_2) \cdot P(B/A_2) + \cdots + P(A_n) \cdot P(B/A_n) \quad (4)$$

or equivalently,

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B/A_i)$$

C. Mapping of the FT into a BN

The mapping of the FT into the BN is based on their identical graphical representation. The second reason is that the two tools are given a strong contribution in fault analysis. The common element is the modeling of the relationships between faults that are represented by variables. The FT constructs by nodes, arcs, and logical gates (AND and OR) to display the interdependent relationships between events and their causal effects, while the BN are used to model the probabilistic relationships between system variables by using nodes and arcs to define the top event.

Also, a BN is based on a conditional probability table to calculate posterior probability of each variable during the quantitative phase. One conditional probability table defines quantitatively the relationship between events, but FT analysis requires two formulas to calculate each logic gates. In the case of new information, updating a BN is easier than a fault tree which requires redoing the all calculation throughout the tree. The combination of these methods makes it possible to obtain a more precise and complete probabilistic analysis of the faults to improve the system reliability.

4. Practical application on an electrical power generator

This section proposes a practical application of the proposed fault analysis approach; the main objective is to determine the degradation indicators of an EPG installed on the power plant production of Boufarik unit in northern Algeria. This plant is made up of three gas turbines totaling an installed capacity of 704.129 MW, and it was commissioned in 2016. It has been connected to the monitoring center of Algerian Electricity Production Company (AEPC) since April 2018.

In the maintenance activities, inspections are made and several repair works are carried out to address issues such as: visual inspection and NDT inspection of all components and replacing destroyed components with new ones. To interpret these found results, this paper presents a concrete example of probabilistic analysis using a combined approach based on FT and BN. It explains how to apply this approach to find out the causes that lead to the EPG faults, and enables us to predict the probability of occurrence of the top event, helps us in simplifying and understanding the results obtained easily, knowing the weaknesses areas in the system, and help taking actions to improve the reliability of the EPG.

Table 1: Basic faults and events of electric generators

Machine	Faults (D)	Causes (C)		
ELECTRICAL GENERATOR	Vibration (D1)	Bearing assembly fault (C1)		
		Vibration monitoring system (C2)	Faulty vibration probe (C21)	
			The monitor (C22)	
		Imbalance fault (C3)		
		Coupling problem (C4)		
	Generator overheating (D2)	Fire (C5)		
		Filters clogged (C6)		
		Cooling fan failure (C7)		
		Dirty air inlet screens (C8)		
		Internal air passage clogged (C9)		
	High temperature accompanied by vibrations (D3)	Electrical protection relays (C10)	Synchronization check (C101)	
			Generator differential (C102)	
			Power return (C103)	
			Loss of excitement (C104)	
			Time over current relay (C105)	
			Over- / underfrequency relay (C106)	
			Overvoltage relay(C107)	
			Undervoltage relay (C108)	
			Ground overvoltage relay (C109)	
		Rotor cooling air holes clogged (C11)		
		Shaft cooling fan broken (C12)		
		Voltage regulator problem (C13)		
	High temperature at the bearings (D4)	Lubrication system (C14)	Oil quality (C141)	
			Motor (C142)	
			Pumps (C143)	Auxiliary pump (C1431)
				Mechanical pump (C1432)
				Emergency pump (C1433)
			Oil circuit (Piping) (C144)	
			Oil pressure (C145)	High pressure (C1451)
				Low pressure (C1452)
		Dirty oil inlet filters (C15)		
		Faulty temperature sensor (C16)		

In *Fig. 4*, the FT model is established where EPG fault is the top event. Four common faults are intermediate events ($D1, D2, D3$ and $D4$). Also, Ci $(i = 1, \dots, 16)$ code is used to represent basic events.

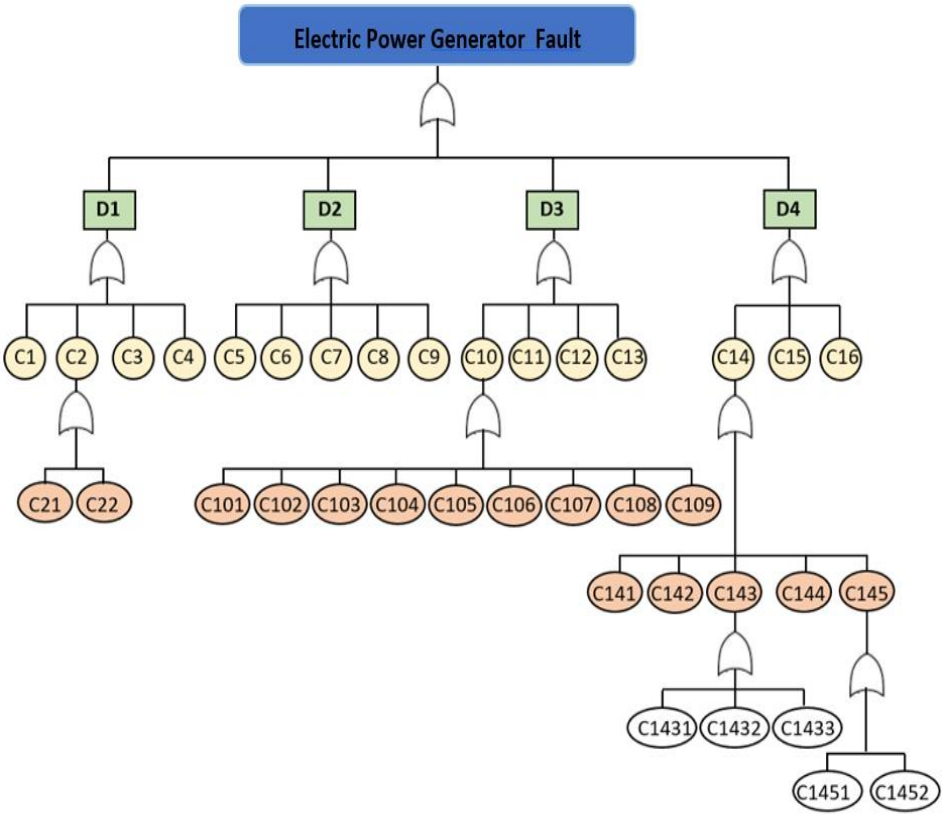


Figure 4: Fault tree diagram of an electrical generator

To complete the transposition of the FT in the space of probabilities, the following parameters must also be provided:

- If the cause D has no direct cause, $P(D)$ will be defined. In the case where the cause D takes two states: true and false, we have to define the probabilities of the two logical values $P(D = True)$ and $P(D = False)$.
- Also, If the effect EPG has a single direct cause D, we have to define $P(G/D)$, i.e., the four values $P(EPG = T/D = T), P(EPG = T/D = F), P(EPG = F/D = T), P(EPG = F/D = F)$.
- If the effect EPG has two direct causes $D1$ and $D2$ we have to define $P(EPG/D1, D2)$, that is to say the eight values:

$P(EPG = T/D1 = T, D2 = T), P(EPG = T/D1 = T, D2 = F), P(EPG = T/\ D1 = F, D2 = T), P(EPG = T/D1 = F, D2 = F), P(EPG = F/D1 = T, D2 = T), P(EPG = F/D1 = T, D2 = F), P(EPG = F/D1 = F, D2 = T), P(EPG = F/D1 = F, D2 = F).$

How to compute the probability of the main event $D1$, which has the secondary causes $C1, C2, C3$ and $C4$? First, we compute the probabilities of all events produced by secondary causes for example $C2$ which in turn contains two secondary causes $C21$ and $C22$, whose values are given in the Table 2.

$$\begin{aligned} P(C2 = T) &= P(C2/C21, C22) \\ &= P(C2 = T/C21 = T, C22 = T).P(C21 = T).P(C22 = T) \\ &+P(C2 = T/C21 = T, C22 = F).P(C21 = T).P(C22 = F) \\ &+P(C2 = T/C21 = F, C22 = T).P(C21 = F).P(C22 = T) \\ &+P(C2 = T/C21 = F, C22 = F).P(C21 = F).P(C22 = F) \end{aligned} \quad (5)$$

$= (1 \times 0.0003 \times 0.0002) + (1 \times 0.0003 \times 0.9998) + (1 \times 0.9997 \times 0.0002) + (0 \times 0.9997 \times 0.9998) = 0.00000006 + 0.00029994 + 0.00019994 + 0 \approx 0.0005.$

Therefore, the value of $P(C2)$ is 0.0005.
Having $C1, C2, C3$ and $C4$, we can easily compute the value of $D1$, that is,

$$\begin{aligned} P(D1 = T) &= P(D1/C1, C2, C3, C4) = \\ &= P(D1{=}T/C1{=}T, C2{=}T, C3{=}T, C4{=}T)\cdot P(C1{=}T)\cdot P(C2{=}T)\cdot P(C3{=}T)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}T, C2{=}T, C3{=}T, C4{=}F)\cdot P(C1{=}T)\cdot P(C2{=}T)\cdot P(C3{=}T)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}T, C2{=}T, C3{=}F, C4{=}T)\cdot P(C1{=}T)\cdot P(C2{=}T)\cdot P(C3{=}F)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}T, C2{=}T, C3{=}F, C4{=}F)\cdot P(C1{=}T)\cdot P(C2{=}T)\cdot P(C3{=}F)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}T, C2{=}F, C3{=}T, C4{=}T)\cdot P(C1{=}T)\cdot P(C2{=}F)\cdot P(C3{=}T)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}T, C2{=}F, C3{=}T, C4{=}F)\cdot P(C1{=}T)\cdot P(C2{=}F)\cdot P(C3{=}T)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}T, C2{=}F, C3{=}F, C4{=}T)\cdot P(C1{=}T)\cdot P(C2{=}F)\cdot P(C3{=}F)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}T, C2{=}F, C3{=}F, C4{=}F)\cdot P(C1{=}T)\cdot P(C2{=}F)\cdot P(C3{=}F)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}F, C2{=}T, C3{=}T, C4{=}T)\cdot P(C1{=}F)\cdot P(C2{=}T)\cdot P(C3{=}T)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}F, C2{=}T, C3{=}T, C4{=}F)\cdot P(C1{=}F)\cdot P(C2{=}T)\cdot P(C3{=}T)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}F, C2{=}T, C3{=}F, C4{=}T)\cdot P(C1{=}F)\cdot P(C2{=}T)\cdot P(C3{=}F)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}F, C2{=}T, C3{=}F, C4{=}F)\cdot P(C1{=}F)\cdot P(C2{=}T)\cdot P(C3{=}F)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}F, C2{=}F, C3{=}T, C4{=}T)\cdot P(C1{=}F)\cdot P(C2{=}F)\cdot P(C3{=}T)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}F, C2{=}F, C3{=}T, C4{=}F)\cdot P(C1{=}F)\cdot P(C2{=}F)\cdot P(C3{=}T)\cdot P(C4{=}F) \\ &+ P(D1{=}T/C1{=}F, C2{=}F, C3{=}F, C4{=}T)\cdot P(C1{=}F)\cdot P(C2{=}F)\cdot P(C3{=}F)\cdot P(C4{=}T) \\ &+ P(D1{=}T/C1{=}F, C2{=}F, C3{=}F, C4{=}F)\cdot P(C1{=}F)\cdot P(C2{=}F)\cdot P(C3{=}F)\cdot P(C4{=}F) \\ &\approx 0.0019 \end{aligned} \quad (6)$$

In the same way, we calculate the remaining causes ($D2, D3$ and $D4$). The difficulty arises from the expanding number of combinations for which the probabilities need to be defined.

Table 2: A priori and a posteriori probabilities of the basic events ($C1-C16$)

Basic Events			A priori probabilities	A posteriori Probabilities		Probability of failure
C1			0.0009	0.0009		D1= 0.0019
C2	C21		0.0003	0.0005		
	C22		0.0002			
C3			0.0002	0.0002		
C4			0.0003	0.0003		
C5			0.0001	0.0001		D2=0.0016
C6			0.0004	0.0004		
C7			0.0006	0.0006		
C8			0.0003	0.0003		
C9			0.0002	0.0002		
C10	C101		0.0001	0.0017		D3=0.0026
	C102		0.0002			
	C103		0.0003			
	C104		0.0001			
	C105		0.0002			
	C106		0.0004			
	C107		0.0004			
	C108		0.0005			
	C109		0.0002			
C11			0.0001	0.0001		
C12			0.0003	0.0003		
C13			0.0005	0.0005		
C14	C141		0.0003	0.0009	0.0039	D4= 0.0054
	C142		0.0006			
	C143	C1431	0.0008	0.0014		
		C1432	0.0001			
		C1433	0.0005			
	C144		0.0009	0.0009		
	C145	C1451	0.0003	0.0007		
		C1452	0.0004			
C15			0.0009	0.0009		
C16			0.0006	0.0006		

Four primary variables ($D1, D2, D3$ and $D4$) represent the factors that conduct to the top event (electric power generator fault), each of these variables is connected to other variables in hierarchical order. In addition, the BN contains a

quantitative description of the behavior of the variables, which are expressed using a posteriori probability of the prior data previously stored by the experts.

From the Table 2, by using the values obtained from the four events calculated previously ($D1, D2, D3$ and $D4$), the top event can be calculated as follows:

$$P(EPG) = P(EPG/D1, D2, D3, D4) = \sum_{i=1}^{4} P(EPG/Di).P(Di), \quad (7)$$

therefore

$$P(EPG) = P(EPG/D1).P(D1) + P(EPG/D2).P(D2) + P(EPG/D3).P(D3) + P(EPG/D4).P(D4)$$

From relationship (4), we have

$$P(EPG/Di).P(Di) = P(EPG \cap Di), for\ i = 1,..,4. \quad (8)$$

From Table 2, we have

$$P(EPG \cap D1) = 0.0019, P(EPG \cap D2) = 0.0016, P(EPG \cap D3) = 0.0026$$

and $P(EPG \cap D4) = 0.0054$.

It follows from these results that

$$P(EPG) = 0.0019 + 0.0016 + 0.0026 + 0.0054 = 0.0115.$$

From the BN of *Fig. 5*, we have calculated that the probability of occurrence of the undesirable event, which is the shutdown of EPG, is 1.2%. This probability is temporarily acceptable in terms of quantity, but from an economic point of view, and also given the importance of electricity in the life of the population, it is necessary to address the sources of the faults before they create other problems, by identifying the weaknesses and taking corrective measures to reduce the probabilities of faults. Therefore, we must return to the FT to find out the causes of top impact on the EPG, using Table 2, which gives us an approximate view of the probability of each event occurring (quantitative description of each event). The contribution of each of the main events ($D1, \ldots, D4$) to the total failure rate can be expressed in percent.

For instance, the probability of 0.0019 of the first main event "$D1$" represents a 17% contribution to the total failure rate of 1.15%. According to the results shown in Table 2, we conclude that the prediction of occurrence of a posteriori probability of bearings ($D4$) due to the high temperature and lubrication system, the most important factor as it ranks first in EPG fault approximately with 49%. High temperature accompanied by vibrations ($D3$) also plays a significant role, but to a lesser extent as they contribute about 23% to EPG fault. EPG overheating

($D2$) and vibration ($D1$) contribute less to the system fault, by rates of 14% and 17% respectively.

Due to the importance and strategic role of the EPG in the community, in order to improve its operational safety, the maintenance team is required to reduce the occurrence probability of the top events. This is done by focusing on the main events, such as reducing vibrations and temperature increases at the bearings, by conducting preventive checks to reduce them, such as checking the lubricating oil system (pumps, filters, and pipes). Also, it is important to take corrective actions for other events that are less harmful to the system. By making these corrections, it is possible to raise the efficiency of the EPG and reduce the exorbitant maintenance costs, in addition to increasing the energy production, and this is what the power units aspire for.

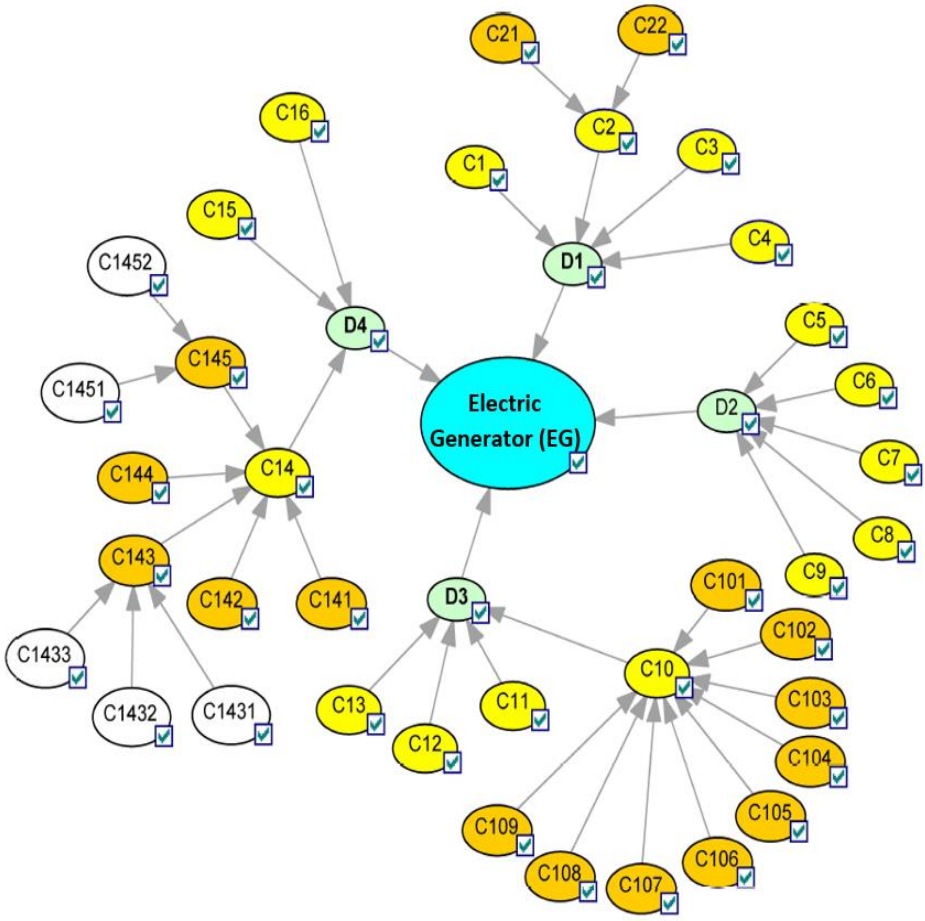


Figure5: Bayesian network of electric power generator fault

5. Conclusion

In this study, a probabilistic analysis of EPG faults was performed, firstly using FT analysis to know all the main and secondary causes that could lead to EPG fault. After listing all faults in Table 1 based on initial data collected by operators and maintenance experts, by the proposed approach of integrating the FT and BN it becomes easy for us to calculate all the a posteriori probability leading to the prediction of occurrence of the EPG fault. The results obtained, shown in Table 2, indicate that EPG failure is primarily attributed to the high temperature at the bearings ($\boldsymbol{D4}$), accounting for approximately 49% of failures. Subsequently, high temperature accompanied by vibrations ($\boldsymbol{D3}$) is identified as another significant factor, contributing to failure at an approximate rate of 23%. In contrast, generator overheating ($\boldsymbol{D2}$) and vibrations ($\boldsymbol{D1}$) are recognized as causative factors at lower rates, estimated at 14% and 17%, respectively. Based on the results obtained, the proposed method simplifies complex industrial models, especially those that are challenging for failure analysis. This method helps to identify fragile branches in the FT, which is crucial to understanding the main causes of system failure. Also, it allows the maintenance team to evaluate potential faults and take preventive measures to avoid or reduce them. In conclusion, the goal of this approach is to improve the reliability of the EPG and of the entire system.

References

[1] Alrifaey, M., Sai Hong, T., Supeni, E. E., As' Arry, A., and Ang, C. K. "Identification and prioritization of risk factors in an electrical generator based on the hybrid FMEA framework". *Energies.*, vol. 12, no. 4, p. 649, 2019.

[2] Mohanty, J K., Dash, P. R., and Pradhan, P. K.,"FMECA analysis and condition monitoring of critical equipments in super thermal power plant", *International Journal of System Assurance Engineering and Management.*, vol. 11, pp. 583–599, 2020.

[3] Kang, J., Sun, L., and Soares, CG., "Fault Tree Analysis of floating offshore wind turbines", *Renewable Energy*, vol. 133, pp. 1455-1467, 2019.

[4] Wilarso, W., Azharul, F., Pahmi, M. A., Mugisidi, D., and Ilman, K. A., "Analysis of engine piston damage to the generator set using the fishbone analysis method", *AIP Publishing LLC.* in *AIP Conference Proceedings.*, vol. 2578, no. 1, p. 070004, Nov. 2022.

[5] Ferrero Bermejo, J., Gómez Fernández, J. F., Olivencia Polo, F., and Crespo Márquez, A., "A review of the use of artificial neural network models for energy and reliability prediction. A study of the solar PV, hydraulic and wind energy sources", *Applied Sciences.*, vol. 9, no. 9, p. 1844, 2019.

[6] Fernández, J. C., Corrales, L. B., Benítez, I. F., and Núñez, J. R., "Fault Diagnosis of Combustion Engines in MTU 16VS4000-G81 Generator Sets Using Fuzzy Logic: An Approach to Normalize Specific Fuel Consumption" in *Intelligent Computing Systems: 4th International Symposium, ISICS 2022, Santiago, Chile, March 23–25, 2022, Proceedings.*, pp. 17-29, Cham: Springer International Publishing. March, 2022.

[7] Tuerxun, W., Chang, X., Hongyu, G., Zhijie, J., and Huajian, Z., "Fault diagnosis of wind turbines based on a support vector machine optimized by the sparrow search algorithm", *IEEE Access.*, vol. 9, pp. 69307–69315, 2021.

[8] Ferencz, J., and Kelemen, A., "Particle Swarm Optimization of a Hybrid Energy Storage System" *Acta Universitatis Sapientiae, Electrical and Mechanical Engineering.,* vol.13, no.1, pp. 82–100. Dec 2021.

[9] Das, B. K., Hassan, R., Tushar, M. S. H., Zaman, F., Hasan, M., and Das, P. "Techno-economic and environmental assessment of a hybrid renewable energy system using multi-objective genetic algorithm: A case study for remote Island in Bangladesh", *Energy Conversion and Management.*, vol. 230, p. 113823. 2021.

[10] Akhtar, I., and Kirmani, S., "An application of fuzzy fault tree analysis for reliability evaluation of wind energy system", *IETE Journal of Research.*, vol. 68, no. 6, pp. 4265–4278, 2022.

[11] Wakijo, W., Purnomo, J., and Susilo, R., "FTA and FMECA analysis for determine critical components of diesel generator cumminsKTA 38D", in *STTAL postgraduate-international conference*, vol. 6, no. 1, Oct. 2022.

[12] Barozzi, M., Contini, S., Raboni, M., Torretta, V., Casson Doreno, V., and Copelli, S., "Integration of recursive Operability Analysis, FMECA and FTA for the quantitative biogas plants: Role of procedural errors and components failures*", Journal of Loss Prevention in the Process Industries.*, vol. 71, p.104468, 2021.

[13] Martinez-Monseco, F J., "An approach to a maintenance plan for a turbine of hydroelectric power plant. Optimisation based in RCM and FMECA analysis", *Journal of Applied Research in Technology and Engineering.*, vol. 2, no. 1, pp. 39–50, 2021.

[14] Hamamoto, A. H., Carvalho, L. F., Sampaio, L. D. H., Abrão, T., and Proença Jr, M. L. "Network anomaly detection system using genetic algorithm and fuzzy logic", *Expert Systems with Applications*, vol. 92, pp. 390–402, 2018.

[15] Shang, P., Dong, H., Li, X., and Ren,W., "Research on fault diagnosis method of 750kV substation based on Bayesian network and fault recording information fusion", *Journal of Physics: Conference Series, IOP*., vol. 1550, no. 5, p. 052020, 2020.

[16] Yanfu, W., and Min, X., "Approach to Integrate Fuzzy Fault Tree with Bayesian Network", *Procedia engineering.*, vol. 45, pp. 131–138, 2012.

[17] Brahim, I. B., Addouche, S. A., Mhamedi, A. E., and Boujelbeneet,Y., "Build a Bayesian network from FMECA in the production of automotive parts: diagnosis and prediction*". IFAC-Papers online.*, vol. 52, no. 13, pp. 2572–2577, 2019.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 60–73

DOI: 10.2478/auseme-2023-0005

Comparative Rheological Investigation of Nanocomposites of Surface Charged Superparamagnetic Iron Oxide Nanoparticles with Polyethylene Glycol

Taraneh JAVANBAKHT

Department of Computer Science, University of Quebec in Montreal, 201 President Kennedy Street, Montreal, Quebec Canada, e-mail: javanbakht.taraneh@courrier.uqam.ca

Manuscript received January 21, 2023; revised, June 07, 2023

Abstract: This paper focuses on a new investigation of the rheological properties of the nanocomposites of surface charged superparamagnetic iron oxide nanoparticles with polyethylene glycol. Both studied nanocomposites showed the steady-state behavior at 20 °C and 40 °C. Moreover, the increase of viscosity versus shear strain, shear rate or time for the nanocomposites was small at 60 °C. The effect of the coating of nanoparticles with the polymer was comparable for different nanocomposites. The data presented in this paper can provide the required information for the preparation of assemblies of nanocomposites with polymers.

Keywords: Rheological properties, nanocomposites, surface charged SPIONs, PEG, mechanical properties.

1. Introduction

Rheological investigation of nanomaterials is an important issue as these materials with various mechanical properties can be applied in diverse fields of science and engineering. Moreover, the surface functionalization of these materials can modify these properties such as viscosity, shear rate, etc. The viscous nature and solid-like behavior of these materials have made them important candidates at different levels with low filler loading and high concentration [1-3]. The rheological properties of these materials or their composites can determine their processing performance for their preparation. It is worth noting that their rheological properties can be modified according to their weight change when a few amount of these materials is increased. In each polymer composite, one or several nanomaterials are mixed with one or several polymers. Computer modeling and simulation are two methods for the investigation of the rheological properties of these materials. In these methods,

the mechanisms at the molecular level are explored for improving the dispersion of nanoparticles in matrices. In addition, the information on the effects of nanoparticles on the chain conformation and glass transition temperature of the samples is searched [4-5].

The rheological properties of a polymer matrix and the presence of branched structures in its structure are affected by its molecular weight [6-8]. These behaviors also depend on the values of shear rate. The stiffening of polymer chains and agglomeration as well as the increase in the viscosity of nanocomposites can happen because of the increase of the nanomaterials loading in the polymeric matrix at low shear rate values due to the confinement of the polymer chains with the embedded nanofiller [9-12]. At high shear, the differences between nanocomposites and corresponding matrices become less significant due to several phenomena such as the shear thinning behavior, the wall slip phenomenon and the unaffected material viscosity after changing the nanomaterial content in the samples. It has been explored that the nanocomposites with high molar mass and low melt flow index could show a decrease in their viscosity values in the high shear rate region. Moreover, the nanocomposites that contain a small amount of nanomaterials can have lower viscosity than that of the unfilled matrix due to the higher viscosity of the polymeric matrix at high shear stress. During recent years, several nanofillers with different chemical nature, shape and morphology have been investigated among which are the nanoparticles such as metals, carbon-based fillers and ceramics [13-18].

Polymer nanocomposites containing nanosized fillers having large surface areas are considered as better materials than other composites with microsized fillers. To maximize these enhancements, fillers should be well-dispersed in nanocomposites [19]. There are several techniques for the improvement of the quality of filler dispersions in these nanocomposites such as intercalation of polymers from solution, *in situ* intercalation, or melt intercalation [20-22]. The quality of filler dispersions can also be improved with the use of compatibilizers [23-25], nanofiller surface treatments [26-28] or the application of an electric field to clay nanocomposites [29-30].

The comparative analysis of the rheological properties of the nanocomposites of different surface charged superparamagnetic iron oxide nanoparticles (SPIONs) coated with polyethylene glycol (PEG) has not been reported, yet. The results of this article can be used for the preparation improvement and applications of these nanocomposites in science and engineering.

2. Materials and methods

PEG (MW= 8000) was purchased from Sigma Aldrich. SPIONs were synthesized as described previously [31]. For the preparation of bare SPIONs, an aqueous solution of ferrous chloride (5 mL, 0.045 M) and ferric chloride (0.0375 M) was added to diethyleneglycol (250 mL). Sodium hydroxide was added to the mixture, then the mixture, with the final concentration of 0.375 M, was heated at 170 °C during 15 minutes. After maintaining its temperature at the same temperature during an hour, it was cooled to 60 °C. The collection of SPIONs was performed using a neodymium magnet, then the nanoparticles were washed with a nitric acid solution (1 M) [31]. For the synthesis of negatively charged SPIONs, 3-(triethoxysilyl)propylsuccinic anhydride (TEPSA) (14.2 mL, 50 mmol) were added to nanoparticles. The solvent was DMF (100 mM of iron in 100 mL). After the addition of water (8.6 mL) and TMAOH (5 mL, 1M) at room temperature, the solution was heated at 100 °C during 24 hours. After the addition of acetone/ether (50/50), SPIONs were precipitated and collected with a magnet, washed with acetone and dispersed in water. Then, excess of additives was removed by filtration using a membrane with a cut-off of 30 kDa. [31] For the preparation of positively charged SPIONs, N-[3-(trimethoxysilyl)propyl] ethylenediamine (TPED) was grafted onto SPIONs with the addition of TPED (25 mmol, 5.4 mL) to a suspension of nanoparticles (100 mL, concentration of iron: 25 mM) at 50 °C. The mixture was stirred for 2 hours, then cooled at room temperature. For the filtration of suspension, a membrane with a cut-off of 30 kDa was used, then SPIONs was centrifuged at 16 500 g for 45 minutes [31].

The nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONs-PEG were prepared as explained in the previous work [31]. After mixing PEG (4.8 g) with deionized water during 15 minutes at room temperature, four portions of the polymer mixture were prepared: there was no SPIONs in one portion, the three other portions were considered for the preparation of nanocomposites of the polymer with these nanoparticles, each of them having 1.2 g of the polymer, which was dissolved in water (5 mL). The solutions of nanocomposites (1%) were prepared by adding 12 mg of bare, positively charged or negatively charged SPIONs separately to each of the three portions of the PEG solution. The samples were separately mixed during 15 minutes at room temperature [31].

The samples of each nanocomposite were prepared and analyzed in triplicate. An Anton Paar MCR-302 rheometer was used for the rheological investigation of nanocomposites. The measurements in triplicate were performed at 20 °C, 40 °C and 60 °C. The analysis of parameters such as mean values, standard deviations and statistical significance was performed with the QtiPlot software [32-33].

3. Results and Discussion

Fig. 1 shows the viscosity variations of the nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONs-PEG versus shear rate.

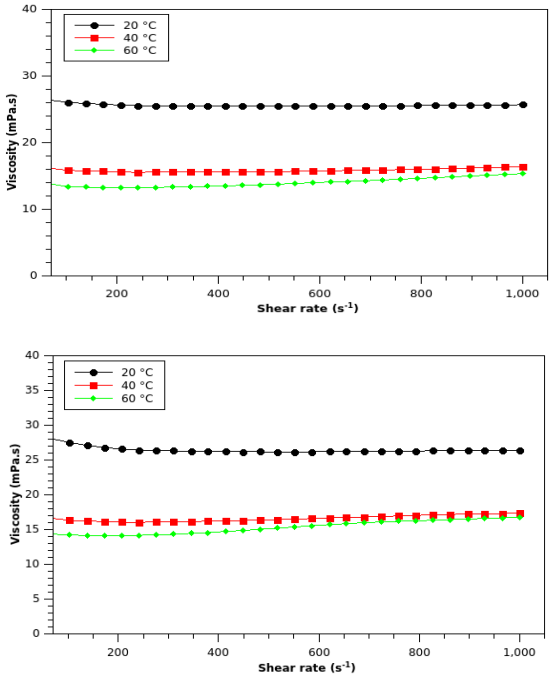


Figure 1: Viscosity (mPa·s) of the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG versus shear rate

The viscosity of the nanocomposites was constant with shear rate at 20 °C and 40 °C, but it had a small increase at 60 °C. As expected, for all the samples the drop in their viscosity values was observed when the temperature increased.

Fig. 2 shows the viscosity variations of the nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONs-PEG versus shear strain at 20 °C, 40 °C and 60 °C. As shown in this figure, the similarity in the behavior or each type of nanocomposite with the same type of nanocomposite in *Fig. 1* was observed. In other words, both nanocomposites showed a steady-state behavior at 20 °C and 40 °C, but their viscosity increased a bit versus shear strain at 60 °C.

Fig. 3 shows the viscosity variations of the nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONS-PEG versus time at 20

°C, 40 °C and 60 °C. The small increase of viscosity versus time at 60 °C and the steady-state behavior of the nanocomposites was observed similar to the results presented in the previous figures.

The changes of torque versus shear strain for both nanocomposites at 20 °C, 40 °C and 60 °C are shown in *Fig. 4*. The torque values increased with the increase of the shear strain of the samples. As shown in this figure, the torque did not increase with the same slope at different temperatures, as it was higher at 20 °C, but it decreased when the temperature increased to 40 °C and 60 °C. In other words, less torque values were observed for both nanocomposites when the temperature increased.

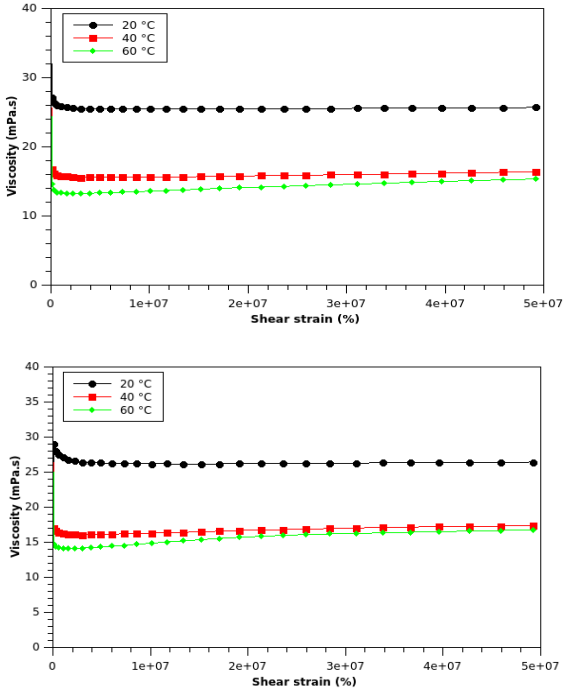


Figure 2: Viscosity of the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG versus shear strain

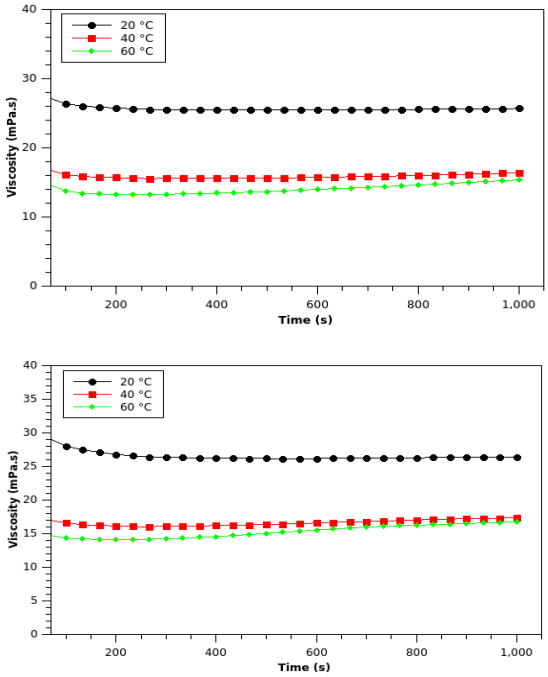


Figure 3: Viscosity of the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG versus time

As observed in the previous figures, the data for 40 °C and 60 °C were close to each other. Therefore, we can conclude that the rheological properties of both nanocomposites changed significantly from 20 °C to 40 °C and changed a bit with the increase of temperature.

The changes of torque versus time at 20 °C, 40 °C and 60 °C are shown in *Fig. 5* for both nanocomposites. The torque increased with time with a constant slope for each nanocomposite. A linear change of torque versus time was observed at each temperature. In other words, the increase of torque was constant with time for the nanocomposites. Therefore, no change in the torque increase was applied over time on each nanocomposite. So, the constant increase of torque versus time was expected.

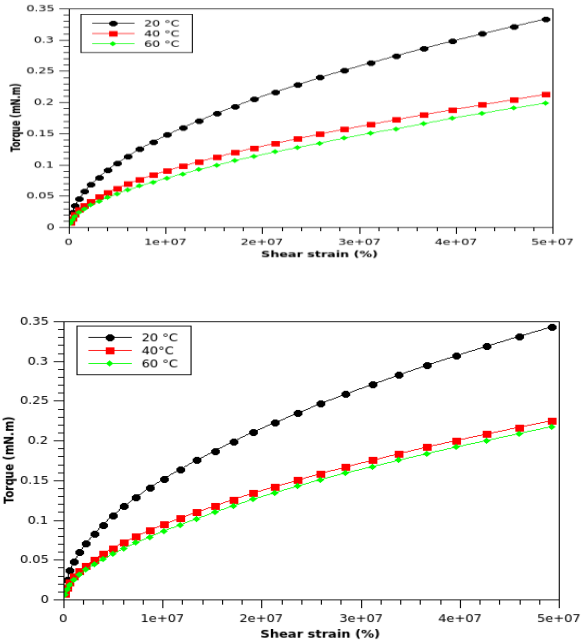
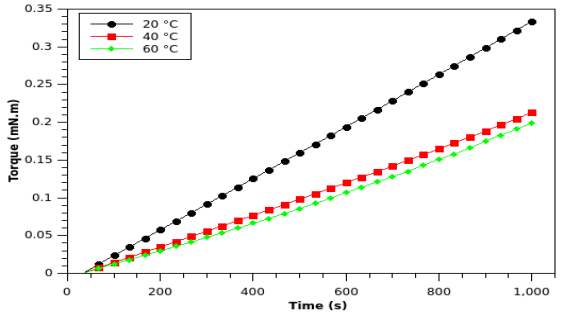
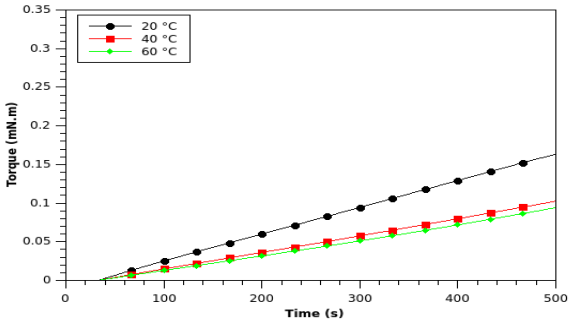


Figure 4: Torque versus shear strain for the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG



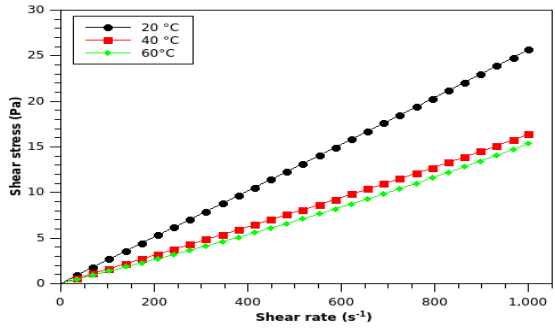


b)

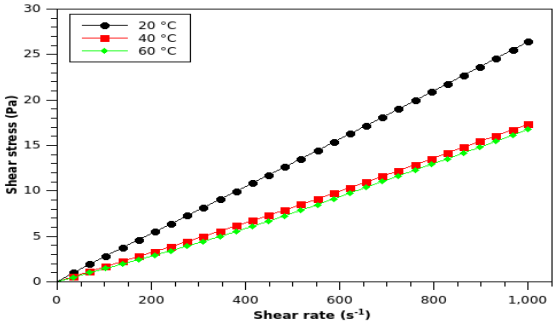
Figure 5: Torque versus time for the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG

Fig. 6 shows the shear stress values versus shear rate at 20 °C, 40 °C and 60 °C for the nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONs-PEG.

As shown in *Fig. 6*, the changes of the shear stress applied on the samples were linear and the slope of the increase of shear stress with shear rate was the same for each nanocomposite at 40 °C and 60 °C, which indicated the same constant rate of shear stress at these temperatures when it changed with shear rate. Moreover, the slope was higher at 20 °C, which showed that the change of the shear stress applied on each nanocomposite of SPIONs with PEG versus shear rate was higher at a lower temperature.

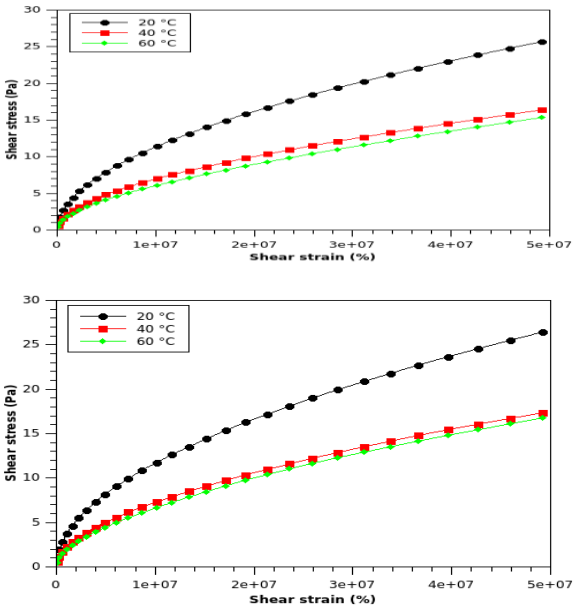


a)



b)

Figure 6: Shear stress of the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG versus shear rate



a)

b)

Figure 7: Shear stress of the nanocomposites of a) positively charged SPIONs-PEG and b) negatively charged SPIONs-PEG versus shear strain

Fig. 7 shows the shear stress applied on the nanocomposites of positively charged SPIONs-PEG and negatively charged SPIONs-PEG versus shear strain at 20 °C, 40 °C and 60 °C.

For both nanocomposites non-linear changes were observed for the shear

stress. Moreover, the shear stress values increased with the increase of the shear strain. Almost the same slopes were observed for the samples for the increase of shear stress with shear strain at 40 °C and 60 °C, which indicated that they had nearly the same increase rate of the shear stress versus shear strain, but the slope observed in the figure was higher at 20 °C, which showed that a higher change of the shear stress of each nanocomposite versus shear strain at a lower temperature.

In this paper, the variations of the shear stress with shear rate and shear strain were investigated for explaining how these parameters changed for the nanocomposites of positively surface charged SPIONs-PEG and negatively surface charged SPIONs-PEG. As observed, although shear stress increased with shear rate for both nanocomposites, their viscosity values versus shear rate, shear strain or time did not show a significant change at different temperatures.

The molecular structure of surface charged SPIONs have been investigated previously with amine and carboxyl groups on the surface of positively charged and negatively charged SPIONs, respectively [34]. In the current research work, the similarity in the results obtained for the samples showed that the difference in their surface charge did not have a significant impact on their rheological properties. This rheological investigation was performed for providing the information required for the determination of the properties of these nanocomposites and their further improvement in future studies.

Several studies have been performed on nanomaterials [35-39], biomaterials [40-44] and construction materials [45-46] with diverse applications in science and engineering. Some of these materials have shown non-Newtonian behavior maintainable with their preparation in polymeric matrices [47-48]. The optimization of the properties of these materials can provide important information for the improvement of their mechanical properties for further applications. The Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is an appropriate method that has been widely applied for the optimization and prediction of various characteristics [49-54]. Some investigations on the nanocomposites of iron oxide nanoparticles with polymers have shown that these materials would be appropriate for the function improvement of electrical devices [55-56]. The present study can provide new insight to the investigation of their properties for the manufacture and optimization of these devices.

4. Conclusion

This paper aimed to investigate the rheological properties of surface-charged SPIONs that have found their diverse applications in science and engineering during recent years. The samples of two types of nanocomposites, positively

surface charged SPIONs-PEG and negatively surface charged SPIONs-PEG, were investigated. The results in this study showed that the rheological properties of the nanocomposites did not depend on their surface charge but depended on temperature, because they changed with the temperature increase. The steady-state behavior of both nanocomposites was observed at 20 °C and 40 °C. Moreover, a small increase of viscosity versus shear strain, shear rate or time for the samples was observed at 60 °C. The shear stress increased with shear rate and shear strain in all the measurements with higher slopes at 20 °C, which decreased with the increase of temperature. The results of this investigation can be used for the preparation improvement of these nanocomposites as well as the correlation of their physical and mechanical properties.

Acknowledgements

The author gratefully thanks Professor Ingo Salzmann for the provision of PEG and Professor Sophie Laurent and Dr. Dimitri Stanicki for the synthesis of SPIONs.

References

[1] Pelto, J., Heino, V., Karttunen, M., Rytöluoto, I., and Ronkainen, H., "Tribological performance of high density polyethylene (HDPE) composites with low nanofiller loading", *Wear*, vol. 460, 203451, Nov. 2020.

[2] Choudhury, A., Bhowmick, A. K., Ong, C., and Soddemann, M., "Effect of various nanofillers on thermal stability and degradation kinetics of polymer nanocomposites", *Journal of Nanoscience and Nanotechnology*, vol. 10, no. 8, pp. 5056–5071, Aug. 2010.

[3] Wu, W., Liu, L., Goksen, G., Demir, D., and Shao, P., "Multidimensional (0D-3D) nanofillers: Fascinating materials in the field of bio-based food active packaging", *Food Research International*, vol. 157, 111446, Jul. 2022.

[4] Chevigny, C., Dalmas, F., Di Cola, E., Gigmes, D., Bertin, D., Boué, F., and Jestin, J., "Polymer-grafted-nanoparticles nanocomposites: Dispersion, grafted chain conformation, and rheological behavior", *Macromolecules*, vol. 44, no. 1, pp. 122–133, Dec. 2011.

[5] Peng, W., Ranganathan, R., Keblinski, P., Akcora, P., and Ozisik, R., "Viscoelastic and dynamic properties of polymer grafted nanocomposites with high glass transition temperature graft chains", *Journal of Applied Physics*, vol. 126, no. 19, 195102, Nov. 2019.

[6] Arrigo, R., and Malucelli, G., "Rheological behavior of polymer/carbon nanotube composites: An overview", *Materials*, vol. 13, no. 12, 2771, Jun. 2020.

[7] Dobkowski Z., "Application of rheological techniques for investigations of polymer branched structures", *Fluid Phase Equilibria*, vol. 152, no. 2, pp. 327–336, Oct. 1998.

[8] Wang, L., Jing, X., Cheng, H., Hu, X., Yang, L., and Huang, Y., "Rheology and crystallization of long-chain branched poly(l-lactide)s with controlled branch length", *Ind. Eng. Chem. Res.*, vol. 51, no. 33, pp. 10731–10741, Jul. 2012.

[9] Nuryawan A., Abdullah, C. K., Hazwan, C. M., Olaiya, N. G., Yahya, E. B., Risnasari, I., Masruchin, N., Baharudin, M. S., Khalid, H., and Khalil, H. P. S. A., "Enhancement of oil

palm waste nanoparticles on the properties and characterization of hybrid plywood biocomposites", *Polymers*, vol. 12, no. 5, 1007, Apr. 2020.

[10] Zahedi, M., Tabarsa, T., Ashori, A., Madhoushi, M., and Shakeri, A., "A comparative study on some properties of wood plastic composites using canola stalk, Paulownia, and nanoclay", *Journal of Applied Polymer Science*, vol. 129, no. 3, Dec. 2013.

[11] Karak, N., "Fundamentals of nanomaterials and polymer nanocomposites", chapter 1, *Nanomaterials and Polymer Nanocomposites: Raw Materials to Applications*, Elsevier, pp. 1–45, 2019.

[12] Yu, W., Wang, J., and You, W., "Structure and linear viscoelasticity of polymer nanocomposites with agglomerated particles", *Polymer*, vol. 98, pp. 190–200, Aug. 2016.

[13] Akca, E., and Gursel, A., "A review on the matrix toughness of thermoplastic Materials", *Periodicals of Engineering and Natural Sciences*, vol. 3, no. 2, pp. 1–8, Aug. 2015.

[14] Lanfant, N. P., and Alglave H. L., "*Manufacturing process of a thermoplastic material part incorporating metal fillers*", Patent FR3111585A1, Jun. 2020.

[15] Mhike, W., Focke, W. W. and Asante, J. K. O., "Rotomolded antistatic and flame-retarded graphite nanocomposites", *Journal of Thermoplastic Composite Materials*, vol. 31, no. 4, 089270571771263, Ju. 2017.

[16] Kim, K.-W., Kim, D.-K., Han, W., and Kim, B.-J., "Comparison of the characteristics of recycled carbon fibers/polymer composites by different recycling techniques", *Molecules*, vol. 27, 5663, Sep. 2022.

[17] Yang, B., "*Swelling of carbon nano-filler modified polydimethylsiloxane*", PhD thesis, University of Alberta, 2018.

[18] Antunes, R. A., de Oliveira, M. C. L., Ett, G., and Ett. W., "Carbon materials in composite bipolar plates for polymer electrolyte membrane fuel cells: A review of the main challenges to improve electrical performance", *Journal of Power Sources*, vol. 196, no. 6, pp. 2945–2961, Mar. 2011.

[19] Akpan, E. I., Shen, X., Wetzel, B., and Friedrich, K., "Design and synthesis of polymer nanocomposites, Polymer Composites with Functionalized Nanoparticles, Synthesis", *Properties and Applications Micro and Nano Technologies*, Elsevier, pp. 47–83, 2019.

[20] Shen, Z., Simon, G. P., and Cheng, Y.-B., "Comparison of solution intercalation and melt intercalation of polymer–clay nanocomposites", *Polymer*, vol. 43, no. 15, pp. 4251–4260, Jun. 2002.

[21] Chen, Y., Bai, W., Chen, J., Chen, X., Zhao, J., Wei, F., Jian, R., Zheng, X., and Xu, Y., "In-situ intercalation of montmorillonite/urushiol titanium polymer nanocomposite for anti-corrosion and anti-aging of epoxy coatings", *Progress in Organic Coatings*, vol. 165, 106738, Apr. 2022.

[22] Di Y., Iannace, S., Maio, E. D., and Nicolais, L., "Nanocomposites by melt intercalation based on polycaprolactone and organoclay", *Journal of Polymer Science Part B: Polymer Physics*, vol. 41, pp. 670–678, Feb. 2003.

[23] Mittal, V., and Chaudhry, A. U., "Effect of amphiphilic compatibilizers on the filler dispersion and properties of polyethylene—thermally reduced graphene nanocomposites", *Journal of Applied Polymer*, vol. 132, 42484, Jun. 2015.

[24] Martínez-Gómez, A., Quiles-Díaz, S., Enrique-Jimenez, P., Flores, A., Ania, F., Gómez-Fatou, M. A., and Salavagione, H. J., "Searching for effective compatibilizing agents for the preparation of poly(ether ether ketone)/graphene nanocomposites with enhanced properties", *Composites Part A: Applied Science and Manufacturing*, vol. 113, pp. 180–188, Oct. 2018.

[25] Patti, A., Acierno, D., and Russi, P., "Influence of filler dispersion and interfacial resistance on thermal conductivity of polypropylene/carbon nanotubes systems", *Proceedings*, vol. 1914, 030014, Dec. 2017.

[26] Wood, W., "*Processing, wear, and mechanical properties of polyethylene composites prepared with pristine and organosilane-treated carbon nanofibers*", PhD theseis, Washington State University, Dec. 2012.
[27] Musanje, L., and Ferracane, J. L., "Effects of resin formulation and nanofiller surface treatment on the properties of experimental hybrid resin composite", *Biomaterials*, vol. 25, no. 18, pp. 4065–4071, Aug. 2004.
[28] Amini, M., Hasheminejad, K., and Montazeri, A., "Engineering the shape memory parameters of graphene/polymer nanocomposites through atomistic simulations: On the effect of nanofiller surface treatment", *Smart Materials and Structures*, vol. 31, 025010, Dec. 2021.
[29] Ock, H. G., Ahn, K. H., and Lee, S. J., "Effect of electric field on polymer/clay nanocomposites depending on the affinities between the polymer and clay", *Journal of Applied Polymer Science*, vol. 133, 43582, Mar. 2016.
[30] Rozynek, Z., Silva, S. M. D. L., Fossum, J. O., da Silva, G. J., de Azevedo, E. N., Mauroy, H., and Plivelic, T. S., "Organoclay polypropylene nanocomposites under different electric field strengths", *Applied Clay Science*, vol. 96, pp. 67–72, Jul. 2014.
[31] Javanbakht, T., Laurent, S., Stanicki, D., and David, E., "Related physicochemical, rheological, and dielectric properties of nanocomposites of superparamagnetic iron oxide nanoparticles with polyethyleneglycol", *Journal of Applied Polymer Science*, vol. 137, no. 3, pp. 48280–48289, Aug. 2019.
[32] Tamhane, D., and Anantharaman M. R., "Design and fabrication of a simple and inexpensive measurement probe for the evaluation of thermal conductivity of nanofluids", *Nanofluids*, vol. 6, no. 2, pp. 390–396, 2017.
[33] Belosi, F., Ferrari, S., Poluzzi, V., Santachiara, G., and Prodi, F., "Comparison between two different nanoparticle size spectrometers", *Journal of the Air and Waste Management Association*, vol. 63, no. 8, pp. 918–925, Aug. 2013.
[34] Javanbakht, T., Laurent, S., Stanicki, D., and Frenette, M., "Correlation between physicochemical properties of superparamagnetic iron oxide nanoparticles and their reactivity with hydrogen peroxide", *Canadian Journal of Chemistry*, vol. 98, no. 10, pp. 601–608, Oct. 2020.
[35] Javanbakht, T., Hadian, H., and Wilkinson, K. J., "Comparative study of physicochemical properties and antibiofilm activity of graphene oxide nanoribbons", *Journal of Engineering Sciences*, vol. 7, no. 1, pp. C1–C8, Apr. 2020.
[36] Javanbakht, T., "Investigation of rheological properties of graphene oxide and its nanocomposite with polyvinyl alcohol", *Ukrainian Journal of Mechanical Engineering and Materials Science*, vol. 7, no. 1-2, pp. 23–32, Apr. 2021.
[37] Javanbakht, T., and David, E., "Rheological and physical properties of a nanocomposite of graphene oxide nanoribbons with polyvinyl alcohol", *Journal of Thermoplastic Composite Matererials*, vol. 35, no. 5, 0892705720912767, Mar. 2020.
[38] Javanbakht, T., Laurent, S., Stanicki, D., and Salzmann, I., "Rheological properties of superparamagnetic iron oxide nanoparticles", *Journal of Engineering Sciences*, vol. 8, no. 1, pp. C29–C37, Jul. 2021.
[39] Farahnaky, A., Dadfar, S. M. M., and Shahbazi, M., "Physical and mechanical properties of gelatin–clay nanocomposite", *Journal of Food Engineering*, vol. 122, pp. 78–83, Feb. 2014.
[40] Zhuravkov, M., and Romanova, N., "Determination of physical and mechanical properties of biomaterials on base of the nanoindentation technologies and fractional order models", *Russian Journal of Biomechanics*, vol. 20, no. 1, pp. 5–22, Mar. 2016.
[41] Javanbakht, T., Ghane-Motlagh, B., and Sawan, M., "Comparative study of antibiofilm activity and physicochemical properties of microelectrode arrays", *Microelectronic Engineering*, vol. 229, 111305, May 2020.

[42] Ghane-Motlagh, B., Javanbakht, T., Shoghi, F., Wilkinson, K. J., Martel, R., and Sawan, M., "Physicochemical properties of peptide-coated microelectrode arrays and their *in vitro* effects on neuroblast cells", *Materials Science and Engineering C*, vol. 68, pp. 642–650, Nov. 2016.

[43] Javanbakht, T., and Sokolowski, W., "Thiol-ene/acrylate systems for biomedical shape-memory polymers, Shape memory polymers for biomedical applications", Sawston, Cambridge: Woodhead Publishing, chapter 8, pp. 157–166, 2015.

[44] Sukaryo, S. G., Purnama, A., and Hermawan, H., "Structure and Properties of Biomaterials, Biomaterials and Medical Devices", Springer, pp. 1–22, 2016.

[45] Djavanbakht, T., Carrier, V., André, J. M., Barchewitz, R., and Troussel, P., "Effets d'un chauffage thermique sur les performances de miroirs multicouches Mo/Si, Mo/C et Ni/C pour le reyonnement X mou", *Journal de Physique IV*, vol. 10, no. 10, pp. 281–287, Sep. 2000.

[46] Kuzmin, S. A., Egorova, A. D., Krasilnikov, D. A., and Emelianov, Z. V., "Durability of construction materials modified by polymeric additives", *Procedia Structural Integrity*, vol. 20, 2019, pp. 278–283.

[47] Hojjat, M., Etemad, S. G., Bagheri, R., and Thibault, J., "Rheological characteristics of non-Newtonian nanofluids: Experimental investigation", *Int. Commun. Heat Transf*, vol. 38, no. 2, pp. 144–148, Feb. 2011.

[48] Chhabra, R. P., Richardson, J. F., "Non-Newtonian Flow and Applied Rheology: Engineering Applications", Butterworth-Heinemann: Oxford, UK, 2011.

[49] Jamshed, W., Eid, M. R. Aissa, A., Mourad, A., Nisar, K. S., Shahzad, F., Saleel, C. A., and Vijayakumar, V., "Partial velocity slip effect on working magneto non-Newtonian nanofluids flow in solar collectors subject to change viscosity and thermal conductivity with temperature", *PLOS ONE*, vol. 16, no. 11, e0259881, Nov. 2021.

[50] Pawanr, S., Tanishk, T., Gulati, A., Garg, G. K., and Routroy, S., "Fuzzy-TOPSIS based multi-objective optimization of machining parameters for improving energy consumption and productivity", *Procedia CIRP*, 102, 2021, pp. 192–197.

[51] Javanbakht, T., and Chakravorty, S., "Prediction of human behavior with TOPSIS", *Journal of Fuzzy Extension and Applications*, vol. 3, no. 2, pp. 109–125, Apr. 2022.

[52] Midilli, Y. E., and Parshutin, S., "Design of experiments vs. TOPSIS to select hyperparameters of neural attention models in time series prediction", *Information Technology and Management Science*, vol. 23, pp. 27–34, Dec. 2020.

[53] Chakraborty, S., "TOPSIS and modified TOPSIS: A comparative analysis", *Decision Analytics Journal*, vol. 2, 100021, Mar. 2022.

[54] García, V., Marqués, A. I., Cleofas-Sánchez, L., and Sánchez, J. S., "Model selection for financial distress prediction by aggregating TOPSIS and PROMETHEE rankings", *International Conference on Hybrid Artificial Intelligence Systems*, 2016, pp. 524–535.

[55] Newati, S., Singh, V., and Khan, R. A., "Componential modeling led construction of an enzyme biosensor reinforced with iron oxide nanoparticles onto the self-assembled monolayers", *Journal of Bionanoscience*, vol. 8, no. 1, Feb. 2014.

[56] Saravanakumar, K, Issac, J. S., Dhanaselvam, J., Rajesh, R., Singh, A. B., and Geetha, K. "Fe_3O_4/TiO_2/graphene hybrid nanocomposite to improve the lifespan of distribution transformers", *International Conference on Electronic Circuits and Signalling Technologies*, vol. 2325, 2022, 012016.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 74–85

DOI: 10.2478/auseme-2023-0006

Reactive Magnetron Sputtering Control Based on an Analytical Condition of Stoichiometry

Róbert Rossi MADARÁSZ[1], András KELEMEN[2]

[1] Doctoral School of Applied Mathematics and Applied Informatics, John von Neumann Faculty, Óbuda University, Budapest, e-mail: madarasz.rossi@uni-obuda.hu
[2] Department of Electrical Engineering, Faculty of Technical and Human Sciences, Sapientia-Hungarian University of Transylvania, Tg. Mureș, e-mail: kandras@ms.sapientia.ro
https://orcid.org/0000-0002-6433-2712

Manuscript received June 15, 2023; revised September 15, 2023

Abstract: The conditions for stoichiometric thin film deposition by reactive magnetron sputtering include the existence of a given ratio between the flux of the sputtered metallic atoms and the flux of the reactive gas molecules on the surface of the substrate. To meet this condition, a relationship based on the Berg model is formulated between the partial pressure of the reactive gas, the target coverage, and the sputtering current density. Given that the target coverage can be estimated online from the sputtering voltage, it is possible to create a control structure where the reactive gas partial pressure is controlled to obtain stoichiometric thin film deposition. Simulation results are presented based on the dynamic model of the sputtering process.

Keywords: Reactive magnetron sputtering, Berg model, state estimation, stoichiometry control.

1. Introduction

Reactive Magnetron Sputtering is the process in which a special thin coating layer is formed on a substrate to enhance its mechanical, chemical, and optical properties. During the process, noble gas ions bombard the surface of a target material that together with the applied reactive gas forms the coating compound. This is achieved by applying a negative potential to the target and by placing strong magnets behind it, in order to facilitate higher rate of noble gas ionization.

Amongst many beneficial thin film layers is the photocatalyst titania (TiO_2, rutile and anatase) which is obtained by sputtering Titanium in an Argon and Oxygen mixture in high vacuum conditions, typically at a pressure of fractions of Pascals [1]. TiO_2 is a highly preferred coating material due to its special optical characteristics and biocompatibility [2]. With its relatively high refractive index

of 2.6 [3] and its good transparency to visible light it is suitable for dielectric interference filters [8].

In a multilayer configuration with SiO_2, it forms an effective antireflective coating. This multilayer combination has recently been used to coat Starlink satellites from SpaceX in order to minimize their reflection of sunlight, which had an unwelcome effect in astronomical observations [4].

These special characteristics can only be obtained if the deposited film is stoichiometric, meaning that there is an exact ratio of gas-to-metal during oxide formation. One way to obtain this is by sputtering in the poisoned mode, where the target surface is completely covered with a layer of compound, which is sputtered and delivered to the substrate surface. If higher sputtering rates are required, sputtering in the metallic mode is needed, where the metallic target particles are sputtered and they either react with the reactive gas on their way to, or once they arrive to the substrate surface. Sputtering in the transition mode is also possible but given the highly unstable nature of the process in this mode, feedback control has to be employed i.e. in the form of partial pressure control, or Plasma Emission Monitoring (PEM) [5].

As observed in [6] attention has to be paid to the distance between the target and the substrate in order to facilitate both the sputtering of metal from the target, as well as the formation of oxides on the substrate. This has been shown to be a criterion for stoichiometric deposition of the oxides of titanium.

Determination of the substrate composition is possible with the help of different ex-situ techniques that can be applied only after film formation. These Techniques include Rutherford Backscattering Spectroscopy (RBS), X-Ray Diffraction (XRD), and X-Ray Photoelectron Spectroscopy (XPS) [8].

This paper focuses on a mathematical approach to determining the target coverage during sputtering in an online manner, to facilitate substrate coverage and stoichiometry control.

The following notation has been applied in the paper:

e- the electron charge ($1.6 \cdot 10^{-19}$ $[C]$);

R – ideal gas constant (8314 $\left[\frac{J}{kmolK}\right]$);

N_A – Avogadro's number ($6.023 \cdot 10^{26}$ $\left[\frac{1}{kmol}\right]$);

m_G- mass of the reactive gas molecule $[kg]$ (for O_2, $m_G = 32$ [a. u.] = $= 32 * 1.66 \cdot 10^{-27}[kg]$);

$k_B = 1.38 \cdot 10^{-23}$ $\left[\frac{J}{K}\right]$ - the Boltzmann constant;

T – the absolute temperature ($T = 300$ $[°K]$);

n_{GG} – the number of reactive gas atoms in a gas molecule ($n_{GG} = 2$ in case of the Oxygen);

n_{GM} – the number of reactive gas atoms in a compound molecule ($n_{GM} = 2$ in case of the TiO_2);

K_{sto} - the ratio of reactive gas molecules to metallic atoms needed to form the compound ($K_{sto} = 1$ in case of TiO_2);
V – the volume of the sputtering chamber $[m^3]$. ($V = 0.08\ [m^3]$);
p_G- the partial pressure of the reactive gas (O_2) in the sputtering chamber;
F_G- flux density of the reactive gas molecules (O_2) on the surfaces inside the sputtering chamber;
F_{IT}- flux density of the sputtering ions on the surface of the target (mainly Ar^+);
F_{MS}- flux density of sputtered metal atoms on the surface of the substrate;
θ_T- target coverage i.e. the fraction of the target surface covered by compound molecules;
θ_S- substrate coverage i.e. the fraction of the substrate area covered by compound molecules;
θ_C- the fraction of the condensation area corresponding to the wall of the sputtering chamber covered by compound molecules;
q_{in}- mass flow of the input reactive gas $\left[\frac{kg}{s}\right]$;
q_p- pumping mass flow of the vacuum pump $\left[\frac{kg}{s}\right]$;
A_T- the target area ($A_T = 3600 \cdot 10^{-6}\ [m^2]$);
A_C- the condensation area ($A_C = 16800 \cdot 10^{-6}\ [m^2]$);
A_S- the substrate area ($A_S = 600 \cdot 10^{-6}\ [m^2]$);
η_M- sputtering yield (efficiency) of the metal (number of Ti atoms sputtered by an incident Ar^+ion) ($\eta_M = 0.5$) [7];
η_{MG}- sputtering yield of the metal-gas compound (TiO_2) i.e. the number of compound molecules sputtered by one Ar^+ ion ($\eta_{MG} = 0.017$) [7];
α_{GT}, α_{GS}- sticking coefficients of the O_2 molecule to the metallic fraction of the target and to the metallic fraction of the substrate, respectively (the probability that the atoms of a reactive gas molecule reaching the metallic surface, are trapped in a metal-gas compound molecule) ($\alpha_{GT} = \alpha_{GS} = 1$) [7];
α_{GC} gettering efficiency of the sputtering chamber's surface (of the area covered by sputtered Ti atoms) ($\alpha_{GC} = 0.8$);
N_G- the surface density of the reactive gas atoms trapped within the compound molecules; (in our particular case the simulations have been carried out with a value of $N_G = 19.4 \cdot 10^{18}\ \left[\frac{atoms}{m^2}\right]$, based on the surface density of the compound molecules of $N_{TiO_2} = 9.7 \cdot 10^{18}\ \left[\frac{molecules}{m^2}\right]$ [8];)
I_d –the intensity of the discharge current;
J- the discharge current density on the surface of the target;
$Y_{SEE,M}$ – ion induced secondary electron emission yield of the metal (Ti) ($Y_{SEE,M} = = 0.114$) [9];
$Y_{SEE,MG}$ – ion induced secondary electron emission yield of the surface covered with metal-gas compound (TiO_2) ($Y_{SEE,MG} = 0.08$ [10], [11]);

2. Analytical study of the conditions needed for stoichiometric sputtering

As mentioned in [6] the production of stoichiometric films on the substrate supposes a controlled ratio of incoming metallic and reactive gas particle fluxes. Assuming a Maxwellian velocity distribution, the flux density of the reactive gas molecules on any surface inside the sputtering chamber depends on the partial pressure *p* of the reactive gas according to the well-known relation [12].

$$F_G = \frac{p_G}{\sqrt{2\pi m_G k_B T}} \tag{1}$$

In a simplified approach, we assume that the flux F_{MS} of the metal atoms on the substrate surface is proportional to the flux of the metal atoms outsputtered from the target surface and to the ratio of the condensation and target surface areas. Here, the condensation surface A_C is introduced as a surface at the distance of the substrate from the target, delimited by the solid angle defined by the jet of outsputtered metallic atoms.

The desired ratio of these flux densities can be formulated as

$$\alpha_{GS}(n_{GG}F_G) = F_{MS}K_{sto} \tag{2}$$

The condition of stoichiometry in case of TiO_2 is $K_{sto} = 1$.

This yields the following condition

$$n_{GG}\alpha_{GS}\frac{p_G}{\sqrt{2\pi m_G k_B T}}A_C = F_{IT}\eta_M(1-\theta_T)A_T K_{Sto}. \tag{3}$$

This can further be rearranged to

$$p_G = \frac{1}{n_{GG}\alpha_{GS}}F_{IT}\eta_M(1-\theta_T)\frac{A_T}{A_C}K_{Sto}\sqrt{2\pi m_G k_B T}. \tag{4}$$

The target coverage θ_T is a state variable of the reactive magnetron sputtering process, the dynamic model of which can be formulated as:

$$\begin{cases} \frac{dp_G}{dt} = \frac{R\cdot T}{N_A\cdot m_G\cdot V}(q_{in} - q_p - \\ -m_G(\alpha_{GT}F_G(1-\theta_T)A_T + \alpha_{GC}F_G(1-\theta_C)(A_C - A_S) + \alpha_{GS}F_G(1-\theta_S)A_S)) \\ \frac{d\theta_T}{dt} = \frac{1}{N_G}(n_{GG}\alpha_{GT}F_G(1-\theta_T) - F_{IT}\eta_{MG}\theta_T) \\ \frac{d\theta_S}{dt} = \frac{1}{N_G}\left\{F_{IT}\left(\frac{A_T}{A_C}\right)[\eta_{MG}\theta_T(1-\theta_S) - \eta_M\theta_S(1-\theta_T)] + n_{GG}\alpha_{GS}F_G(1-\theta_S)\right\} \\ \frac{d\theta_C}{dt} = \frac{1}{N_G}\left\{F_{IT}\left(\frac{A_T}{A_C}\right)[\eta_{MG}\theta_T(1-\theta_C) - \eta_M\theta_C(1-\theta_T)] + n_{GG}\alpha_{GC}F_G(1-\theta_C)\right\} \end{cases} \tag{5}$$

This system is similar to [13], [14], but the substrate and condensation surface coverages are handled as separate state variables from the consideration that the

conditions of bond formation on the substrate are different from the conditions on the wall of the chamber.

Thus, the state vector is defined as:

$$\boldsymbol{x} = [p_G, \theta_T, \theta_S, \theta_C]' \tag{6}$$

In order to determine the partial pressure of the reactive gas that satisfies the condition (4) of stoichiometry, the target coverage estimation can rely on the above dynamic model, or as a simplified approach, can be based on the discharge voltage measurement, as presented in Chapter 4.

In steady state the desired operating point can be determined using the Berg model [12].

$$\begin{cases} q_{in} - q_p - -m_G(\alpha_{GT}F_G(1-\theta_T)A_T + \\ +\alpha_{GC}F_G(1-\theta_C)(A_C - A_S) + \alpha_{GS}F_G(1-\theta_S)A_S) = 0 \\ n_{GG}\alpha_{GT}F_G(1-\theta_T) - F_{IT}\eta_{MG}\theta_T = 0 \\ F_{IT}\left(\frac{A_T}{A_C}\right)[\eta_{MG}\theta_T(1-\theta_S) - \eta_M\theta_S(1-\theta_T)] + \\ +n_{GG}\alpha_{GS}F_G(1-\theta_S) = 0 \\ F_{IT}\left(\frac{A_T}{A_C}\right)[\eta_{MG}\theta_T(1-\theta_C) - \eta_M\theta_C(1-\theta_T)] + \\ +n_{GG}\alpha_{GC}F_G(1-\theta_C) = 0 \end{cases} \tag{7}$$

Compared to [12], (7) is modified as mentioned for the dynamic model, but this has no influence on the relation between the target coverage and the partial pressure.

Thus, from the second equation of (7) it results:

$$\theta_T = \frac{n_{GG}\alpha_{GT}F_G}{n_{GG}\alpha_{GT}F_G + F_{IT}\eta_{MG}}. \tag{8}$$

The reactive gas flux density F_G is given in (1) versus the partial pressure, while for determining the sputtering ion flux density we'll consider two approaches.

In a simplified version,

$$F_{IT} \cong \frac{J}{e} = \frac{I_d}{A_T e}, \tag{9}$$

where the discharge current is an easily measurable quantity.

In a more accurate approach, the current due to the ion induced secondary electron emission, and thus the effect of the target coverage, is taken into account.

According to [15],

$$F_{IT} = \frac{J}{e}\frac{1}{1+Y_{SEE,M}(1-\theta_T)+Y_{SEE,MG}\theta_T} \tag{10}$$

Hence, by substituting (10) into the second equation of (7), it results

$$\left(Y_{SEE,M} - Y_{SEE,MG}\right)\theta_T^2 + \left(Y_{SEE,MG} - 2Y_{SEE,M} - 1 - \frac{J}{e}\frac{\eta_{MG}}{n_{GG}\alpha_{GT}F_G}\right)\theta_T + \left(1 + Y_{SEE,M}\right) = 0 \tag{11}$$

Fig. 1 shows a numerical example for the compound TiO_2 ($I_d = 0.7\ [A]$, $J = \frac{I_d}{A_t} = 194.4\ \left[\frac{A}{m^2}\right]$, $\eta_M = 0.5$, $\eta_{MG} = 0.017$). Right- and left-hand sides of (4) are plotted taking into account both versions (9) and (10). Thus, the partial pressure at which the ratio of sputtered metallic and reactive gas particles corresponds to stoichiometry can be determined by the intersection of these curves. It can be noticed that this pressure corresponds to a location on the p-q curve that is very close to the tipping point, where the target tips over to the poisoned state in case of increasing reactive gas flow.

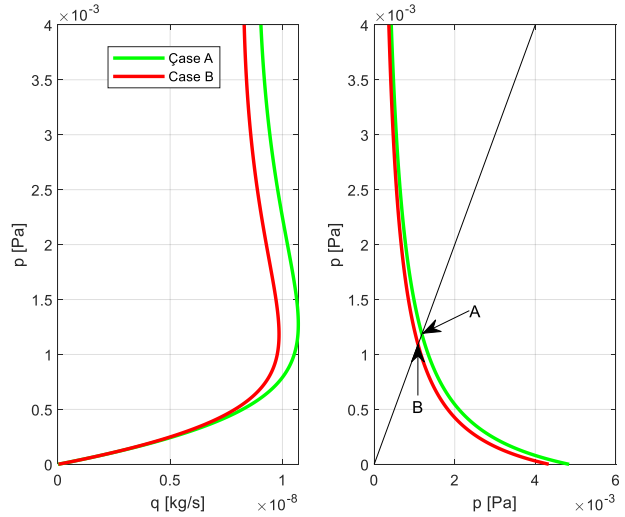


Figure 1: a.) The p-q characteristics of the system modelled in case *A* using (9) and in case *B* using (10)
b.) The graphical solution of (4) in the cases corresponding to (9) and (10)

Since the outsputtered metallic flux from the target, and the reactive gas flux reaching the substrate can be influenced by p and J, it can be assumed that the conditions needed for stoichiometry on the substrate surface can be reached by controlling the target conditions. For this to work, a way of determining and controlling the target coverage is needed. These tasks are covered in Chapter 3 and 4.

3. Target coverage control by manipulating the reactive gas partial pressure

For a given sputtering ion flux density F_{IT}, the relationship between the partial pressure of the reactive gas and target coverage can be found by inserting (1) into (9).

$$\theta_T = \frac{n_{GG}\alpha_{GT}\frac{p_G}{\sqrt{2\pi m_G k_B T}}}{n_{GG}\alpha_{GT}\frac{p_G}{\sqrt{2\pi m_G k_B T}} + F_{IT}\eta_{MG}}, \tag{12}$$

where we assume that F_{IT} is controlled by means of the discharge current in a fast control loop.

It can be seen in *Fig. 2* that the relationship (12) is a monotonic one. This means that the partial pressure can be used to directly influence the target coverage.

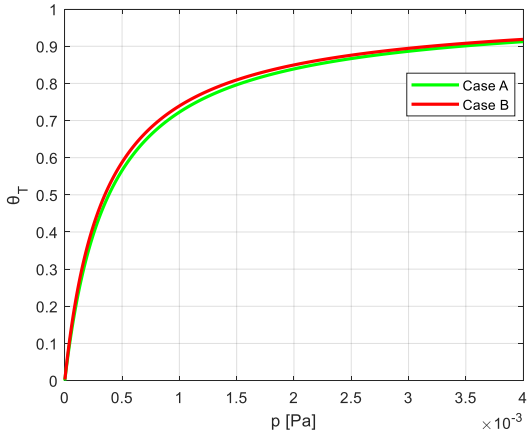


Figure 2: Target coverage as a function of reactive gas partial pressure

Any desired target coverage can be reached by keeping the reactive gas partial pressure at a corresponding value. This of course is only possible with the help of some feedback type control, since not all pressure values can be observed in an open loop manner, due to the hysteresis of the p-q characteristic curve.

4. Target coverage estimation based on the discharge voltage

A method is needed to determine the target coverage in an online manner, so that it can be used for control purposes. Since the target is used as an active

electrode, its voltage, i.e. the discharge voltage between the target and the sputtering chamber, can be measured. The target voltage depends on the target conditions, being influenced by the secondary electron emission of the different materials that form the target and are exposed. In [15] a linear dependency of the Y_{SEE} (secondary electron emission yield) on the poisoned $Y_{SEE,MG}$ and metallic $Y_{SEE,M}$ target fractions, has been assumed. This is formulated in (13).

$$Y_{SEE} = Y_{SEE,M}(1-\theta_T) + Y_{SEE,MG}\theta_T \tag{13}$$

In [10] and [11] the measured discharge voltage of different target materials both in metallic and poisoned mode has been reported. It has been observed, that target electrodes made of Al, Mg, Ce, and Y have lower discharge voltages when they are poisoned, but the opposite is true for the target electrodes made of Ag, Au, Cr, Cu, Nb, Pt, Re, Ta, Ti.

Due to the fact that $U_d \sim \frac{1}{Y_{SEE}}$ [16], Strijckmans [15] rewrites (14) in the following form

$$\frac{1}{U_d} = \frac{1}{U_M}(1-\theta_T) + \frac{1}{U_{MG}}\theta_T, \tag{14}$$

where U_M and U_{MG} are the discharge voltages at a given discharge current and Ar pressure for a metallic target and a completely poisoned target, respectively. The exact method for determining these voltages is presented in detail in [10] and [11]. This allows for the target coverage to be calculated in an online manner during sputtering by monitoring the target discharge voltage U_d

$$\theta_T = \frac{\frac{1}{U_d(I_d,p_{Ar})} - \frac{1}{U_M(I_d,p_{Ar})}}{\frac{1}{U_{MG}(I_d,p_{Ar})} - \frac{1}{U_M(I_d,p_{Ar})}}. \tag{15}$$

This is further examined by Depla in [16] by testing different targets to see whether the target composition can be figured out from the target voltage. The results show a near exact match between the calculated and measured data. This supports (13), since a pure target partially covered by a compound behaves in the same way as any given surface of a target made of different constituents, given that the alloy is homogeneous.

5. Control structure for stoichiometric thin film deposition

A summary of the goals for a productivity-oriented control structure that can be used to produce films of a given stoichiometry is given below:

1. The controller can keep the system stable, preventing unwanted target poisoning.

2. The controller can manipulate inputs in such way as to reach a prescribed coating speed.
3. The controller can maintain stoichiometric coating of the substrate throughout the entire process.

In US patent 6,106,676 [17] a method for quick poisoning suppression is presented. It employs a cascade-type controller, which manipulates the reactive gas flow and the discharge current. The inner loop controls the partial pressure of the reactive gas by means of the discharge current, since the current controller is much faster than the mass flow controller. The outer, slower loop controls the deposition rate (function of the discharge current) by means of the reactive gas flow. The change of the partial pressure resulting from the manipulation of the gas flow is compensated by the variation of the discharge current.

A slight modification to [17] has been presented in [18], where instead of an optical emission sensor and feedback loop, the partial pressure is the quantity being controlled.

The ideas introduced in Chapters 2, 3, and 4, present a way to influence the condition of stoichiometric deposition by manipulating the partial pressure of the reactive gas. This requires in the control structure a block, which calculates the prescribed p_{Gref} based on the actual target coverage. The proposed control structure can be seen in *Fig. 3*.

The offline parameter identification block is needed to determine the sticking coefficients of the system [8].

In (4) the value of θ_T is needed, which can either be estimated according to the methods presented in (5) and (15).

The controller has been simulated with a current reference input that has a ramp, step transitions, and a constant portion. The results can be seen in *Fig. 4*.

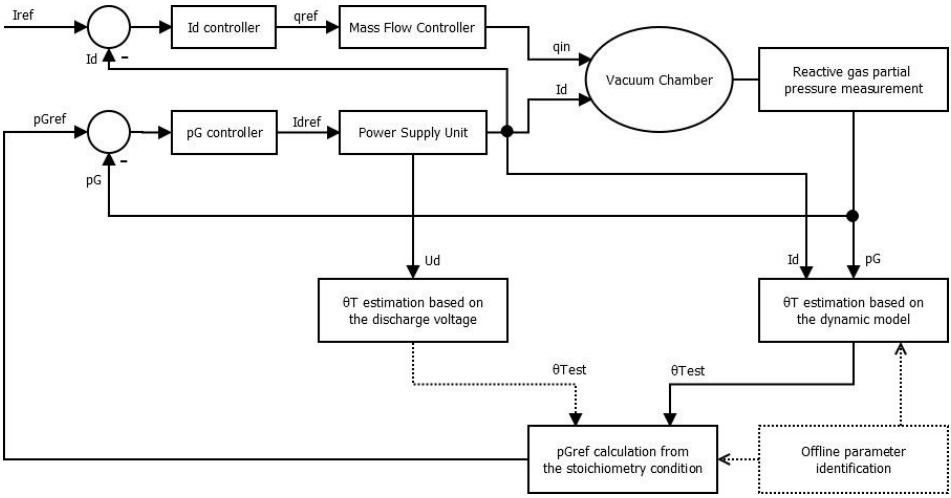


Figure 3: The proposed control structure for stoichiometric sputtering

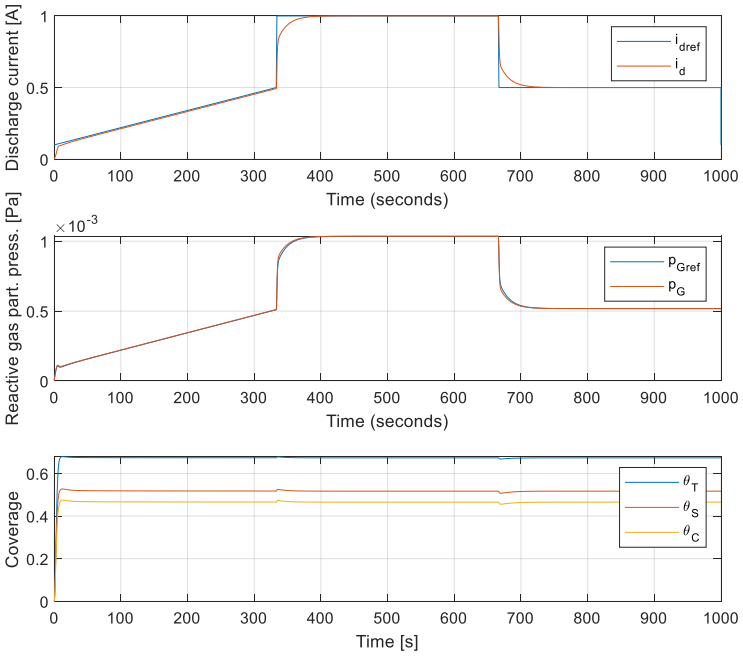


Figure 4: Simulation results obtained using the proposed control structure

6. Conclusions

An analytical study of the conditions needed for stoichiometric layer growth on the substrate surface is presented. It is shown that there is a monotonic relationship between the target coverage and the partial pressure of the reactive gas. Results are gathered from the literature that show that the target voltage, along with state estimation, can be used to calculate the target coverage. Combining these findings, a method is presented that facilitates stoichiometric film growth on the substrate surface.

A control structure employing two control loops is proposed, that can keep the system stable, while controlling the sputtering power, and which leads to stoichiometric film growth on the substrate surface.

References

[1] Luttrell, T, Halpegamage, S., Tao, J, Kramer, A., Sutter, E., and Batzill, M., "Why is anatase a better photocatalyst than rutile? - Model studies on epitaxial TiO2 films", *Sci. Rep.*, vol. 4, pp. 1–8, 2014, doi: 10.1038/srep04043.

[2] López-Huerta, F. *et al.*, "Biocompatibility and surface properties of TiO2 thin films deposited by DC magnetron sputtering", *Materials (Basel).*, vol. 7, no. 6, pp. 4105–4117, 2014, doi: 10.3390/ma7064105.

[3] Möls, K. *et al.*, "Influence of phase composition on optical properties of TiO2: Dependence of refractive index and band gap on formation of TiO2-II phase in thin films", *Opt. Mater. (Amst).*, vol. 96, no. August, p. 109335, 2019, doi: 10.1016/j.optmat.2019.109335.

[4] SpaceX, "Brightness Mitigation Best Practices for Satellite Operators", 2022, [Online]. Available:https://api.starlink.com/public-files/BrightnessMitigationBestPracticesSatelliteOperators.pdf

[5] Tomaszewski, H. *et al.*, "TiO2 films prepared by DC magnetron sputtering from ceramic targets", *Vacuum*, vol. 68, no. 1, pp. 31–38, 2002, doi: 10.1016/S0042-207X(02)00279-8.

[6] Schiller, S., Heisig, U., Steinfelder, K., and Strümpfel, J., "Reactive D.C. Sputtering with the Magnetron-Plasmatron for Tantalum Pentoxide and Titanium Dioxide Films," in *International Conference on Metallurgical Coatings, San Diego, California, U.S.A.*, San Diego, California: Elsevier Sequoia S.A., Lausanne, 1979, pp. 369–375.

[7] Christie, D. J., "Making Magnetron Sputtering Work: Modelling Reactive Sputtering Dynamics, Part 2", *SVC Bull.*, vol. 2015, no. Spring, pp. 30–33, 2015, [Online]. Available: http://www.svc.org/DigitalLibrary/documents/2016_Spring_DJC.pdf

[8] Kelemen, A. and Madarász, R. R., "Reactive magnetron sputtering: An offline parameter identification method," *SACI 2021 - IEEE 15th Int. Symp. Appl. Comput. Intell. Informatics, Proc.*, pp. 357–362, 2021, doi: 10.1109/SACI51354.2021.9465630.

[9] Depla, D., Buyle, G., Haemers, J., and De Gryse, R., "Discharge voltage measurements during magnetron sputtering", *Surf. Coatings Technol.*, vol. 200, no. 14–15, pp. 4329–4338, 2006, doi: 10.1016/j.surfcoat.2005.02.166.

[10] Depla, D., Haemers, J., and De Gryse, R., "Discharge voltage measurements during reactive sputtering of oxides", *Thin Solid Films*, vol. 515, no. 2 SPEC. ISS., pp. 468–471, 2006, doi: 10.1016/j.tsf.2005.12.256.

[11] Depla, D., Heirwegh, S., Mahieu, S., Haemers, J., and De Gryse, R., "Understanding the discharge voltage behavior during reactive sputtering of oxides", *J. Appl. Phys.*, vol. 101, no.

1, 2007, doi: 10.1063/1.2404583.

[12] Berg, S., Nyberg, T., Blom, H., and Nender, C., "Computer modeling as a tool to predict deposition rate and film composition in the reactive sputtering process composition in the reactive sputtering process", *J. Vac. Sci. Technol. A*, vol. 16, no. 3, pp. 1277–1285, 1998, doi: 10.1116/1.581274.

[13] György, K., Kelemen, A., and Papp, S., "Modeling and Stability Analysis of the Nonlinear Reactive Sputtering Process", in *The 5th Edition of the Interdisciplinarity in Engineering International Conference "Petru Maior" University of Tîrgu Mureş*, 2011, pp. 11–15.

[14] Christie, D. J., "Making Magnetron Sputtering Work: Modelling Reactive Sputtering Dynamics, Part 1", *SVC Bull.*, vol. 2014, no. Fall, pp. 24–27, 2014, [Online]. Available: http://www.svc.org/DigitalLibrary/documents/2016_Spring_DJC.pdf

[15] Strijckmans, K., "Modeling the Reactive Magnetron Sputtering Process- PhD Thesis", Universiteit Gent, 2015.

[16] Depla, D., Tomaszewski, H., Buyle, G., and De Gryse, R., "Influence of the target composition on the discharge voltage during magnetron sputtering", *Surf. Coatings Technol.*, vol. 201, no. 3–4, pp. 848–854, 2006, doi: 10.1016/j.surfcoat.2005.12.047.

[17] Terry, R., Gibbons, K., and Zarrabian, S., "United States Patent- 6,106,676", 2000.

[18] Madarász, R. R., and Kelemen, A., "Stoichiometry control of the two gas reactive sputtering process", in *IEEE Joint CINTI - MACRo 2019*, 2019.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 86–97

DOI: 10.2478/auseme-2023-0007

Numerical Study of Bottom Shape Effect on the Mixing for Stirred Tank

Abdelghani BELHANAFI[1], Touhami BAKI[2]

[1] Department of Mechanical Engineering, Faculty of Mechanics, Mechanical Apply Laboratory, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), El Mnaouer, PO Box 1505, Bir El Djir 31000, Oran, Algeria, e-mail: abdelghani.belhanafi@univ-usto.dz
[2] Faculty of Mechanics, Gaseous Fuels and Environment Laboratory, University of Sciences and Technology of Oran Mohamed Boudiaf (USTO-MB), El Mnaouer, BP1505, Bir El Djir 31000, Oran, Algeria, e-mail: touhami.baki@univ-usto.dz

Manuscript received September 01, 2023; revised October 28, 2023

Abstract: In this work, computational fluid dynamics (CFD) was used to investigate the effect of the bottom shape on the flow field and mixing characteristics in an agitated vessel stirred by six-pitched blade turbine (PBT6). The simulation was based on the resolution of the Navier-Stokes equations using standard k-ε turbulence model. Hydrodynamic behavior of the stirred vessel with four types of bottom shapes was investigated, a flat bottom and three different semi-elliptical bottoms. The results show reasonably satisfactory agreement with the experimental data. The CFD simulation also showed that the semi-elliptical bottom shapes had a significant influence on the velocity profiles and on the turbulent kinetic energy distributions of the local flow generated at the bottom region of the vessel.

Keywords: Bottom shape, computational fluid dynamics (CFD), hydrodynamic, tank height, turbulent flow.

1. Introduction

Mechanical agitation occurs in several industrial processes. Its field of application is very wide, from the mixing of miscible fluids to the manufacture of pharmaceutical products. Stirring techniques are parameters to take into account when optimizing a process.

In the literature, the study of agitation systems has been initiated in several works. We can mention the work of Hao et al. [1] who studied the influence of bottom shape on hydrodynamics and particle suspension in a DTB crystallizer. The effect of eccentrically located position of the impeller on the mixing time was investigated by Koji et al. [2] in agitated vessels of different bottom shapes. They

concluded that the bottom shapes do not affect the mixing time but give a significant influence on power consumption. Experimental studies have been performed by Ghionzoli et al [3] and Kondoo et al. [4] to evaluate the influence of bottom roughness and optimum bottom shape on solid particle dispersion in stirred vessels. They concluded that the distribution of the solids concentration in the tank was significantly influenced by the local flow formed at the vessel base.

The comparative study has been carried out by Foukrach et al. [5] on the influence of the vessel shape on the hydrodynamic performances in a tank mechanically agitated by a Rushton turbine and by Ammar et al. [6] on the design effect for three different vessel shapes: a flat-bottomed cylindrical vessel, a dished bottomed cylindrical vessel and a closed spherical vessel. Gong et al. [7] studied the effect of different geometric parameters in stirred vessels with a flat square base equipped with four blades impeller on the particles concentration in a solid-liquid system. Binxin. [8] evaluated six turbulence models of a non-Newtonian fluid in a mechanically stirred anaerobic digester with inclined turbine blades (PBT) by using (standard k-ε, realizable k-ε, RNG k-ε, Standard k-ω, SST k-ω and Reynolds Stress Model). He found that the standard turbulence models k-ω and the realizable k-ε performed better than the other models. Numerical and experimental studies carried out by Antonija et al. [9] investigated the mobile type effect of agitation and its position in a cooling catalyst on the enhancement of the kinetic of the crystal borax decahydrate. The used geometry was a cylindrical flat bottom tank equipped with four baffles placed at 90° with a distance ratio between the impeller and the bottom which varies from 0.1 to 0.5 for three types of turbines and four different blades. For a low viscous Newtonian fluid and turbulent flow regime, Taca and Paunescu [10] studied experimentally the power input in a spherical closed vessel stirred by a Rushton turbine or six pitched blade impeller. A comparative study was conducted by Jie et al. [11] on the influence of the bottom shape and of the baffles' length on the velocity distribution in transitional and turbulent flow, in two agitated tanks (flat bottom and dished bottom), equipped with a Rushton turbine. They found that the shapes of the bottom tank have significant effect on the flow pattern, as well as on the velocity profiles below the impeller.

LDV method was used by Aubin et al. [12] to measure the one single phase turbulent flow in a tank stirred by a down and an up-pumping modes using two different turbines. They observed the circulation efficiency and it was shown to be clearly superior for up-pumping impellers (especially for the PBT) than for the down-pumping or reverse mode operations. Using a PIV and LDA systems, Aubin et al. [13], Gabriele et al. [14] and Petrícek et al. [15] studied the effect of the down and up-pumping direction of the impeller in a stirred tank. From these anterior studies, it is clear that the design investigation of the bottom shape for stirred tank is very useful.

The present work aims to determine the influence of geometrical parameters of the stirred vessel with down pumping direction from a pitched blade turbine (PBT 45°) on the hydrodynamic structure of the flow. Moreover, we are interested in moving the lowest point of the tank located at the center dished bottom towards the turbine by reducing the bottom height with the same distance.

2. Geometry model

As presented in *Fig. 1*, four geometrically scaled vessels are chosen with different tank heights equal to: $H+h_3$=0.190 m named case 1, $H+h_2$=0.177 m named case 2, $H+h_1$=0.164 m named case 3 and H=0.152 m named case 4. The first system configuration is similar to the one presented by Aubin et al. [12]. In these conditions, the tank height is equal to the diameter $H=T$ for each tested geometry, and is completely filled with water. In each case, a six pitched blade turbine PBT45° has been used, with diameters d_1=0.095m, d_2=0.088m, d_3=0.082m, d_4=0.076m, respectively.

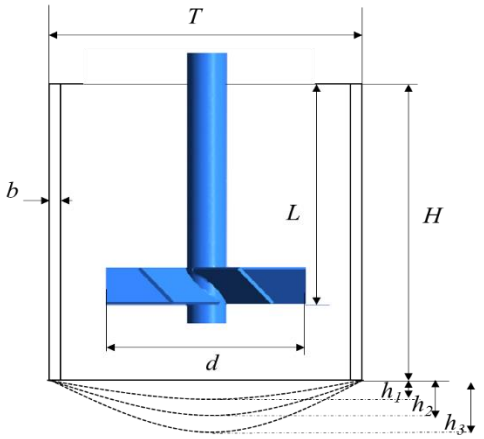


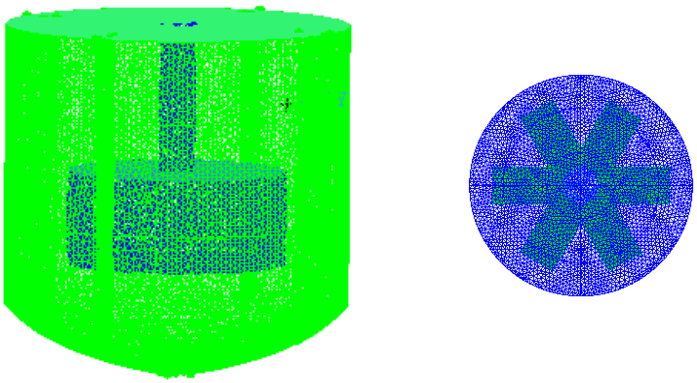
Figure 1: Vessel configuration

3. CFD simulation

Three-dimensional computational fluid dynamic modeling was carried out. to investigate, with the help of a CFD software (Ansys CFX), the flow field in an agitated vessel stirred by six-pitched blade turbine (PBT6). The computer tool Ansys ICEM CFD is used to mesh the computational domain by tetrahedral cells

(*Fig. 2*). The domain is divided into two distinct zones. Tank walls and baffles constitute the stationery zone (*Fig. 2a*) discretized with 530208 nodes. A rotating zone (*Fig. 2b*) describes the rotational motion of the fluid around the impeller. This zone is discretized with 188773 nodes. The boundary conditions are introduced in ANSYS CFX-Pre, using the multiple referential approach (MRF).

In this approach, the interface between the two regions is treated by the method called frozen rotor and the flow fields are connected at the interior surfaces (interface) separating the two domains.



a. Stationary zone *b. Rotating zone*

Figure 2: Tetrahedral mesh generation

4. Equation

The power Number (*Np*) is an essential parameter to estimate agitated tank. It is given as:

$$N_P = \frac{P}{\rho N^3 d^5} \quad (1)$$

In (1), *P* is the power consumption, *d is* turbine diameter and *N* is the impeller rotational velocity. The Reynolds number Re is the ratio between the viscous and inertia forces:

$$Re = \frac{\rho N d^2}{\mu} \quad (2)$$

In (2), ρ and μ are the density and dynamic viscosity of the working fluid.

The dimensionless radial and axial coordinates *r** and *z** are defined respectively as:

$$r^* = {}^{2r}/_{T} \; and \; z^* = {}^{z}/_{T} \tag{3}$$

In (3), r is radius of the tank, z is axial coordinate and T is vessel diameter. The velocity at the edge of the blade is defined as:

$$V_{tip} = \pi N d \tag{4}$$

The dimensionless radial velocity component is

$$V^* = \frac{V}{V_{tip}} \tag{5}$$

5. Numerical results

5.1 Validation

First, we considered necessary to validate the CFD model, for this purpose we have referred to the work of Aubin et al. [12]. With the exactly same geometrical conditions we predicted the variation of radial profile of turbulent kinetic energy k^* along the vessel dimensionless radius r^* (*Fig. 3*) for Reynolds number equal with 4.5×10^4. At axial position $z/H = 0.057$, it can be seen that predicted profiles have a similar form to the experimental one and there the agreement is quite good. As observed on this figure, the maximum wake of the dimensionless turbulent kinetic energy extended with the up-pumping direction flow is caused by the inclined angle of the blade impeller.

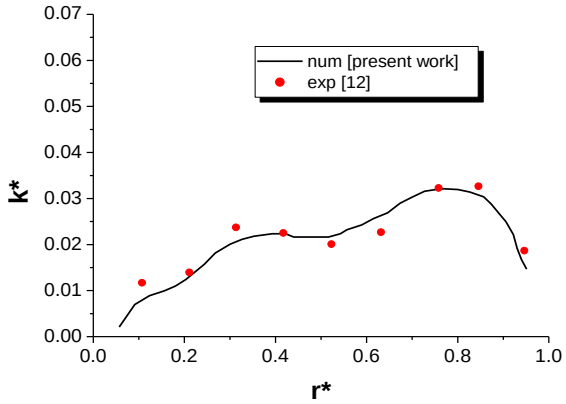


Figure 3: Radial profiles of dimensionless turbulent kinetic energy at axial position z/H=0.057

5.2 Effect of the bottom shape of tank

5.2.1 Tangential Velocity

The distribution of the turbulent viscosity is presented in the r–z plane (*Fig. 4*). In the two configurations corresponding to case 1 (*Fig. 4a*) and case 4 (*Fig. 4d*), the wake of the maximum values of the turbulent viscosity appears on the mechanical source and develops within the fluid to reach the sidewall of the tank. Indeed, reducing the height of the convex bottom creates larger amplitude circulation loops in the upper part of the tank with a dished bottom, for both systems case 2 and case 3 (*Fig. 4b* and *4c*, respectively).

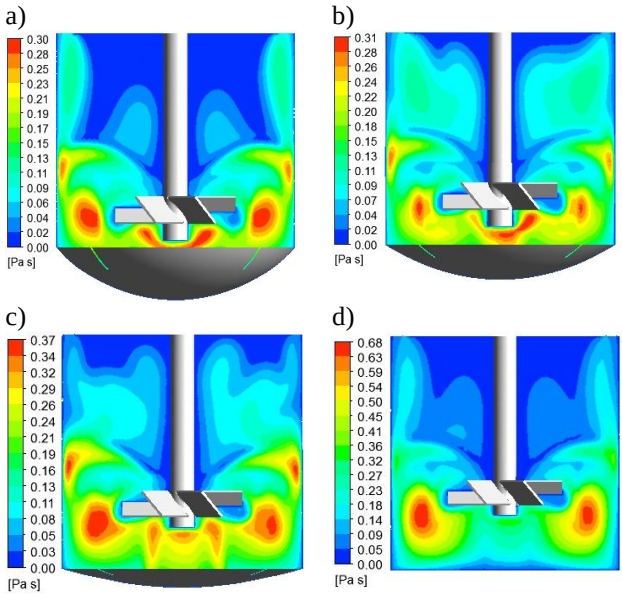


Figure 4: Velocity contours in the r-z plane

5.2.2 Turbulent kinetic energy

The distribution of the turbulent kinetic energy is shown in *Fig. 5*, on the vertical plane. The turbulent kinetic energy keeps a high value at the proximity of the side wall of the tank for the 3rd and 4th configurations (*Fig. 5c* and *5d*). This is explained by the reduction of the distance between the impeller and the bottom center of tank. In the first configuration with dished bottom corresponding to case 1 (*Fig. 5a*), the turbulent kinetic energy is very low with 0.07 m^2/s^2 compared to the three other cases. It can be noted that as the vessel diameter

decreases, there is a continuous increase in turbulent kinetic energy between the impeller blade tip and vertical wall of the tank.

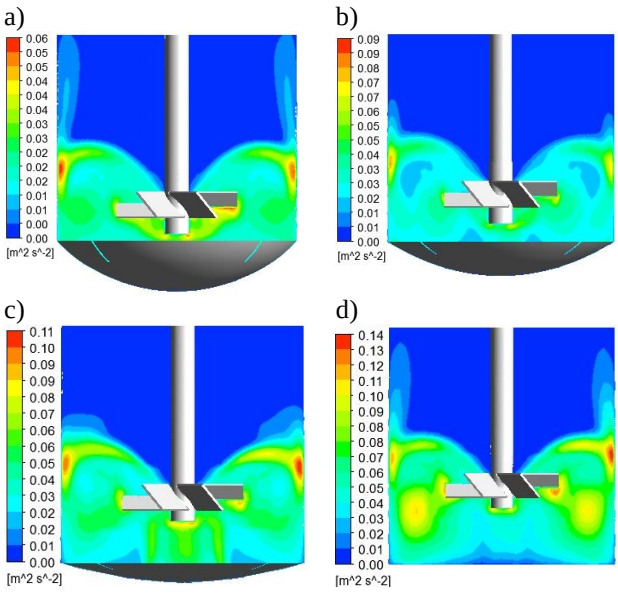


Figure 5: Turbulent kinetic energy contours r-z plane

5.2.3 Radial profiles of the dimensionless turbulent kinetic energy

For a location just below the pitched blade turbine at z=15mm, the profiles of dimensionless turbulent kinetic energy are plotted on *Fig. 6* for four cases. Particularly, in case 4 corresponding to a tank with flat bottom, the height values of turbulent kinetic energy are obtained along the vessel radius. This is due to the small distance between the impeller and effects of the flat bottom shape of the tank. But in the other three tank configurations with dished bottom, the turbulent kinetic energy profile decreases gradually and becomes negligible at sidewall. The magnitude of *TKN* increases with respect to the decrease of tank diameter and close turbine to the vessel base with a small ratio of *d/T*.

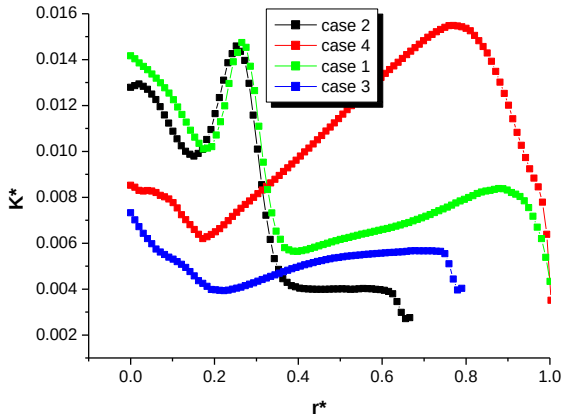


Figure 6: Dimensionless turbulent kinetic energy at z= 15mm, *N*=300 rpm

5.2.4 Flow patterns

The flow velocity vectors generated for all studied configurations are shown in *Fig. 7*. These figures indicate the dominance of a radial jet originating from the turbines which divides into two jets when it reaches the side wall of the tank. The descending jet creates an intense recirculation zone at the bottom, while the ascending jet decreases gradually in the configurations corresponding to case 1, 2 and 4 (*Fig. 7a*, 7b and *7d* respectively).

A very larger circulation loop is formed near the vertical solid wall in the tank with a dished bottom i.e. case 3 (*Fig. 7c*). This can be explained by higher down-pumping flow effects of the pitched blade turbine and the shape of vessel base.

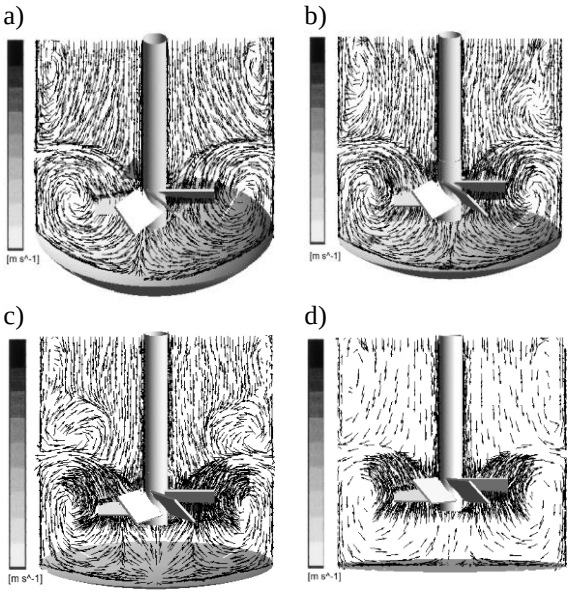


Figure 7: Flow patterns in r-z plane

5.3 Dimensionless velocity profiles

5.3.1 Axial velocity component

The profiles plotted on *Fig. 8* give the distribution of dimensionless axial velocity for four cases studied at $z^*= 0.13$. We can distinguish two different zones: negative and positive zone of the axial component. This dynamic phenomenon is due to the down pumping effect of the six pitched bade turbine toward the bottom of the tanks. In the first configuration (case 1), the axial component is very low compared to the other configurations (cases 2, 3 and 4 respectively), because of the bottom shape effect of each tank and the large distance between the impeller and vessel base. Arguably, the axial velocity component is influenced by the impeller's clearance from the tank bottom.

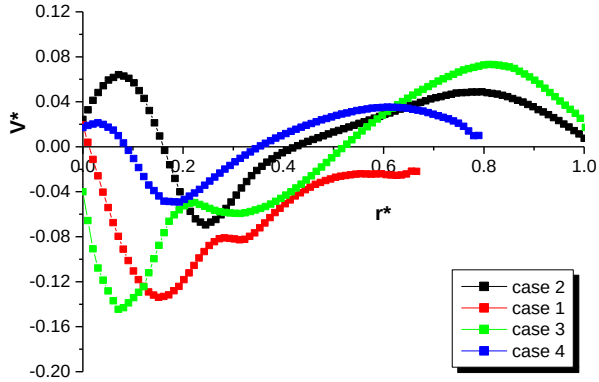


Figure 8: Axial velocity at axial position z*= 0.13 (below the impeller), *N*=300 rpm.

5.3.2 Radial velocity component

Fig. 9 shows a parabolic shape of the dimensionless radial component along the vessel's radius (in the space between shaft and vertical side wall). At *z*= 0.13, the profiles of the radial velocity are distributed according to the configurations studied in decreasing way: case 3, 1, 4 and 2 respectively. This velocity profile is due to the reduction of the bottom height of the tank. The radial component is more significant for two configurations: case 1 and case 3, both with a dished bottom tank. In the case 4 with flat bottom of the tank the velocity is very low. The negative values of the velocity indicate the existence of recirculation loops.

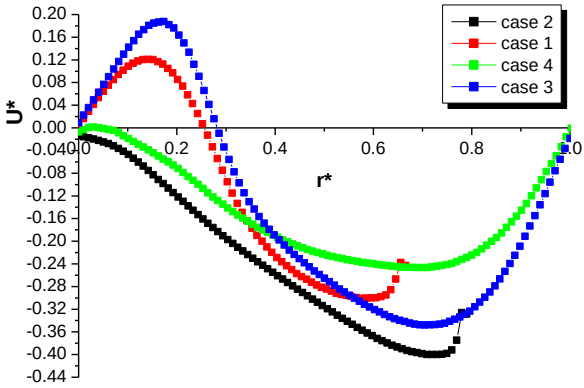


Figure 9: Radial velocity at axial position z*= 0.13 (below the impeller), *N*=300 rpm.

6. Conclusion

The goal of the investigation in this work was to determine the effect of the distance between the turbine and the bottom of the tank with different geometries. From the presented results the following conclusions can be deduced.

In planes r-z the reduction in bottom clearance for all tank configurations leads to significant growth in the parameters of turbulent viscosity and turbulent kinetic energy. A closer bottom to the turbine results in the intense movement of turbulent viscosity towards the solid walls of the tank and a notable improvement in the upper part for both configurations, namely case 2 and 3. It minimizes the stagnation of the turbulent kinetic energy at the vessel base. It has been clearly noticed that the axial and radial velocities are directly affected by impeller size and tank design. The impeller clearance from the bottom tank with the down pumping direction of pitched blade turbine plays also an important role by improving the operating conditions of stirring and mixing.

We can conclude that the design of the bottom shape of the cylindrical tank with down pumping direction of a pitched blade turbines is a very important parameter in optimizing the mixing systems.

References

[1] Pan, H., Li, J., Jin, Y., Yang, B., and Li, X., "Numerical investigation of the effect of bottom shape on the flow field and particle suspension in a DTB crystallizer", *International Journal of chemical engineering*, pp. 1-11, 2016, DOI:10.5545/sv-jme.2011.013.

[2] Koji, T., and Shigihara, D. T. Y., "Laminar mixing in eccentric stirred tank with different bottom", *Journal of chemical engineering of Japan*, 44, pp. 931-935, 2012.

[3] DOI:10.5545/sv-jme.2011.013.

[4] Ghionzoli, A., Bujalski, W., Grenville, R. K., Nienow, A. W., Sharpe, R. W., and Paglianti, A., "The effect of bottom roughness on the minimum agitator speed required to just fully suspend particles in stirred vessel", *Chemical Engineering Research and Design*, 85, pp. 685–690, 2007, DOI:10.5545/sv-jme.2011.013.

[5] Konodo, S., Motoda, M., Takahashi, K., and Horiguchi, H., "The influence of the bottom shape of an agitated vessel stirred by dual impellers on the distribution of solid concentration"., *Journal of Chemical Engineering of Japan*, 40, pp 617–621, 2007, DOI:10.5545/sv-jme.2011.013.

[6] Foukrach, M., Bouzit, M., Ameur, H., and Kamla, Y., "Influence of the vessel shape on the performance of a mechanically agitated systeme", *Chemical papers*, 73, pp 469–480, 2019,

[7] DOI:10.5545/sv-jme.2011.013.

[8] Ammar, M., Driss, Z., Chtourou, W. and Abid, M. S., "Effect of the Tank Design on the Flow Pattern Generated with a Pitched Blade Turbine", *European Journal of Mechanics* B/Fluids, 29, pp 236–245, 2010, DOI:10.5545/sv-jme.2011.013.

[9] Gong, H., Huang, F., Li, Z., Gao, Z., and Derksen, J. J., "Mechanisms for drawdown of floating particles in a laminar stirred tank flow", *Chemical Engineering Journal*, 346, pp. 340–350, 2018. DOI:10.5545/sv-jme.2011.013.

[10] Binxin. W., "CFD investigation of turbulent models for mechanical agitation of non Newtonian fluids in anaerobic digester", *Water Research*, 45, pp. 2082–2094, 2011, DOI:10.5545/sv-jme.2011.013.

[11] Antonija, K. Akrap, M. and Kuzmanic. N., "Effect of impeller type and position in a batch cooling crystallizer on the growth of borax decahydrate crystals", *Chemical Engineering Research and Design*, 91, 274–285, 2013, DOI:10.5545/sv-jme.2011.013.

[12] Taca, C. D., and Paunescu, M., "Power input in closed stirred vessels", *Chemical Engineering Science*, 56, pp 4445-4450, 2001, DOI:10.5545/sv-jme.2011.013.

[13] Jie, D., Hu, B., Pacek, A. W., Yang, X., and Miles, N. J., "The Effect of bottom shape and baffle length on the flow field in stirred tanks in turbulent and transitional flow", *International Journal of Mechanical*, pp. 1–10, 2016, DOI:10.5545/sv-jme.2011.013.

[14] Aubin, J., Fletcher, D. F., Xuereb. C., "Modeling turbulent flow in stirred tanks with CFD: the influence of the modeling approach, turbulence model and numerical scheme", *Experimental Thermal and Fluid Science*, 28, pp. 431–445, 2004, DOI:10.5545/sv-jme.2011.013.

[15] Aubin, J., Mavros, P. D., Fletcher, F., Bertrand, J., and Xuereb. C., "Effect of axial agitator configuration (up-pumping, down-pumping, reverse rotation) on flow patterns generated in Stirred Vessels", *Trans IChemE*, 79, pp 845–56, 2001, DOI:10.5545/sv-jme.2011.013.

[16] Gabriele, A., Tsoligkas, A. N., Kings I. N., and Simmons, M. J. H., "Use of PIV to measure turbulence modulation in a high throughput stirred vessel with the addition of high Stokes number particles for both up-and down-pumping configurations", *Chemical Engineering Science*, 66, pp. 5862–5874, 2011, DOI:10.5545/sv-jme.2011.013.

[17] Petrícek, R., Labík, L., Moucha, T., Brucato, A., and Scargiali., F., "Gas-liquid mass transfer rates in unbaffled tanks stirred by PBT: scale-up effects and pumping direction", *Chemical Engineering Research and Design*, 137, pp 265–272, 2018, DOI:10.5545/sv-jme.2011.013.

Acta Universitatis Sapientiae
Electrical and Mechanical Engineering, 15 (2023) 98–107

DOI: 10.2478/auseme-2023-0008

Development of a Methodology for Assessing the Quality of Drilling Machines

Abdessalam CHEKIR[1], Samia NEMOUCHI[2],
Elias HADJADJ AOUL[3]

[1] Mechanical Genetics Department, University of August 20,
1955 Skikda21000 Algeria, e-mail: abdesalam18@hotmail.fr
[2] Mechanical Genetics Department, University of August 20,
1955 Skikda21000Algeria, e-mail: Nemouchisamia@hotmail.com
[3]Electromechanical Systems Laboratory, Department of Electromechanical,
Badji Mokhtar – Annaba University, Annaba, 23000, Algeria, e-mail: hadjadj.elias@yahoo.fr

Manuscript received December 08, 2022, revised July 24, 2023

Abstract: Optimization is one of the best techniques used in industries to assess the quality of the drilling machine at a lower cost. This work is based on the rotary drilling mode with different methods based on the physicomechanical properties of the rock, machine setting parameters, and tool geometry parameters. The aim of the research presented is to develop quality of methods for determining the parameters of the rational operating regime of the rotary drilling machine during its operation in geological and mining conditions.

Keywords: Setting parameter, axial force, drilling speed, rock, quality.

1. Introduction

The engineering community attempted to influence the use of energy to drill rock formations as early as the late 1940s when it developed some drilling machines [1], [2], [3], [4], [5]. So far, drilling a hole in rock with a machine tool is one of the most common operations, it is important in geotechnics, construction, in the mining and hydrocarbon industry (gas and oil) [6].

The most commonly used drilling methods for blasting are percussion drilling, roto-percussion drilling and rotary drilling [7]. Drilling technology plays a significant role in the Algerian mining industry. Optimizing drilling parameters is an integral part of the economic success or failure of any mining operation. The use of the appropriate drilling machine is always associated with the efficiency of excavation mining and overall project success. Operators and manufacturers are constantly exploring ways to reduce costs and increase

productivity by improving the drilling penetration rate and decreasing drill bit wear [8], energy consumption, and the vibrations produced in the stem. Penetration speed is generally one of the most important factors in planning drilling in mineral deposits and estimating costs, therefore, it is necessary to predict the penetration speed to assess the total drilling costs using an accurate equation. This equation can be used to determine the type of drilling most suitable for certain conditions. In this work a method is presented for evaluating the drilling rig quality based on rock properties [9], [10] and operational variables such as rotational speed, thrust, blowing frequency, with the purpose to improve the quality [11], [12] of the drilling parameters of the machine, to increase the quality of the drilling rig and to enhance the production [13], [14], [15] at the lowest cost.

2. Nomenclature

P_{ax}: Axial force, P_{ax}=800-1500 [kgf];
d: Tool diameter, d=16 [cm];
Δ: Coefficient relating to the intermittence of the cutting edge of the bit; Δ=0,15;
m: Number of disconnected pair; m=1;
σ_d: Specific rock resistance;
f: The hardness of the rock, f=7;
C_{ft}: Coefficient of the geometric shape of the cutting edge;
nr: Rotation speed, nr=80 [1/min];
V_f: Drilling speed, [m/min];
β: Determines the position of the weathered rock surface, β=30°;
L: Maximum tooth wear, L=4 [mm];
r: Tool clearance, r = 7,5 [mm];
R: Tool radius, R=8 [cm];
C_f: Coefficient of friction, C_f=0,5;
σ_{com}: Compressive strength;
L: Footage of the drilled hole, [m];
T_p: Duration of a shift [min];
K_{exp}: Operating coefficient;
T_f: Productive working time of the drill [min];
T_{org}: Waste of time because of work organization, [min];
T_{aux}: Loss of time in carrying out auxiliary work downtime of the drill due to its imperfection, [min];
C: Cost of one meter of the drilled hole, [DA/m];
C_{ou}: Tool price, [DA];
H: The footage drilled of holes relating to a tool, [m];
C_p: Expenses relating to the operation of the drilling machine; [DA/post].

3. Theoretical determination of the indices characterizing the drilling process

The best choice of a drilling machine depends mainly on the operating conditions, rock properties and quality of the machine. The researchers carried out operating tests and laboratory tests in order to determine the operating indices and design characteristics of the machine. The researchers Rakov and Peretoltchin studied the performance of drilling machines [16].

According to the theoretical method of Rakov [16], the drilling speed can be determined using the following formula (1).

$$V_f = 0.01 \cdot n_r \cdot \frac{2.4 \cdot P_{ax} \cdot C_{ft}}{\sigma_d \cdot d \cdot (1-\Delta) \cdot m} \tag{1}$$

According to the theoretical method of Peretoltchin [16], the drilling speed can be determined using the following formula (2).

$$V_f = 0.01 \cdot n_r \cdot \left(\frac{P_{ax}}{m \cdot \delta_{comp}} - \frac{R-r}{2 \cdot R} \cdot L \right) \cdot \frac{\cos B \cdot (\sin \beta - C_f \cdot \cos \beta)}{C_f \cdot K \cdot (R-r)} \tag{2}$$

During the experiments we vary the axial force and calculate each time the drilling speed by the following formula:

$$V_f = \frac{L}{T_f} \tag{3}$$

4. Drilling rig productivity

Theoretical productivity is the number of meters of the hole drilled per unit of time.

$$Q_{theo} = 60 \cdot V_f \cdot T_p \tag{4}$$

The operating productivity depends on the degree to which the technical possibilities of a drilling rig are used in concrete operating conditions.

$$Q_{exp} = Q_{theo} \cdot k_{exp} \cdot T_p \tag{5}$$

$$k_{exp} = \frac{1}{1 + \left(\frac{Tax + Torg}{L} \right) \cdot V_f} \tag{6}$$

5. Results and discussion

In Table 1 and Table 2 there are presented the results for a variable axial force, obtained by the application of formulae Rakov (1), Peretoltchin (2), and experimentally.

Table 1: The variation of the drilling speed as a function of the axial force by the two formulas (Rakov, Peretoltchin)

P_{ax} [kgf]	$V_{f\ rakov}$ [m/min]	$V_{f\ pert}$ [m/min]
0	0	0
800	0.195	0.114
850	0.207	0.122
900	0.219	0.13
950	0.231	0.139
1000	0.243	0.147
1050	0.256	0.155
1100	0.268	0.163
1150	0.28	0.171
1200	0.292	0.18
1250	0.304	0.188
1300	0.316	0.196
1350	0.329	0.204
1400	0.341	0.212
1450	0.441	0.276
1500	0.457	0.286

Table 2: Result of the experimental study

Test nr	Pax (kgf)	Vf (m/min)
1	800	0.114
2	850	0.12
3	900	0.1122
4	950	0.124
5	1000	0.134
6	1050	0.137
7	1100	0.149
8	1150	0.152
9	1200	0.161
10	1250	0.173
11	1300	0.17
12	1350	0.192
13	1400	0.1876
14	1450	0.196
15	1500	0.2

For the treatment of the results the regression analysis by the method of least squares is used, assuming that the relationship between the drilling speed and the axial forces is represented by a straight line.

Table 3: Results of the experimental study obtained by linear regression

P_{ax} [kgf]	V_f [m/min]
800	0.1073
850	0.1141
900	0.1209
950	0.1277
1000	0.1344
1050	0.1412
1100	0.148
1150	0.1547
1200	0.1615
1250	0.1683
1300	0.1751
1350	0.1818
1400	0.1886
1450	0.1954
1500	0.2022

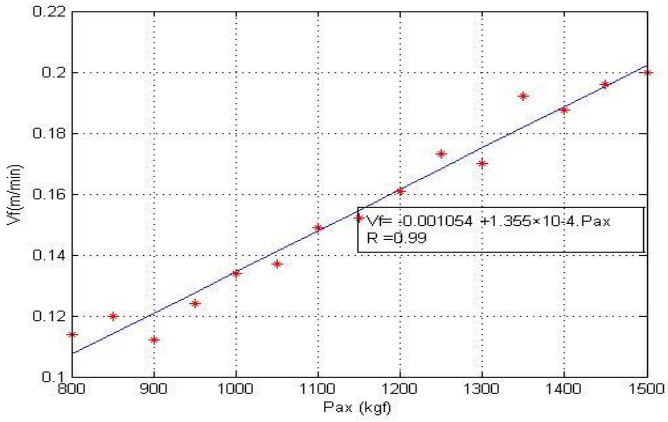


Figure 1: The variation of the experimental drilling speed as a function of the axial force

Table 4: Productivity calculated by the two methods (Rakov; Peretoltchin)

P_{ax} [kgf]	Q_{exp_Rakov} [m/post]	Q_{exp_Pert} [m/post]
0	0	0
800	78.74	49.42
850	80.881	51.76
900	84.395	54.731
950	86.17	56.819
1000	87.182	58.613
1050	88.115	60.241
1100	89.091	61.768
1150	78.873	57.59
1200	64.325	50.07
1250	54.059	43.926
1300	49.556	41.359
1350	46.769	39.665
1400	43.624	37.606
1450	42.087	36.604
1500	39.326	34.78

Table 5: Experimental results obtained

P_{ax} [kgf]	V_f [m/min]	$Q_{thé}$ [m/post]	K_{exp}	Q_{exp} [m/post]
800	0.1073	51.504	0.906	46.656
850	0.1141	54.768	0.887	48.536
900	0.1209	58.032	0.879	51.008
950	0.1277	61.296	0.862	52.832
1000	0.1344	64.512	0.841	54.248
1050	0.1412	67.776	0.821	55.644
1100	0.148	71.04	0.802	56.97
1150	0.1547	74.256	0.719	53.384
1200	0.1615	77.52	0.6056	46.95
1250	0.1683	80.784	0.514	41.52
1300	0.1751	84.048	0.467	39.25
1350	0.1818	87.264	0.432	37.698
1400	0.1886	90.528	0.396	35.849
1450	0.1954	93.79	0.374	35.078
1500	0.2022	97.056	0.343	33.35

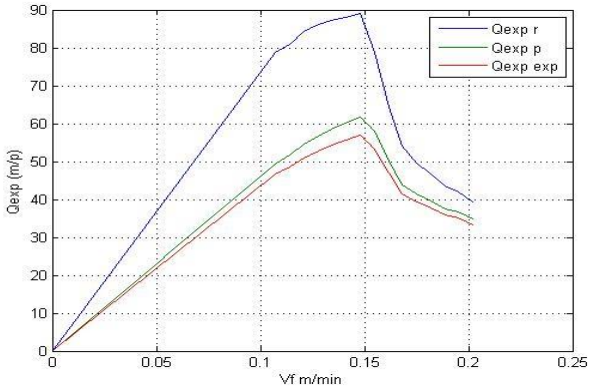


Figure 2: Comparison of operating productivity results

6. Criteria for choosing the rational operating regime of drilling machines

When carrying out mining works, as a general criterion of the effectiveness of any method of drilling, one can take the cost price of one meter of the hole drilled. This is used for determination of the parameters of a rational regime of rotary drilling having the best results and determines the optimum settings for the rotary drilling machine.

$$C = \frac{C_p}{Q_{exp}} + \frac{C_{ou}}{H} \tag{7}$$

Table 6: Cost price results

P_{ax} [kgf]	V_f [m/min]	H [m]	Q_{exp} [m/post]	C [DA/m]
800	0.1073	284	46.656	234.6
850	0.1141	248	48.576	226.48
900	0.1209	242	51.008	215.95
950	0.1277	219	52.832	209.36
1000	0.1344	196	54.248	204.88
1050	0.1421	178	54.152	200.706
1100	**0.148**	**165**	**54.412**	**196.895**
1150	0.1547	109	53.384	214.158
1200	0.1615	68	50.072	249.943
1250	0.1683	49	40.744	288.3
1300	0.1751	42	34.224	308.305
1350	0.1818	38	30.624	32328
1400	0.1886	34	26.072	342.569
1450	0.1954	32	22.56	351.886
1500	0.2022	29	20.568	372.539

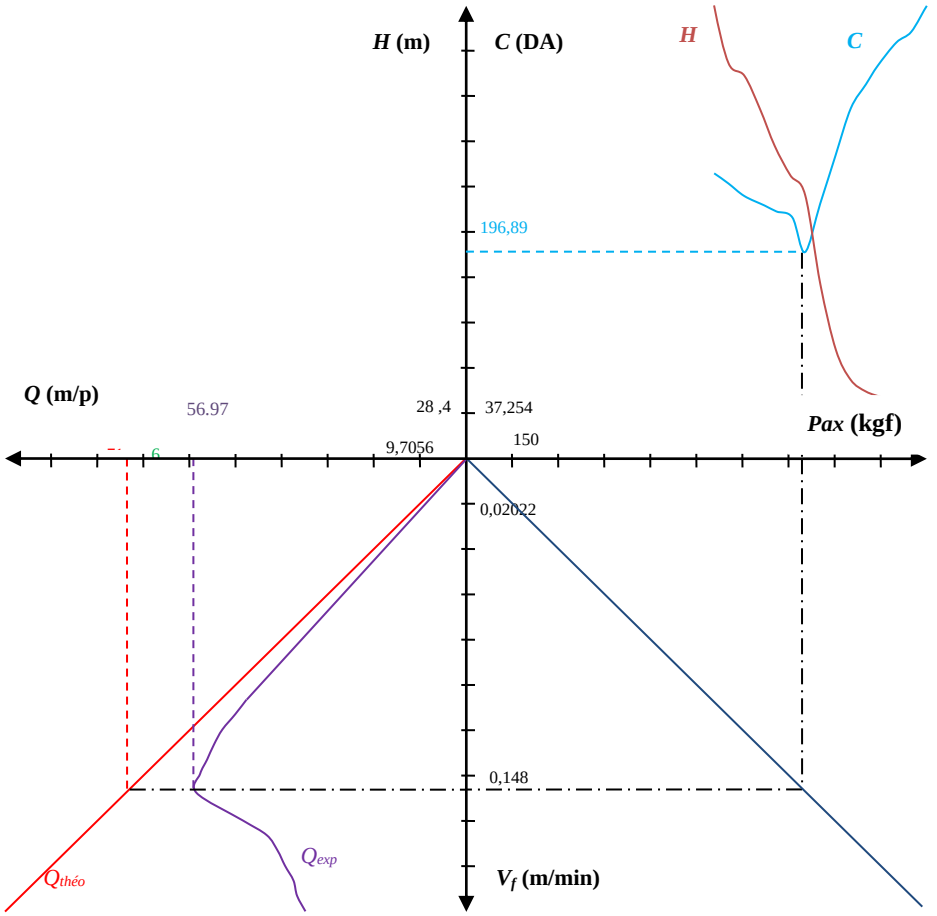


Figure3: Nomogram of experimental results

7. Conclusion

The selection methodology recommended through the technical-economic model allows the best alternative, hence the assertion that the method of; Peretoltchin is better than that of Rakov for the same conditions of the iron quarry of Ouenza (Algeria).

The optimization of this drilling method leads us to an optimal drilling cost. The rational parameters of the operating regime of drilling machines are:

Table 7: Optimal settings

P_{ax} [kgf]	V_f [m/min]	Productivity		C [DA/m]
		$Q_{thé}$ [m/p]	Q_{exp} [m/p]	
1100	0.148	71.04	54.412	196.895

From the results obtained using the proposed regression model, it can be concluded that there is a strong linear correlation between velocity and axial force ($R \approx 0.99$), and there is a very good relationship between theory and experiment.

References

[1] Mehennaoui, S., Khochemane, L., "Optimization of Setting Parameters of a Rotary Drilling Rig Using SimHydraulics Toolbox", *J. Contemporary Engineering Sciences*, vol. 8, no. 3, 115–120, 2015.

[2] Derdour, F. Z., Kezzar, M., Bannis, O., Khochman, L., "The optimization of the operation parameters of a rotary percussion shilling machine using the Taguchi methods", *World Journal of Engineering*, vol.15, issue 1, pp. 62–69, 2017.

[3] Modi, M., Agarwal, G., Patil, V., Bhatia, U., Pancholi, R., "Parametric Optimization in Drilling of Al–SiC Composite Using Taguchi Method", *International Journal of Scientific & Technology Research*, vol. 8, issue 09, 2019.

[4] Souilah, N., Zahzouh, Z., "Optimum energy calculation for a drill hammer-blow RH571-4W", *J. Fundam. Appl Sci.*, 13(1), pp. 151–171, 2021.

[5] Rabbouche, Z., Khochemane, L, Bennis, O., "Study of the determination of the rational operating regime of percussion drilling machines", *International Journal of Engineering and Technology (IJET)*, vol 9, no. 2, Apr-May 2017.

[6] Kapustin, N., O., Grushevenko, D., A., "Long-term outlook on Russian oil industry facing internal and external challenges", *J. Oil & Gas Science and Technology- Rev. IFP Energies nouvelles* 74, 72, 2019.

[7] Khocheman, L., Hadjadj, A. E., "New Méthodologie Sélection of Shilling Machines", *Arian Journal of Information Technologie* 5 (12), 1310–1312, 2006.

[8] Nemouchi, S., Hadjadj, A. E., "Improvement of the hammer drill performance (Algerian quarries conditions)", *J. Acta Univ. Sapientiae Electr. and Mechan. Eng.*, ol. v12, issue 1 pp. 77–85, 2020.

[9] Yaşar, E., Ranjith, P. G., Viete, D. R., "An experimental investigation into the drilling and physico-mechanical properties of a rock-like brittle material", *Journal of Petroleum Science and Engineering* 76, pp. 185–193, 2011.

[10] Benmoussa, M., Inel, F., "Study and determination of rational operating ratio for percussion drilling machines", *J. Fundam. Appl. Sci.*, 12(2), 755–773, 2020.

[11] Xinghua, L., Tao, J., "Working pose measurement and quality evaluation of rotary drilling rig based on laser tracker", *International Journal for Light and Electron Optics* 187 311–317, 2019.

[12] Suranga, C., Geekiyanage, H., Andrzej Tunkiel, D, S., "Drilling data quality improvement and information extraction with case studies", *Journal of Petroleum Exploration and Production*, vol. 11, pp. 819–837, 2021.

[13] Richard, T., Germai, C., Detournay, E., "A simplified model to explore the robot cause of stick–slip vibrations in shilling system within drag bits", *Journal of Sound and Vibration* vol. 305, pp. 432–456, 2007.
[14] Afeni, T. B., "Optimization of drilling and blasting operations in an open pit mine—the SOMAIR experience", *J. Mining Science and Technology* 19, pp. 0736–0739, 2009.
[15] Huang, B.-W., Kuang, J.-Y., Yu, P., "Effect of crack on Trilling vibration", Journal of Sound and Vibration vol. 322, pp. 1100–1116, 2009.
[16] Khochemane, L., "Optimisation des paramètre des machine de forage à mollettes (dans les condition de la carrière de OUENZA)", *State doctoral thesis*, 2007.