DEFENCE S&T TECHNICAL BULLETIN

VOL. 15 NUM. 2 YEAR 2022 ISSN 1985-6571

CONTENTS

Ministry of Defence Malaysia

SCIENCE & TECHNOLOGY RESEARCH INSTITUTE FOR DEFENCE (STRIDE)

AIMS AND SCOPE

The Defence S&T Technical Bulletin is the official journal of the Science & Technology Research Institute for Defence (STRIDE). The journal, which is indexed in, among others, Scopus, Index Corpenicus, ProQuest and EBSCO, contains manuscripts on research findings in various fields of defence science & technology. The primary purpose of this journal is to act as a channel for the publication of defence-based research work undertaken by researchers both within and outside the country.

WRITING FOR THE DEFENCE S&T TECHNICAL BULLETIN

Contributions to the journal should be based on original research in areas related to defence science & technology. All contributions should be in English.

PUBLICATION

The editors' decision with regard to publication of any item is final. A manuscript is accepted on the understanding that it is an original piece of work that has not been accepted for publication elsewhere.

PRESENTATION OF MANUSCRIPTS

The format of the manuscript is as follows:

a) Page size A4
b) MS Word format
c) Single space
d) Justified
e) In Times New Roman, 11-point font
f) Should not exceed 20 pages, including references
g) Texts in charts and tables should be in 10-point font.

Please e-mail the manuscript to:

1) Gs. Dr. Dinesh Sathyamoorthy (dinesh.sathyamoorthy@stride.gov.my)
2) Dr. Mahdi bin Che Isa (mahdi.cheisa@stride.gov.my)

The next edition of the journal (Vol. 16, Num. 1) is expected to be published in April 2023. The due date for submissions is 11 January 2023. **It is strongly iterated that authors are solely responsible for taking the necessary steps to ensure that the submitted manuscripts do not contain confidential or sensitive material.**

The template of the manuscript is as follows:

TITLE OF MANUSCRIPT

Name(s) of author(s)

Affiliation(s)

Email:

ABSTRACT

Contents of abstract.

Keywords: *Keyword 1; keyword 2; keyword 3; keyword 4; keyword 5.*

1. TOPIC 1

Paragraph 1.

Paragraph 2.

1.1 Sub Topic 1

Paragraph 1.

Paragraph 2.

2. TOPIC 2

Paragraph 1.

Paragraph 2.

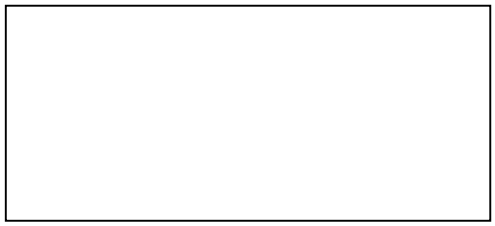


Figure 1: Title of figure.

Table 1: Title of table.

Content	Content	Content
Content	Content	Content
Content	Content	Content
Content	Content	Content

Equation 1 (1)
Equation 2 (2)

REFERENCES

Long lists of notes of bibliographical references are generally not required. The method of citing references in the text is 'name date' style, e.g. 'Hanis (1993) claimed that...', or '…including the lack of interoperability (Bohara *et al.*, 2003)'. End references should be in alphabetical order. The following reference style is to be adhered to:

Books

Serra, J. (1982). *Image Analysis and Mathematical Morphology*. Academic Press, London.

Book Chapters

Goodchild, M.F. & Quattrochi, D.A. (1997). Scale, multiscaling, remote sensing and GIS. *In* Quattrochi, D.A. & Goodchild, M.F. (Eds.), *Scale in Remote Sensing and GIS*. Lewis Publishers, Boca Raton, Florida, pp. 1-11.

Journals / Serials

Jang, B.K. & Chin, R.T. (1990). Analysis of thinning algorithms using mathematical morphology. *IEEE T. Pattern Anal.*, **12**: 541-550.

Online Sources

GTOPO30 (1996). *GTOPO30: Global 30 Arc Second Elevation Data Set*. Available online at: http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html (Last access date: 1 June 2009).

Unpublished Materials (e.g. theses, reports and documents)

Wood, J. (1996). *The Geomorphological Characterization of Digital Elevation Models*. PhD Thesis, Department of Geography, University of Leicester, Leicester.

ELECTROCHEMICAL BEHAVIOUR AND CURRENT CAPACITY STUDIES OF AS-CAST AND HEAT-TREATED Al-Zn-Mg-xSn ALLOYS IN TROPICAL SEAWATER

Mahdi Che Isa*, Nik Hassanuddin Nik Yusoff, Mohd Subhi Din Yati, Mohd Moesli Muhammad & Hasril Nain

Maritime Technology Division, Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*Email: mahdi.cheisa@stride.gov.my

ABSTRACT

In this paper, the effect of heat treatment on the electrochemical behaviour and current capacity of as-cast and heat-treated of Al-5.5Zn-2.0Mg-xSn alloys (where x = 0.1 and 2.0 %wt.) was studied using potentiodynamic polarisation, open circuit potential (OCP), electrochemical impedance spectroscopy (EIS) and current efficiency measurement. The phases present in the fabricated alloys were determined using an X-ray diffractometer (XRD), while corroded surface morphologies of the samples were studied with the aid of a scanning electron microscope (SEM). The results showed that the presence of higher Sn content (2.0 %wt.) in the Al-5.5Zn-2.0Mg-xSn alloy promoted the formation and growth of Mg_2Sn intermetallic, both in the as-cast and heat-treated conditions. The OCP, corrosion potential and EIS results of the heat-treated alloys showed improvements as compared to the as-cast condition. The current efficiency of the as-cast and heat-treated alloy increased from 81.14 to 84.37% for Al-5.5Zn-2.0Mg-0.1Sn (%wt.) and from 82.55 to 85.87% for Al-5.5Zn-2.0Mg-2.0Sn (%wt.).

Keywords: *Aluminium alloys; heat treatment; open circuit potential (OCP); electrochemical impedance; current capacity.*

1. INTRODUCTION

The steadily growing interest in aluminium alloys in various industries and engineering applications, including aerospace, transportation, marine and defence industries, are connected to its good mechanical properties (Hirsch & Al-Samman, 2013; Zakaria *et al.*, 2013; Dursun & Soutis, 2014). Aluminium also possesses attractive corrosion properties, such as high negative potential and high theoretical electric capacity (2,980 Ah/kg), which makes it an attractive choice as sacrificial anode materials in batteries and marine corrosion control applications (Mahdi *et al.*, 2011; Khireche *et al.*, 2014; Liu *et al.*, 2014; Farooq *et al.*, 2019).

However, the theoretical current capacity advantage of aluminium has not been realised in practice, as pure aluminium and its alloys tend to passivate, and thus do not function effectively as an anode material. For example, in cathodic protection systems, the alloying element, known as activator, must be added to the aluminium alloy to shift the electrode potential to a more active value sufficiently to establish the potential difference of the anode and cathode (Breslin *et al.*, 1993, Barbucci *et al.*, 1997; Bruzzone *et al.*, 1997; Salinas *et al.*, 1999; Flamini & Saidman, 2012; Pourgharibshahi & Lambert, 2016). A lot of effort has been spent on heat treatment processes in aluminium alloys (Campana & Pilone, 2009; Xu *et al.*, 2012), as it is commercially important. Homogenisation and subsequent ageing treatment of aluminium alloys are known to offer considerable improvement in mechanical properties (Gupta *et al.*, 2001; Berg *et al.*, 2001; Shibli *et al.*, 2007; Wang *et al.*, 2020).

Nonetheless, such heat treatment may have a great impact on electrochemical behaviour due to the modification of macrostructures and microstructures caused by the presence of solid solutions, segregate second-phase particles, intermetallic compounds and inclusions during heat treatment processes (Acer *et al*., 2016; Azarniya *et al*., 2019). Heat treatment of aluminium alloys also leads to the precipitation of intermetallic phase along the grain boundaries or in the sub-grain (Chen *et al*., 2021; He *et al*., 2019). The size and morphology of the intermetallic precipitates depend on the ageing temperature, time and specific processing route (Maloney *et al*., 1999; Li *et al*., 2020). These metallurgical features are expected to be directly related to the anode operating potential and its efficiency as these phases may act either as an effective cathode or a physical barrier to corrosion. Although several studies have been reported on the relation to heat-treated aluminium alloys (Dai *et al*., 2020; Babu *et al*., 2021), the effect of heat treatment on corrosion behaviour (Xu *et al*., 2012; Liu *et al*., 2021), as well as the role of intermetallic or secondary phases in electrochemical behaviour of aluminium alloys (Fang *et al*., 2017, Li *et al.*, 2021; Birbilis & Buchheit, 2005; Chen *et al*., 2021), the electrochemical behaviour of aluminium alloys with tin (Sn) addition as the activator is still less reported in the open literature.

Therefore, the purpose of this study is to examine the effect of Sn addition on current capacity and electrochemical properties of aluminium-zinc-magnesium (Al-Zn-Mg) alloys in as-cast and heat-treated condition subjected to tropical marine environment.

2. MATERIALS & METHODS

2.1 Sample Preparation, Phase Analyses and Surface Morphology.

A nominal composition of Al-5.5Zn-2.0Mg-xSn alloys were prepared by mixing high purity metals of Al, Zn, Mg and Sn (99.99%) supplied by Aldrich Chemical Co., USA. The melting of these elements was performed in an inert atmosphere melting furnace apparatus up to temperature of 800 °C for 10 min. Then they were cast into steel moulds and cooled to room temperature. The homogenisation (H) process was performed at 550 °C for 24 h in normal atmosphere followed by air quench to room temperature of 27 °C. The homogenised alloys were aged artificially (AA) in an air circulated oven at 150 °C for 3 h. The phases presence in the Al-Zn-Mg-Sn alloys were determined using an X-ray diffractometer (XRD) (Bruker D8-Advanced) housed at the School of Applied Physics, National University of Malaysia (UKM). The diffractogram was generated using a Cu Kα (λ = 1.543 Å) radiation source at scanning rate of 0.002 °/s with 2θ from 20 to 60º. The elemental composition of the alloy was analysed using a wavelength dispersive X-ray fluorescence (WDXRF) spectrometer (Bruker SP 4 Pioneer, WDXRF) equipped with a Rh X-ray tube and 4 kW generator. The X-ray generator was operated at voltage of 20-50 kV and current of 5-20 mA. The surface morphology of the alloy was characterised using a scanning electron microscope (SEM, Leo VP 1430).

2.2 Electrochemical Characterisation

In order to obtain the working electrodes, specimens with 1.6 cm in diameter and 2 mm in thickness were cut from the ingots previously prepared. Prior to the electrochemical tests, they were ground using SiC paper up to 2,400 grit and followed by diamond paste using 1 µm finish up to mirror quality, rinsed with acetone and cleaned with ethanol in an ultrasonic bath. Electrochemical measurements were conducted with a Gamry Reference 3000 potentiostat / galvanostat / zero resistance ammeter controlled by GAMRY Framework V5.65, with the output data analysed using the Echem Analyst V5.65 software. A three-electrode cell with high-density graphite as a counter electrode and a saturated calomel (SCE) reference electrode was used. All the tests were performed in tropical marine seawater taken from the Teluk Muroh area in Manjung, Perak, Malaysia. The reference electrode was separated from the cell using a Vycor frit. Before the measurements, the specimens were held in the test solution for 3 h to establish a steady open circuit corrosion potential, after allowing a steady state potential to develop (ASTM 97, 2013). The open circuit potential (OCP)

measurement was carried out continuously for at least 15 h with data recorded for every 10 s. Potentiodynamic polarisation scans were performed at a relatively slow scan rate of 0.5 mV/s, commencing from a potential -0.1 V below the corrosion potential (E_{corr}) value.

All the experiments were conducted at room temperature without temperature control. The solution volume was 1,000 mL and a Teflon knife-edge O-ring was used to expose a specimen area of 1 cm^2. As-cast and heat-treated samples in initial state and after corrosion tests (full polarisation run in tropical seawater) were used to study corrosion induced changes of alloy microstructure and corrosion product morphology. For the electrochemical impedance spectroscopy (EIS) experiments, spectrums were generated at OCP values over a frequency range of 100 kHz to 0.01 Hz. An alternating current (AC) signal (< 10 mV peak to peak) was used for the impedance measurement and the data was analysed using Gamry's accompanied software (Shibli & George, 2007; Mahdi *et al*., 2012).

2.3 Current Capacity Measurements

For the current capacity measurements, both electrodes, namely anode (Al-Zn-Mg-Sn alloy) and cathode (carbon steel) specimens, were polished to 1,200 grit, washed and finally rinsed with acetone. The individual dimensions for the round shaped anode was 0.30 cm in thickness and 1.80 cm in diameter. The cathode with rectangular shape had dimensions of 0.8 x 4.1 x 16.0 cm. The separation or distance between the anode and cathode was 10 cm, with the anode's weight being recorded before the test started. Both the anode and cathode were immersed in a lightproof acrylic tank containing 30 L of filtered tropical seawater medium for about 72 h. A current density was supplied to the anode at amount of 0.5 mA/cm^2 and charge transfer reading recorded using a coulometer after 72 h was taken for anode capacity determination. The anode specimen was then washed with water and a soft brush, and dried. For weight loss determination, anode specimen was dipped in a cleaning solution (50 g/L chromic acid) for 20 s. The samples were then washed with water, rinsed with acetone and dried to determine their weight loss (ASTM, 2013).

3. RESULTS AND DISCUSSION

3.1 Composition and Phase of the Alloys

The chemical composition of the fabricated aluminium alloys analysed using the WDXRF is shown in Table 1. A fair agreement between the nominal and real compositions was established. The iron present in both samples was not deliberately introduced, but probably came from commercial metals used during the alloy preparation.

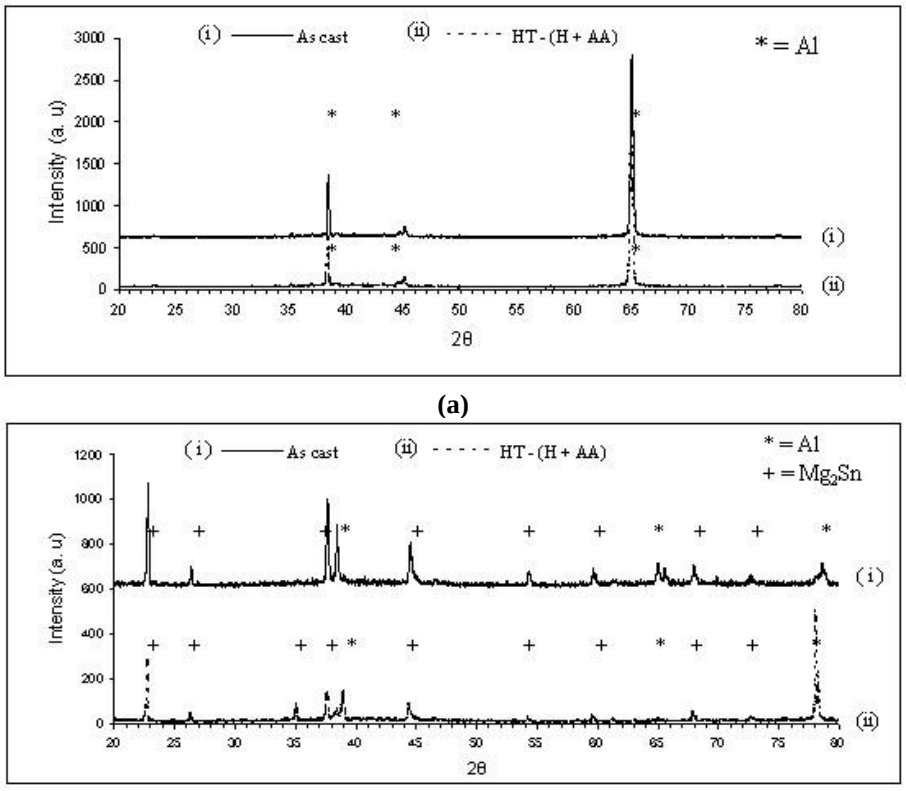
Table 1: Chemical composition of fabricated aluminium alloys determined using the WDXRF.

Al-5.5Zn-2.0Mg-xSn (Nominal wt.%)					Al-5.5Zn-2.0Mg-xSn (WDXRF results)				
Samples	*Al*	*Zn*	*Mg*	*Sn*	*Al*	*Zn*	*Mg*	*Sn*	*Fe*
B-1	Bal	5.5	2.0	0.1	Bal	5.51	1.93	1.82	0.02
B-2	Bal	5.5	2.0	2.0	Bal	5.52	1.94	1.79	0.03

Bal: Balance

It was anticipated that they were intermetallic particles or other individual phases in the alloy. This is due to the fact that Sn has low solid solubility in aluminium, whether at low or high temperature, thus it will be rejected to the grain boundary during the solidification process and act as nucleation site to give a finer grain size of the alloys. An increment in the amount of added Sn would also provide the formation and growth of new intermetallic compounds. These were confirmed through the XRD

analysis that showed Mg_2Sn intermetallic compound presence in the Al-5.5Zn-2.0Mg-2.0Sn alloy, as indicated by the diffraction peaks shown in Figure 1(b). When 2.0 wt.% Sn was added to the Al-Zn-Mg alloy, the XRD results showed that Mg_2Sn intermetallic formed whether the samples were heat-treated or not. Previous studies also showed that other stable intermetallics, such as $MgZn_2$, $Al2Mg_3Zn_3$, $Mg(ZnAl)_2$ and other metastable phases, were found in Al-Zn-Mg alloys, and can also play their roles in physical aspects and electrochemical behaviour (Barbucci *et al.*, 1997; Bruzzone *et al*., 97; Fang *et al*. 2017; Li *et al*., 2021). The XRD peak intensity in Figure 1 showed that less amount of Mg_2Sn intermetallic compound formed when the samples were heat-treated due to the dissolution parts of the intermetallics to solute atoms, and which would enrich the matrix with Zn and Mg atoms.



(a)

(b)

Figure 1: XRD spectrums for Al-5.5Zn-2.0Mg-xSn alloys in as-cast and heat-treated conditions (H + 3 h AA) of: (a) Al-5.5Zn-2Mg-0.1Sn (b) Al-5.5Zn-2Mg-2.0Sn.

3.2 Electrochemical Behaviour

One of the techniques used to monitor the corrosion behaviour of metal or alloy was by recording the OCP as a function of time. The OCP value was recorded every 10 s for 15 h in a single experiment, with the average value taken for three experiments. Figure 2 shows the effect of heat treatment on OCP against time for the Al-5.5Zn-2Mg-xSn (x = 0.1 and 2.0 wt. %) alloys immersed in tropical seawater medium. Initially, they increased towards the positive direction, and after 3 to 4 h, a more stable value was approached. The OCP showed transients for the first 5 h of immersion and reached a wavy pattern with respect to the potential changing by approximately 60-100 mV after about 5 h of immersion. The average OCP values for 15 h of immersion in tropical seawater are shown in Table 2.

The free corrosion potential transient observed for the as-cast and heat-treated aluminium alloys presumably corresponds to higher anodic and cathodic reaction as a result of the dissolution process occurring on the alloy surface. It is also attributed to changes in the properties of the near surface layer of metals or in the properties of surface oxide (Wang *et al*., 2019; Chen *et al*., 2021). The thickness of corrosion products would increase as immersion time increases, which in turn creates a hindrance between the metal surface and ionic species interaction in the solution (Xia *et al.*, 2020). As

a result of high electrical resistance and different thickness of corrosion product layer developed on the metal surface, the corrosion process was unable to proceed with stable corrosion current density.

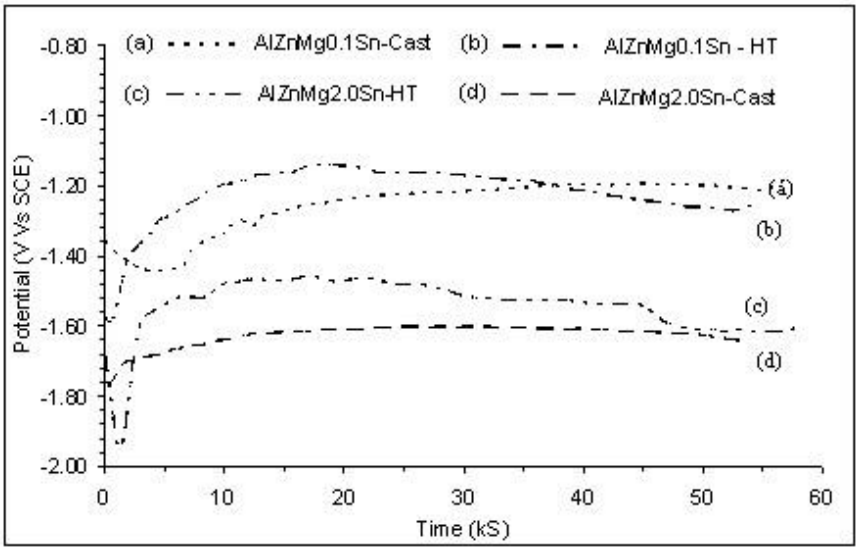


Figure 2: Corrosion potential as a function of time for the Al-5.5Zn-2.0Mg-xSn in tropical seawater medium at 27 ºC. (a) As-cast 0.1 wt. % Sn (b) Heat-treated 0.1 wt. %Sn (c) Heat treated 2.0 wt. %Sn (d) As-cast 2.0 wt. %Sn.

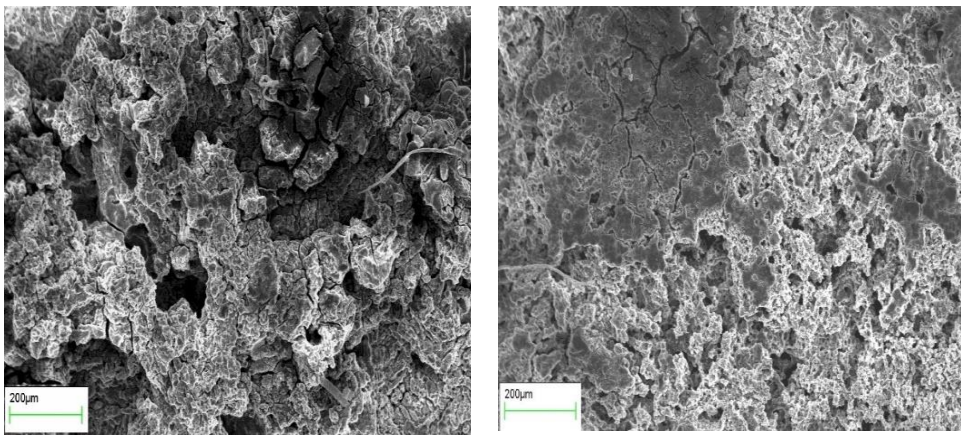
Table 2: Average OCP values for the tested alloys.

Sample	Average OCP (V, SCE)
Al-5.5Zn-2Mg-0.1Sn: As-cast	-1.21 ± 0.09
Al-5.5Zn-2Mg-0.1Sn: Heat-treated	-1.30 ± 0.07
Al-5.5Zn-2Mg-2.0Sn: As-cast	-1.54 ± 0.08
Al-5.5Zn-2Mg-2.0Sn: Heat-treated	-1.63 ± 0.03

The difference in the corrosion products morphology of the Al-5.5Zn-2Mg-2Sn alloy can be seen in Figure 3, where the heat-treated sample showed a more uniform porous layer as compared with the as-cast sample with bigger pits or holes due to localised attacks (Pourgharibshahi & Meratian, 2014). As shown in Table 2, the average OCP value is strongly dependent on the alloy composition and heat treatment, and thus these values can be used to differentiate the condition of the alloys. The as-cast sample also showed more wavy pattern as compared to the heat-treated alloy. This can be ascribed to the presence of Mg_2Sn phase, with the distribution of this cathodic intermetallic believed to be due to lack of homogeneity in the as-cast sample. Another possibility could be from the influence of the intermetallic distribution in the surface area exposed to the solution during the OCP measurement, which can produce different reproducibility as reported by previous researchers (Aballe *et al*., 2003; Sperandio *et al*., 2021). As a result of the addition of the homogenisation and artificial ageing processes, the as-cast sample with 2.0 wt.% Sn alloy turned out to be the most active electrode potential based on average OCP value that has shifted to a more significant negative direction.

The numerous Mg_2Sn phases found in the as-cast alloy was believed to lead to less polarisation resistance of the local anode and consequently shifted the OCP in the noble direction. This result agrees with studies carried out by El Shayeb *et al*. (2001), which reported that increased Sn deposited on the electrode surface would move the corrosion potential towards the positive direction. At this moment, it is quite difficult to explain or justify the role of Mg_2Sn phases in shifting OCP towards the positive direction, but it is believed that ennoblement of OCP is due to the fact that Mg_2Sn has more noble potential than aluminium and this alloy has more anodic component of the local galvanic corrosion process. Although previous studies showed that Zn and Mg could shift OCP of aluminium to more negative potential, but it was strongly believed that the remaining of Mg_2Sn intermetallic in

the alloys give more influence in compensating this effect by continuing to shift the OCP value of heat-treated alloy in the positive direction (Ping *et al*., 2012; Cain *et al.*, 2019; Lee *et al.*, 2020).



(a) (b)

Figure 3: Surface morphology on the Al-5.5Zn-2.0Mg-2.0Sn alloy after the potentiodynamic test: (a) As-cast (b) Heat-treated

On the one hand, Table 2 shows that the OCP values of as-cast alloys are more negative than those heat-treated in the same media. The existence of more negative OCP for as-cast alloys could be due to the presence of secondary phases distributed heterogeneously, such as Mg_2Sn, in the Al-5.5Zn-2.0Mg-2.0Sn alloy. Besides that, the difference in microstructure would contribute to the inhomogeneities in anodic current density as shown by the wavy patterns in Figure 2.

The effect of heat treatment on the dissolution behaviour of as-cast and heat-treated Al-5.5Zn-2.0Mg-xSn alloys are shown in Figure 4. The presence of 2.0 wt.% Sn in the Al-5.5Zn-2.0Mg alloy clearly activated more electrochemical reactions at the aluminium surface relative to the 0.1 wt.% Sn alloy sample, as the corrosion potential shifted to significantly negative values. Potentiodynamic polarisation results with no sign of passive current showed that the presence of Sn in these alloys combined with heat treatment activated the electro-oxidation process when exposed to aggressive chloride containing medium. The activation effect can be seen in Figure 4 with no sign of passive current established during the scanning process.

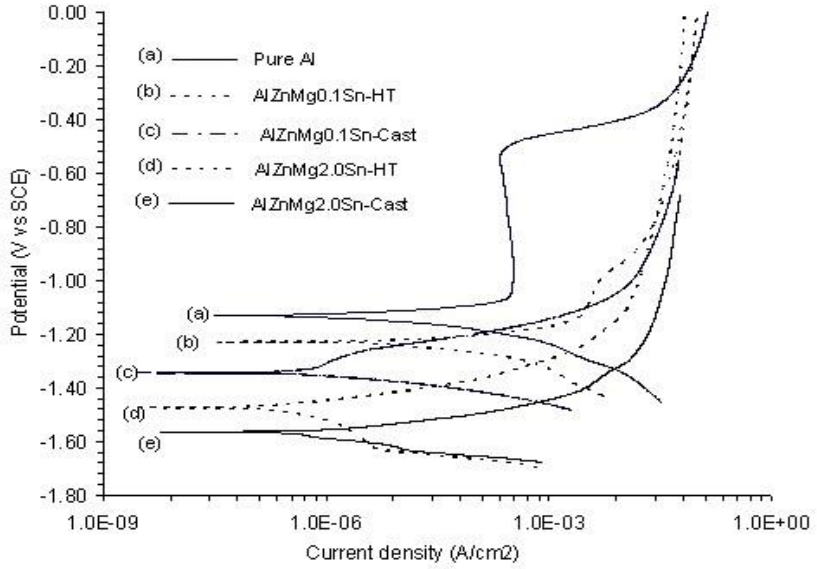


Figure 4: Potentiodynamic plots for as-cast and heat-treated Al-5.5Zn-2Mg-xSn (x = 0.1 & 2.0 wt. %) alloys in tropical seawater medium at 27 ºC with scan rate of 0.5 mV/s.

3.3 Current Capacity & Efficiency

The current capacity measurement is designed to simulate the service operating condition of sacrificial anodes. The basic anode properties obtained by this test are current capacity and current efficiency. The efficiency is defined as the ratio between actual and theoretical values of current capacity, and it is evaluated by the following equation (ASTM, 2013):

$$\text{Efficiency (E \%)} = \frac{\text{Actual current capacity}}{\text{Theoretical current capacity}} \text{ X } 100\% \quad (1)$$

where:

$$\text{Actual current capacity (Ah/kg)} = \frac{\text{Ah}}{\text{kg}} = \frac{\text{Ah}}{(M_1 - M_2)} \text{ X } 1{,}000 \quad (2)$$

where M_1 is the initial mass of aluminium alloy and M_2 is final mass (in g). The theoretical anode capacity is:

$$\text{Theoretical current capacity (Ah/kg)} = \frac{(96{,}480 \text{ C} / 3{,}600 \text{ S})}{\text{Equivalent weight of Al alloys, kg}} \quad (3)$$

The values of current capacity and efficiency for Al-5.5Zn-2.0Mg-xSn were calculated and listed in Table 3. It showed that current capacity and efficiency differs according to the chemical composition and heat treatment given. The presence of Sn, which has been classified as activator in these types of alloys, can cause a significant impact on current capacity values. Higher Sn content was needed for better efficiency in Al-Zn-Mg-Sn alloys. Heat-treated alloys showed better efficiency for the Al-5.5Zn-2.0Mg-xSn alloys. These results could be ascribed to the fact that heat treatment is capable of improving metallurgical properties, such as microstructure, grain size, solid solution and phase presence in the alloy. These in turn will give better electrochemical characteristics, more stable corrosion current density during OCP measurement and deliver higher current capacity.

Table 3: Current capacity efficiency for as-cast and heat-treated Al-5.5Zn-2.0Mg-xSn alloys in aerated tropical seawater at 27 °C and pH 8.1.

Samples (wt%)	Actual current capacity (Ah/kg)	Theoretical current capacity (Ah/kg)	Efficiency (E %)
Al-5.5Zn-2Mg-0.1Sn: As-cast	2279.33	2809.14	81.14
Al-5.5Zn-2Mg-0.1Sn: Heat-treated	2370.07	2809.14	84.37
Al-5.5Zn-2Mg-2.0Sn: As-cast	2262.92	2741.28	82.55
Al-5.5Zn-2Mg-2.0Sn: Heat-treated	2352.29	2741.28	85.81
Pure Al: As-cast	2115.72	2979.9	71.00

For example, heat-treated Al-5.5Zn-0.1Sn (wt.%) gave 2,181.96 Ah/kg current capacity as compared to the as-cast sample, which was 2,116.94 Ah/kg. The lower values of current capacity delivered in 2.0 wt%. Sn indicate that the anodes can supply only a partial amount of current for protection. The missing percentage represents the amount of charge lost that is no longer useful for cathodic protection or charge transfer process. As a result of the addition of active element Mg into the Al-5.5Zn-2.0Sn alloy, it has turned out to be the most active corrosion potential. The results from Table 3 also showed that the addition of 2.0 wt.% Sn to the Al-Zn-Mg alloy has shifted the OCP values to a more significant negative direction, thus it will deliver higher current capacity and efficiency.

Figure 5 illustrates the effect of heat treatment on the impedance response of OCP for the Al-5.5Zn-2.0Mg-2.0Sn samples after 3 h of immersion in tropical seawater medium at room temperature. Inspection of the data reveals that the impedance spectrum consists of a single capacitive semicircle. After the heat treatment process, the Al-5.5Zn-2.0Mg-2.0Sn sample showed a gradual decrease in the diameter of the loop corresponding to the low frequency region. The AC impedance of the Al-5.5Zn-2.0Mg-2.0Sn alloy composed of two parts: (i) a capacitive impedance in the middle frequency region; and (ii) a resistive impedance in the high and low frequency regions. The capacitive behaviour for both as-cast and heat-treated alloys in the middle frequency region might result from the interfacial electrochemical reaction, while the resistive behaviour in the high frequency region is because of solution resistance in the system. The impedance at the low frequency region usually corresponds to the charge transfer reaction (R_{pol}) at the electrolyte-alloy interface. The magnitude of the semicircle inversely measures the rate of dissolution process (corrosion rate) at the metal surface by means of the well-known Stern-Geary equation. Here, R_{pol} represents the corrosion resistance of electrode material in each solution. This means that larger R_{pol} implies lower corrosion rate. The qualitative R_{pol} values obtained from the EIS diagram are in good agreement with current capacity, which shows that heat-treated alloys give better current efficiency.

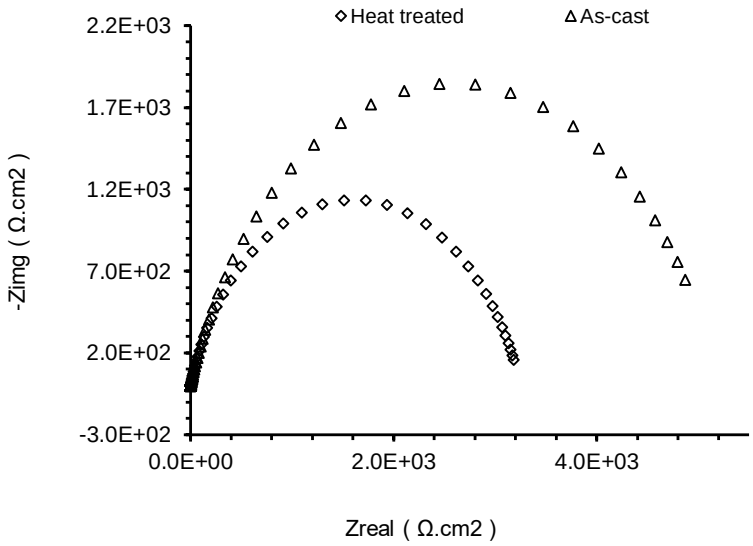


Figure 5: Impedance diagram obtained for the as-cast and heat-treated Al-5.5Zn-2.0Mg-2.0Sn (wt. %) alloys after 3 h of immersion in fresh tropical seawater.

4. CONCLUSION

The results presented in this work clearly showed that the presence of 2.0 wt.% Sn in the Al-5.5Zn-2.0Mg-xSn alloy combined with high temperature heat treatment gave a strong influence on its electrochemical behaviour. Heat treatment of Al-5.5Zn-2.0Mg-xSn at 550 ºC for 24 h was found to be effective in improving electrochemical properties by stabilising the OCP values, shifting the corrosion potential towards a more electronegative direction and activating electrochemical alloy dissolution. The current efficiency of the aluminium alloys was found to increase by increasing the Sn concentration in the alloy. This phenomenon can be correlated with the formation of Mg_2Sn intermetallic phase, which acts as a local cathode, and creates a galvanic action or micro-cell in the alloy that promotes better current capacity when exposed in chloride solution. The EIS spectra also show that the heat-treated Al-5.5Zn-2.0Mg-2.0Sn alloy produced less electrochemical resistance at the solution-alloy interface, which was responsible for reducing resistance to polarisation, and higher charge transfer activities or current efficiency.

ACKNOWLEDGEMENT

The authors would like to thank the Government of Malaysia for financing this R&D project. We also wish to thank the Faculty of Science & Technology, National University of Malaysia, for conducting the X-ray diffractometer (XRD) tests. We are pleased to acknowledge the cooperation and technical assistance given by STRIDE's officers and staff in improving the quality of this manuscript.

REFERENCES

Aballe, A., Bethencourt, M., Botana, F.J., Cano, M.J. & Marcos, M. (2003). Influence of the cathodic intermetallics distribution on the reproducibility of the electrochemical measurements on AA5083 alloy in NaCl solutions. *Corr. Sci.*, **45**: 161-180.

Acer, E., Çadırlı, E., Erol, H., Kırındı, T. & Gündüz, M. (2016). Effect of heat treatment on the microstructures and mechanical properties of Al–5.5Zn–2.5Mg alloy. *Mat. Sci. Eng. A*, **662**: 144-156.

American Society for Testing and Materials (ASTM) (2013). *Test Method for Laboratory Evaluation of Magnesium Sacrificial Anode Test Specimens for Underground Applications*. American Society for Testing and Materials (ASTM), Philadelphia, USA.

Azarniya, A., Taheri, A.K. & Taheri, K.K. (2019). Recent advances in ageing of 7xxx series aluminum alloys: a physical metallurgy perspective. *J Alloys Comp.*, **781**: 945-983.

Babu A.P., Huanga A. & Birbilis, N. (2021). On the heat treatment and mechanical properties of a high solute Al–Zn–Mg alloy processed through laser powder bed fusion process. *Mat. Sci. Eng. A*, **807**: 140857.

Barbucci, A., Cerisola, G., Bruzzone, G. & Saccone, A. (1997). Activation of aluminium anodes by the presence of intermetallic compounds. *Elec. Acta*, **42**: 2369-2380.

Berg, I.K., Gjonnes, J., Hansen, V., Li, X.Z., Waterloo, G. & Wallenberg, L.R. (2001). GP - zones in Al-Zn-Mg Alloys and their Role in Artificial Aging *Acta Mat*., **49**: 3443-3451.

Birbilis, N. & Buchheit, R.G. (2005). Electrochemical characteristics of intermetallic phases in aluminum alloys: an experimental survey and discussion. *J Elec. Soc*., **152**: B140-B151.

Breslin, B., Friery, L.P. & Carroll, W.M. (1993). Influence of impurity elements on electrochemical activated by indium. *Corr*., **49**: 895-902.

Bruzzone, G., Barbucci, A. & Gerisola, G. (1997). Effect of intermetallic compounds on the activation of aluminium anodes. *J Alloy Comp*., **247:** 210-216.

Cain, T.W., Glover, C.F. & Scully, J.R. (2019). The corrosion of solid solution Mg-Sn binary alloys in NaCl solutions. *Elec. Acta*, **297**: 564-575.

Campana, F. & Pilone, D. (2009). Effect of heat treatments on the mechanical behavior of aluminum alloy foams. *Scri. Mat*., **60:** 679-682.

Chen, M., Zheng, X., He, K., Liua, S. & Zhang, Y. (2021). Local corrosion mechanism of an Al-Zn-Mg-Cu alloy in oxygenated chloride solution: Cathode activity of quenching-induced η precipitates. *Corr. Sci.*, **191**: 109743.

Dai, P., Xian, L., Yanqing, Y., Zongde, K., Bin, H., Jinxin, Z. & Jigang, R. (2020). High temperature tensile properties, fracture behaviors and nanoscale precipitate variation of an Al–Zn–Mg–Cu alloy. *Prog. Nat. Sci.: Mat. Int.*, **30**: 63-73.

Dursun, T. & Soutis, C. (2014). Recent developments in advanced aircraft aluminium alloys. *Mat. Des*., **56**: 862-871.

El Shayeb, H.A., Abd El Wahab, F.M. & Zein El Abedin, S. (2001). "Electrochemical behaviour of Al, Al-Sn, Al-Zn and Al-Zn-Sn alloys in chloride solutions containing stannous ions. *Corr. Sci*., **43**: 655-669.

Fang, H.C., Luo, F.H. & Chen, K.H. (2017). Effect of intermetallic phases and recrystallization on the corrosion and fracture behavior of an Al-Zn-Mg-Cu-Zr-Yb-Cr alloy. *Mat. Sci. & Eng. A.*, **84**: 480-490.

Farooq, A., Hamza, M., Ahmed, Q. & Deen, K.M. (2019). Evaluating the performance of zinc and aluminum sacrificial anodes in artificial seawater. *Elec. Acta*, **314**: 135-141.

Flamini, D.O. & Saidman, S.B. (2012). Electrochemical behaviour of Al-Zn-Ga and Al-In-Ga alloys in chloride media. *Mat. Chem Phys*., **36:** 103-111

Gupta, A.K., Llyod, D.J. & Court, S.A. (2001). Precipitation hardening processes in an. Al-0.4%Mg-1.3%Si-0.25%Fe aluminum alloy. *Mat. Sci. & Eng. A*, **301**: 140-146.

He, H., Wu, X.D., Sun, C.R. & Li, L.X. (2019). Grain structure and precipitate variations in 7003-T6 aluminum alloys associated with high strain rate deformation. *Mat. Sci. Eng. A*, **745:** 429-439.

Hirsch, J. & Al-Samman, T. (2013). Superior light metals by texture engineering: optimized aluminum and magnesium alloys for automotive applications. *Acta Mat.*, **61**:818–843.

Khireche, S., Boughrara, D., Kadri, A., Hamadou, L. & Benbrahim, N. (2014). Corrosion mechanism of Al, Al–Zn and Al–Zn–Sn alloys in 3 wt.% NaCl solution. *Corr. Sci*., **87**: 504-516

Lee, D., Kim, B., Baek, S. M., Kim, J., Park, H.W., Lee, J.G. & Park, S.S. (2020). Microstructure and corrosion resistance of a Mg_2Sn-dispersed Mg alloy subjected to pulsed electron beam treatment. *J. Mag. Alloys* **8**: 345-351.

Li, Y., Yunlai, D., Shitong, F., Xiaobin, G., Keda, J., Zhen, Z. & Lin, S. (2021). An in-situ study on the dissolution of intermetallic compounds in the Al–Zn–Mg–Cu alloy. *J. Alloys Comp*. **829**: 154612.

Li, Z., Chen, L., Tang, J., Zhao, G. & Zhang, C. (2020). Response of mechanical properties and corrosion behavior of Al–Zn–Mg alloy treated by aging and annealing: A comparative study. *J. Alloys Comp*., **848:** 156561

Liu, P., Lulu, H., Qinhao, Z., Cuiping, Y., Zuosi, Y., Jianqing, Z., Jiming, H. & Fahe, C. (2021). Effect of aging treatment on microstructure and corrosion behavior of Al-Zn-Mg aluminum alloy in aqueous solutions with different aggressive ions. *J. Mat. Sci. Tech*. **64**: 85-98

Liu, F., Zhang, J., Sun, C., Yu, Z. & Hou, B. (2014). The corrosion of two aluminium sacrificial anode alloys in SRB-containing sea mud. *Corr. Sci*., **83**: 375-381

Mahdi, C.I. Daud, A.R., Mohd Yazid, A., Daud, M., Shamsudin, S.R., Nik Hassanuddin, N.Y., Mohd Subhi, D.Y. & Mohd Moesli, M. (2012). an electrochemical impedance spectroscopy study of Al-Zn and Al-Zn-Sn alloys in tropical seawater. *Key Eng. Mat*., **510-511**: 284- 292.

Mahdi, C.I., Ahmad, M.Y., Daud, A.R. & Daud, M. (2010). The effect of Sn on the impedance behaviour of Al-Zn alloys in natural chloride solution. *Key Eng. Mat*., **442**: 322-329.

Mahdi, C.I., Mohd Subhi, D.Y., Nik Hassanuddin, N.Y., Mohd Fauzi, M.N., Mohd Moesli, M., Osmera, I. & Irwan, M. (2011). Electrical characterization og As-Cast Al-Zn-Sn alloys for corrosion control application in tropical marine environment. *Def. S&T Tech. Bull*., **4:**119-130.

Maloney, S. K., Hono, K., Polmear, I. J. & Ringer, S. P. (1999). The chemistry of precipitates in an aged Al-2.1Zn-1.7 Mg at% alloy. *Scripta Mat.*, **41:** 1031-1038

Ping, W., Jianping, L., Yongchun, G., Zhong, Y., Feng, X. & Jianli, W. (2012). Effect of Sn on microstructure and electrochemical properties of Mg alloy anode materials. *Rare Metal Mat. Eng*. **41**: 2095-2099.

Pourgharibshahi, M. & Lambert, P. (2016). The role of indium in the activation of aluminum alloy galvanic anodes. *Mat. Corr*., **67**: 857-866

Pourgharibshahi, M. & Meratian, M. (2014). Corrosion morphology of aluminium sacrificial anodes. *Mat. Corr*., **65**: 1188-1193

Salinas, D.R., García, S.G. & Bessone, J.B. (1999). Influence of alloying elements and microstructure on aluminium sacrificial anode performance: case of Al-Zn. *J. Appl. Elec*., **29**: 1063-1071.

Shibli, S.M.A. & George, S. (2007). Electrochemical impedance spectroscopic analysis of activation of Al–Zn alloy sacrificial anode by RuO2 catalytic coating. *Appl. Surf. Sci*., **253**: 7510-7515,

Shibli, S.M.A., Jabeera, B. & Manu, R. (2007). Development of high-performance aluminium alloy sacrificial anodes reinforced with metal oxides. *Mat. Lett*., **61**: 3000-3004

Sperandio, G.F., Santos, C.M.L. & Galdino, A.G.S. (2021). Influence of silicon on the corrosion behavior of Al–Zn–In sacrificial anode. *J. Mat. Res. Tech*., **5**: 614-622

Wang, K., Yin, D., Zhao, Y., Atrens, A. & Zhaoa, M. (2020). Microstructural evolution upon heat treatments and its effect on corrosion in Al-Zn-Mg alloys containing Sc and Zr. *J. Mat. Res. Tech*., **9**: 5077-5089

Wang, H., Du, M., Liang, H. & Gao, Q. (2019). Study on Al-Zn-In alloy as sacrificial anodes in seawater environment. *J. Ocean Univ*., **18**: 889-895

Xia, Z., Zhang, W., Yang, X., Chen, T., Zhu, Y. & Ma, H. (2020). Influence of Sn, Cd, and Si addition on the electrochemical performance of Al–Zn–In sacrificial anodes. *Mat. Corr*., **71**: 585-592

Xu, D.K., Birbilis, N. & Rometsch, P.A. (2012). The effect of pre-ageing temperature and retrogression heating rate on the strength and corrosion behaviour of AA7150. *Corr. Sci*., **54**: 17-25.

Zakaria, K.A., Abdullah, S. & Ghazali, M.J. (2013). Comparative study of fatigue life behaviour of AA6061 and AA7075 alloys under spectrum loadings. *Mat. Design*, **49**: 48-57

FAILURE ANALYSIS OF HYDROGEN EMBRITTLEMENT IN SHIP GEARBOX HEX BOLTS

Mohd Moesli Muhammad[1], Sayed Roslee Sayd Bakar[2], Mohd Subhi Din Yati[1], Nik Hassanuddin Nik Yusoff[1] & Mahdi Che Isa[1]

[1]Marine Technology Research Group, Maritime Technology Division (BTM)
[2]Tropical Research Centre (TRC)
Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

Email: moesli.muhammad@stride.gov.my

ABSTRACT

This paper presents the results of a failure investigation of hex bolts in a ship gearbox that failed during cruising. Failure analysis procedures were employed in this investigation. The results showed that the fracture occurred due to brittle failure. Voids were found on the fracture surface with intergranular cracks along with on metal grain boundaries. Hydrogen measurement was carried out and discovered that all the failed samples contained hydrogen in the range of 0.5 to 1.5 ppm. The chemical composition obtained showed that the bolts are low alloy steel with hardness greater than 30 HRC. It was concluded that the brittle failure of the hex bolts was due to hydrogen embrittlement. Hydrogen was produced during the corrosion process of the bolts and diffused into the voids. As a result, during mechanical loading, bolt strength and ductility both decreased.

Keywords: *Failure analysis; brittle; voids; intergranular; hydrogen embrittlement.*

1. INTRODUCTION

Gearbox components are commonly important and extensively used in many engineering systems. In a ship propulsion system, the gearbox is an essential component to transfer power from the main engine to the propeller. Gearbox systems contain meshing teeth on pinions and wheels to drive a shaft that allows speed to be increased or decreased based on operational requirements. Failure of a gearbox system not only results in replacement costs such as bearings or shafts, but can also affect the performance of the propulsion system (Charmont & Samroeng, 2013; Goran *et al*., 2017; Onwuegbuchunam *et al*., 2020). Previous gearbox failure studies have revealed that common components include not only shafts, bearings, bolts and gears, but also lubricants (Hassan & Alam, 2010; Weigang *et al*., 2017). Axial cracks, macro pitting, scoring, wear, fretting corrosion, scuffing, misalignment and false brinelling are common failure modes on bearings and shafts (Charmont & Samroeng, 2013). A gearbox system can be fractured in a variety of ways including fatigue, brittle, ductile, mixed mode and shear fractures. Brittle fracture occurs quickly and with little deformation, whereas ductile fracture occurs slowly and with deformation before a part of the gear breaks. Mixed mode fracture refers to a fracture that is both brittle and ductile, while shear failure occurs when a line of force fails transversally rather than axially such as when a shaft twists. Fatigue is the formation and propagation of cracks as a result of repetitive or cyclic failure below the loads that would cause the materials to yield (ASM, 2005; Abdel & Mahmood, 2016).

In this paper, a study was carried out on a series of fractured hex bolts in a ship gearbox that failed during cruising. The hex bolts were used to connect the engine gearbox's outer shaft to the flange (Figure 1). The failed samples were thoroughly investigated in order to determine the cause of the failure.

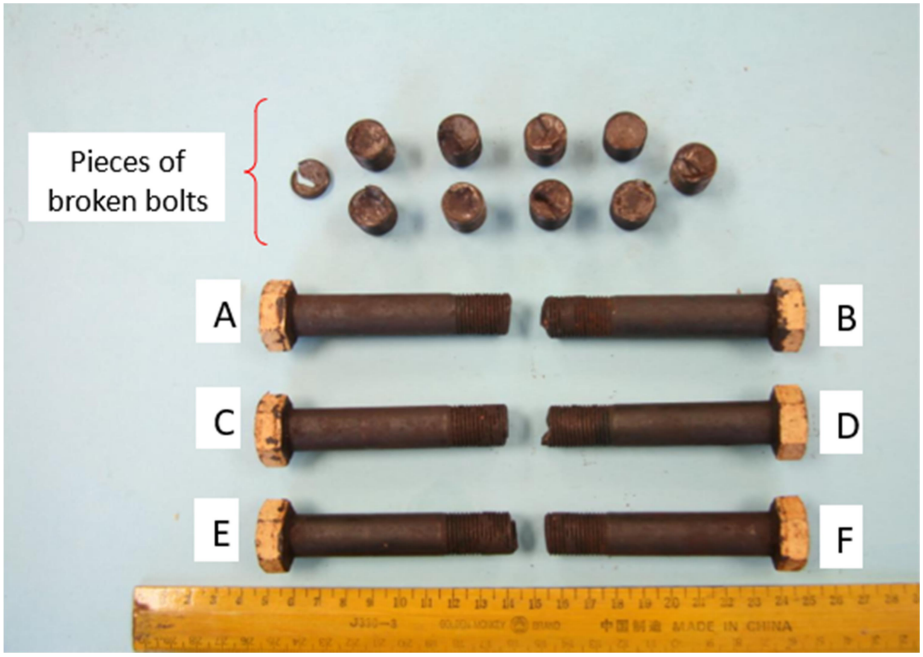


Figure 1: Six failed hex bolts, labelled as A, B, C, D, E and F, as well as pieces of ten broken bolts.

2. METHODOLOGY

Visual inspection of the samples was performed to acquire information on deformation and fracture surface (ASM, 1986; 1992; 1993). The samples were then cut with an abrasive cutter blade for additional analysis. After that, the samples were cleaned for 30 min in methanol using ultrasonic cleaning equipment. Macroscopic examinations were carried out with a Carl Zeis Stemi SV11 stereo microscope, with the post-processing images examined with an Axio Vision image analyser. Metallographic analysis was performed on the cross-sectioned samples of the fractured bolts. Grinding and polishing were used to prepare the samples, which were then exposed to 1% Nital etchant to disclose the microstructure. A scanning electron microscope (SEM) was used to examine the fracture surface at high magnification. A Wilson 2000 Series Rockwell hardness tester was used to measure the mechanical parameters of the samples' hardness (Figure 2). A Shidmadzu EDS 720 energy dispersive X-Ray fluorescent spectroscope was used to determine the chemical compositions.

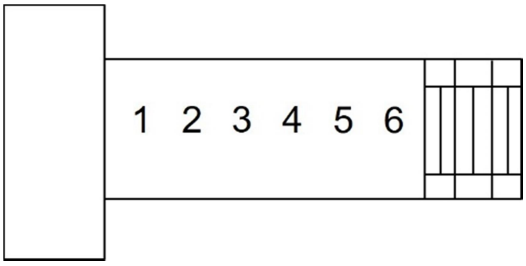


Figure 2: Locations of hardness testing on the failed hex bolt.

A Bruker G4 Phoenix DH hydrogen diffuser was used to measure hydrogen concentration. At high temperatures, this apparatus can detect hydrogen. Each fractured bolt was made up of two 3 g samples. In accordance with ISO (2018), the samples were placed in a quartz tube and heated to 400 and 600 °C. Hydrogen is released from the sample during the heating process and carried through the thermal conductivity cell with nitrogen as a carrier gas. The detector was then used a molecular sieve to filter out any interfering compounds, allowing only hydrogen to be measured.

3. RESULTS

3.1 Visual Examination

The fractured hex bolts were examined visually to study the failure mode. All the samples revealed that the external surfaces were rusted and the fractures took place in the threaded regions. Figure 3 shows the fracture surfaces of the bolts. The surfaces were found to be flat, with very little shear lip or plastic deformation. According to the examination, the flat areas that cover most of the fracture surfaces could suggest the brittle failure of the bolts. This failure demonstrates that the hex bolts failed suddenly with minimal elastic or plastic deformation before rupture.

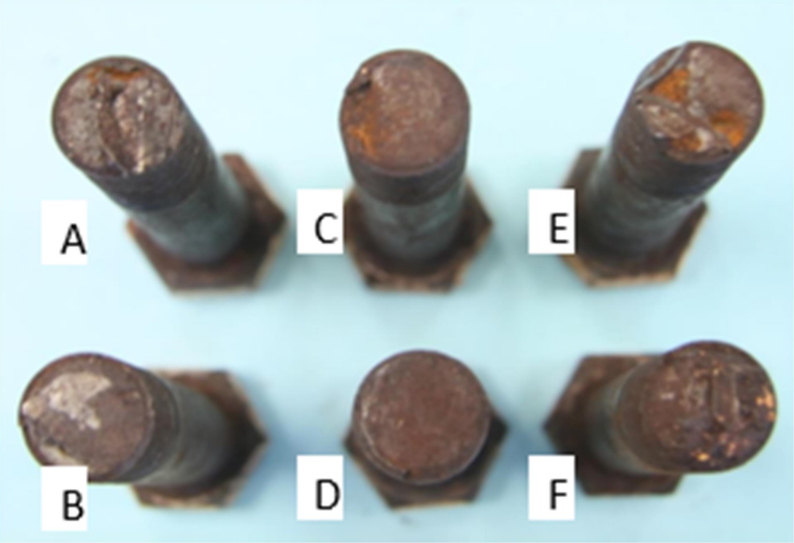


Figure 3: Fracture surfaces of hex bolts.

3.2 Macroscopic Examination

A Carl Zeiss Stemi SV 11 stereo microscope and Axio Vision software for image post-processing analysis were used out to enlarge the fracture surface. The examination shows that the regions of the flat zones are dominated on the fracture surfaces rather than the rough zones. The rough zones indicate that elastic or plastic deformation has occurred. The flat surface of brittle fracture is believed to be the source of the beginning of crack propagation. Due to low magnification of the macroscopic examination, no cracks were found on this flat zone. The images captured on the fracture surfaces of the failed hex bolts of samples B and D are shown in Figures 4(a) and 4(b) respectively.

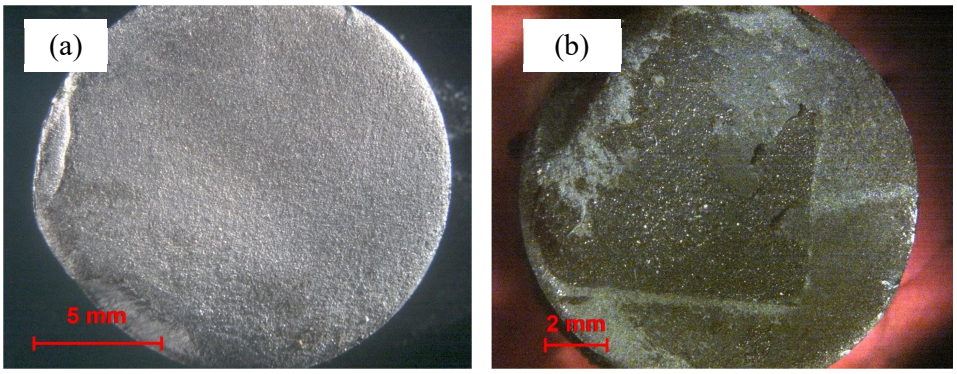


Figure 4: Macroscopic examination on the fracture surface of failed hex bolts: (a) Sample B (b) Sample D.

3.2 Microscopic Examination

The failed samples were examined under an inverted microscope for metallographic investigation and observation under higher magnification using a SEM. The images of metallography are shown in Figure 5. Observation of the microstructure found that individual and cluster voids with different dimensions are visible on the surface. The cluster voids (Figure 5(a)) with elongated and rounded shapes are approximately 200 µm. Figure 5(b) shows that elongated voids of approximately 60 µm in length. Further examination was carried out under higher magnification using the SEM to determine the location of voids in the microstructure. Based on SEM analysis (Figure 6), the voids occurred on the intergranular that can be observed on the grain boundary associated with the secondary crack, which is the effect of crack propagation from the intergranular. This intergranular cracking in the microstructure indicates the possibility of being caused by corrosion or hydrogen embrittlement (Sanchez *et al*., 2015; Le *et al*., 2018).

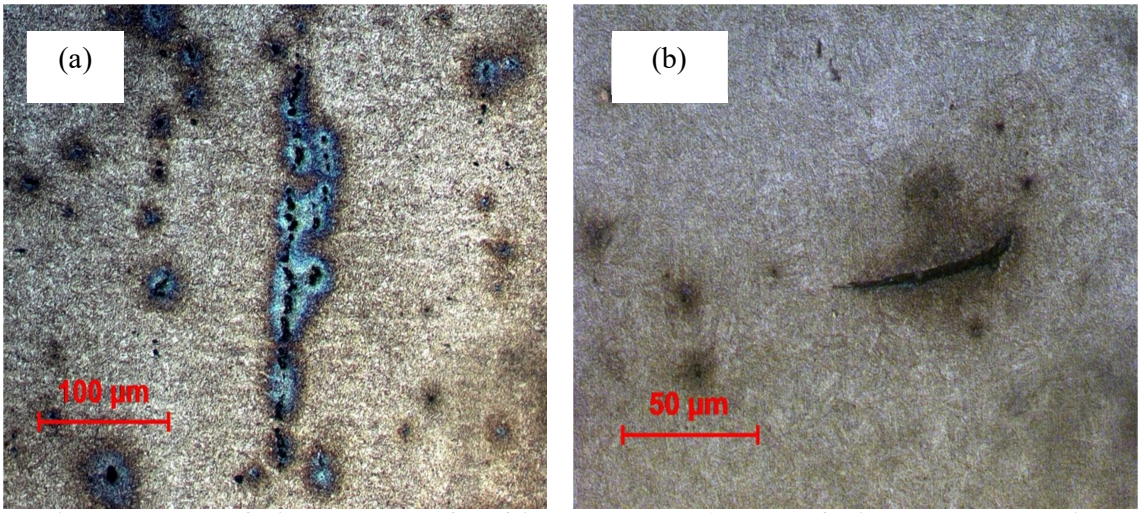


Figure 5: Images of microstructures of the failed samples: (a) A cluster of voids (b) Elongated void.

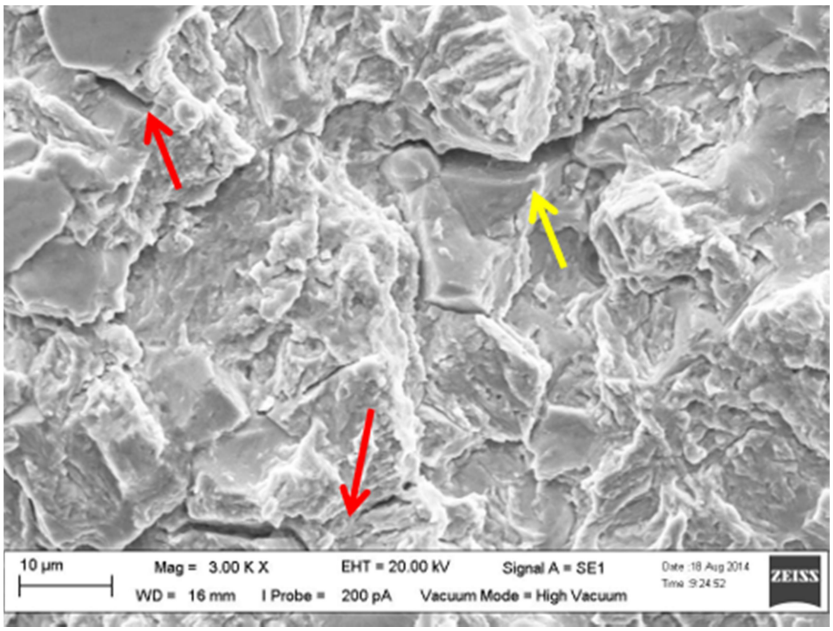


Figure 6: SEM image of the voids. The intergranular cracks are indicated by the red arrows while the yellow arrow indicates a secondary crack.

3.3 Materials Composition

Chemical analysis of the failed samples was carried out using a Shimadzu EDS 720 energy dispersive X-ray fluorescent spectroscope. The test results obtained are shown in Table 1. Based on the chemical analysis, the hex bolts were found to contain chromium (Cr), manganese (Mn) and carbon (C) as alloying elements with Iron (Fe) as a major element in the matrix alloy. The content of Cr and Mn in the steel alloy increases the resistance of the hex bolts to corrosion. The chemical composition results show that the hex bolts are suitable for use in the gearbox and are not the cause of the failure.

Table 1: Chemical composition of the failed hex bolts.

Element	Cr	Mn	C	Fe
Composition (%)	0.32	1.02	0.23	Balance

3.4 Materials Hardness

The hardness of the failed hex bolts was measured by using a Wilson 2000 Series Rockwell hardness tester. The hardness test was carried out in six different locations, with the results presented in Figure 7**.** The range of hardness obtained is 30 to 35 HRC. All the failed bolts show evenly distributed values of hardness except for Sample C, where the values are higher when close to the threaded region (Locations 4 to 6). The highest value of hardness is obtained for Sample B.

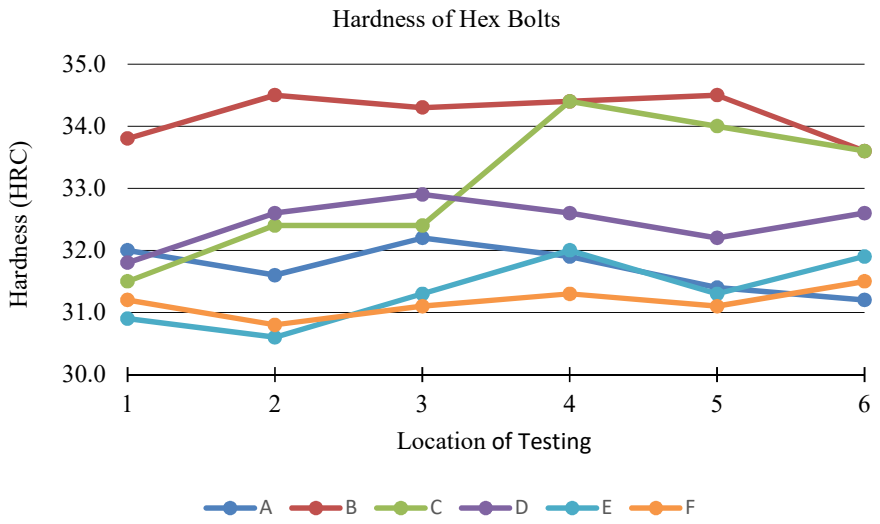


Figure 7: Graph of hardness testing results.

3.5 Hydrogen Concentration

Hydrogen content was measured to determine the concentration of hydrogen in the failed samples. The measurement was carried out using a Bruker G4 Phoenix DH hydrogen diffuser. The concentration of hydrogen was measured at temperatures of 400 and 600 °C. The sum of concentrations is shown in Table 2. The highest concentration of hydrogen was obtained for Sample F followed by Sample A with the values of 1.4216 and 0.5117 ppm respectively. The concentrations for Samples B, C, D and E are in the range of 0.1 to 0.4 ppm.

Table 2: Hydrogen concentration in the failed hex bolts.

Hex Bolt Sample	Parts Per Millions (ppm)		
	400 °C	600 °C	Total
A	0.4970	0.0147	0.5117
B	0.1327	0.2035	0.1362
C	0.2236	0.0601	0.2837
D	0.2214	0.1337	0.3551
E	0.2287	0.1033	0.3320
F	0.8577	0.5639	1.4216

4. DISCUSSION

According to the visual and microscopic examinations, the fracture surfaces of failed hex bolts revealed that the mode of failure was brittle. This is due to the evidence that most of the fracture surfaces were dominated by smooth and flat regions. The brittle failure shows that the hex bolts experienced sudden failure during operation with low intensity of elastic or plastic deformation (rough surface) before the final rupture. This type of brittle failure is undesirable in any mechanical system because it can occur suddenly and without warning, resulting in catastrophic or complete system failure (Bill, 2013; Knott, 2015).

Further investigations of the failed samples were carried out using microscopic examinations, which included metallographic analysis and SEM observation. On the flat regions of the failed samples, cluster and individual microvoids were clearly visible. These voids are believed to be manufacturing defects of the hex bolts. The rounded and elongated voids on the microstructure are the source of high stress concentration. This was widely assumed to be the cause of the failure. These voids exhibit as a crack initiation point and propagate when stress is applied during operation. Under high magnification using SEM, these voids are found between the boundaries of material grains. According to Amit (2017), this phenomenon is categorised as intergranular induced cracking, which occurs in the microstructure of hex bolts due to corrosion and hydrogen embrittlement. Sandeep & Manisah (2019) reported that hydrogen embrittlement can be attacked on the structure or mechanical system when these three factors exist, which are mechanical loading, high strength materials and external environment. The chemical composition in Table 1 indicates that the hex bolt is low alloy steel. The hardness values of the samples are above 30 HRC, indicating that this alloy steel is high strength material. Low alloy steel associated with high strength hardness is among susceptible materials to hydrogen embrittlement as well as nickel and titanium alloys (Motomichi *et al*., 2017; Enyinnaya & Ubong, 2018).

The hydrogen concentration was then measured, with Sample E having the highest value of 1.42 ppm. According to Murakami *et al*, (2010), hydrogen embrittlement can occur at a concentration greater than 1 ppm because hydrogen atoms can diffuse inside a material under atmospheric pressure. Hydrogen is diffused on the tip of voids in this case, which can reduce a material's cohesive strength and ductility.

5. CONCLUSION

The failure analysis was carried out on the failed ship gearbox hex bolts. The visual examination revealed that the failure mode was brittle fracture due to large areas of the flat surfaces as compared to the rough surfaces. Detailed analysis in the macroscopic examination showed that individual and cluster voids exist in the area of the flat zones on the fracture surfaces. Further examination using higher magnification during the microscopic examination indicated that intergranular induced cracking occurred along the material grain boundaries. The chemical composition of the samples

showed that the bolts are low alloy steel, which contain Cr, Mn and C as alloying elements, with Fe as a major element in the matrix alloy. The hardness values of the failed samples were evenly distributed except for Sample C. The hydrogen concentration measurement demonstrated that the highest value was obtained for Sample F followed by Sample A. Based on the results of the investigation, it can be concluded that the brittle failure occurred due to hydrogen embrittlement and it was strongly believed that the hydrogen was generated from corrosion of gearbox bolts.

ACKNOWLEDGEMENT

The authors would like to thank the officers and staff from the various laboratories at the Maritime Technology Division (BTM), Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia for their technical support during investigation works.

REFERENCES

Amit, K., Manish, V. & Vishal, P. (2017). A review on failures of industrial components due to hydrogen embrittlement & techniques for damage prevention. *Int. J. Res. Appl. Sci. Eng. Technol.*, **12**: 1784-1792.

Abdel, S.H.M. & Mahmood, A. (2016). *Handbook of Materials Failure Analysis with Case Studies from the Oil and Gas Industry*. Butterworth Heinemann, Waltham, US.

ASM (American Society for Metals) (1986). *ASM Metals Handbook, Volume 11: Failure Analysis and Prevention*. American Society for Metals (ASM), Ohio, US.

ASM (American Society for Metals) (1992). *ASM Handbook of Case Histories in Failure Analysis, Vol. 1 and Vol. 2*. American Society for Metals, Ohio, US.

ASM (American Society for Metals) (1993). *Handbook of Case Histories in Failure Analysis, Vol. 2*. American Society for Metals, Ohio, US.

ASM (American Society for Metals) (2005). *Failure Analysis of Engineering Structure, Methodology and Case Histories*. Ohio, US.

Bill, E. (2013). The stronger the better is not necessarily the case for fasteners. Fastener Tech., **11**: 29-30.

Charmont, M. & Samroeng, N. (2013). Failure analysis of a two high gearbox shaft. *Procedia Soc. Behav. Sci.,* **88**: 154-163.

Enyinnaya, O. & Ubong, E.,J. S. (2018). Hydrogen related degradation in pipeline steel: A Review. *Int. J. Hydrogen Energy,* **43**: 14584-14617.

ISO (International Organization for Standardization) (2018). *ISO 3690:2018 - Welding and Allied Processes: Determination of Hydrogen Content in Arc Weld Metal*. International Organization for Standardization (ISO), Geneva, Switzerland.

Hassan, S., F. & Alam, M., R. (2010). Failure analysis of gearbox and clutch shaft from a marine engine. *J. Fail. Anal. Preven*., **10**:393-398.

Jotram, P., Gopal, S., & Prakash, K., S. (2015). A study common failure of gears. *Int. J. Innov. Res*., **2**: 2349-6002.

Knott J. (2015). Brittle fracture in structural steels: Perspectives At different size-scales. *Phil. Trans. R. Soc. A* **373**: 20140126.

Le, L., Mojtaba, M., Chun-Qing, L. & Dilan, R. (2018). Effect of corrosion and hydrogen embrittlement on microstructure and mechanical properties of mild steel. *Constr. Build. Mater.*, **170**: 78-90.

Motomichi, K., Eiji. A., Young, L., Dierk, R. & Kaneaki, T. (2017). Overview of hydrogen embrittlement in high-Mn steels. *Int. J. Hydrogen Energy,* **17**: 12706-12723.

Murakami, Y., Kanezaki T. & Mine, Y. (2010). Hydrogen effect against hydrogen embrittlement. *Metall. Mater. Trans. A.*, **41A**: 2548 -25462.

Onwuegbuchunam, D.E., Ogwude, I.C., Igboanusi, C.C., Okeke, K.O. & Azian, N.N. (2020). Propulsion shaft and gearbox failure in marine vessels: A duration model analysis. *J. Transp. Technol.*, **10**: 291-305.

Sanchez, J., Leea, S.F., Martin-Rengel, M.A., Fullea, J., Andrade, C. & Ruiz-Hervías, J. (2015). Measurement of hydrogen and embrittlement of high strength steels. *Eng. Fail. Anal.,* **59**: 467-477.

Sandeep, K.D. & Manish, V. (2018). Hydrogen embrittlement in different materials: A review. *Int. J. Hydrogen Energy,* **43**: 21603-21616.

Weigang, H., Zhiming, L., Dekun, L. & Xue, H. (2017). Fatigue failure analysis of high speed train gearbox housings. *Eng. Fail. Anal*., **73**: 57-71.

EVALUATION OF WELD DEFECT SIGNAL FEATURES USING ULTRASONIC FULL WAVE PULSE ECHO METHOD

Suhairy Sani[1,2], Mohamad Hanif Md Saad[1], Nordin Jamaludin[1], Norsalim Muhammad[3], Siti Fatahiyah Mohamad[4] & Megat Harun Al Rashid Megat Ahmad[2]

[1]Department of Mechanical and Materials Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia (UKM), Malaysia
[2]Industrial Technology Division, Malaysian Nuclear Agency, Malaysia
[3]Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[4]Radiation Processing Technology Division, Malaysian Nuclear Agency, Malaysia

[*]Email: suhairy@nm.gov.my

ABSTRACT

Inspecting weld structures is crucial to determine the quality of welds to prevent failure in daily operations. Ultrasonic testing (UT) is an effective and non-destructive test that has become one of the most suitable methods for detecting defects in welded structures. This work presents a method to determine the most suitable feature extraction using full-wave pulse echo (FWPE) to distinguish the various types of weld defects in single V-joint plates. Six types of certified carbon steel welding materials with induced defects are employed in this study. Normal condition welding (no defects) is used as a reference. A portable ultrasonic equipment generates the FWPE signal using a shear wave angle beam probe. The measured FWPE signals in the configured scanning area are filtered to reduce noise and subsequently recorded to extract the wave characteristics. Manual transverse scan probe movement is performed in the defect areas on both sides of the weld for each welding specimen to obtain the B-scan results. The outcomes revealed that characteristic extraction for the time domain, maximum amplitude, depth, number of peaks and skew can be used to determine the type of defect on the welded specimens tested. Scanning both sides of the weld provides additional information that helps distinguish the type of defect in terms of orientation and position.

Keywords: *Ultrasonic testing (UT); full-wave pulse echo (FWPE); weld defects; signal features extraction; transverse scan probe movement*

1. INTRODUCTION

Structural weld inspection is a crucial method for determining whether the quality of welds meets the quality standards for operation. Defects in welds can affect their condition. This depends on three factors, which are the type, size and location of the defect (Singh, 2012; Consonni *et al.*, 2014, Kim *et al.*, 2021). One of the most widely used non-destructive evaluation (NDE) methods of weld defects is ultrasonic testing (UT). The UT method, consisting of pulse-echo, is the most reported method applied to characterise A-scan weld defects (Hoseini *et al.*, 2013, Hernandez *et al.*, 2018, Sudhamayee, 2019).

The UT pulse-echo method utilises high frequency sound wave pulses that can be transmitted into a weld. Any weld defects will cause echoes in pulses. The signal patterns are then processed to obtain information on the size, shape and orientation of the defects (Seyedtabaii, 2012; Zolfaghari *et al*., 2013, Cai *et al.*, 2020). The two types of A-scan signal patterns used to characterise weld defects are unrectified and full rectification signals. They are commonly known as RF-wave pulse-echo (RFPE) and full-wave pulse-echo (FWPE) respectively. Signal processing classification using ultrasonic pulse-echo mostly utilises RF-wave signals to extract specific features according to the time and frequency domains. This is mostly accomplished through pre-processing, either with Fourier

transform (FT) or Wavelet transform (WT), as the inputs in the classifier (Sambath *et al.*, 2011; Singh *et al.*, 2015; Hu *et al.*, 2018). However, studies using the FWPE method for weld defect characterisation remain considerably scarce, even though this method is used as the standard procedure in practical training of ultrasonic inspection (ISO , 2012). The main reason is possibly due to the unavailability of frequency domain feature extraction from the FWPE enveloped signals compared to the RFPE waveforms. Several reports have highlighted feature extraction using FWPE, but these works did not elaborate on the details regarding the applied algorithms (Zolfaghari *et al.*, 2013, Shahriari *et al.*, 2013).

Feature extraction is a process that determines the key signatures of signal waveforms that cause dimensional reduction of the original dataset while preserving significant information of the signals. The reflected waveform characteristics can be obtained from the reflected amplitude, number of peaks, location of the reflection, symmetrical distribution of the enveloped signal, duration of echo, as well as the rising and falling times of signals. From these physical features, various signal characteristics can be obtained for different types of defects. Signal characteristics can also be statistically formulated as the area under the enveloped signal (Birks *et al*., 1991). All features must have carefully prepared definitions and thresholds so that automated feature extraction and pattern recognition can take place. Threshold selection is based on experience as well as the amount of random and material noise (Lingvall & Stepinski, 1999; Droubi *et al.*, 2017, Ma *et al.*, 2020).

In this study, UT is obtained by inspecting V-welded steel plates using the FWPE technique through manual transverse probe movement. The probe moves forwards and backwards, perpendicular to the weld axis (B-scan). The aim of this study is to determine the most valuable extraction features for comparing different types of weld defects. The extraction features are evaluated in the six most common conditions that are usually encountered in weld joints; i.e., 1) centre crack (CCr); 2) slag inclusion (SI); 3) lack of fusion (LOF); 4) cluster porosity (Po); 5) lack of penetration (LOP); and 6) root crack (RCr) (Qidwai & Bettayeb, 2009, Sambath *et al.*, 2011, Lin *et al.*, 2019). Feature data for normal conditions (N - no defects) is also used as the baseline reference. Table 1 summarises the outcomes of the six conditions due to the transverse probe movement. Their schematic drawings are presented in Figure 1.

Table 1: Characteristics of defect types (IAEA , 2018)

Defect Type Description	Signal Characteristics
Crack (Cr): Irregular and multi-facetted profiles at the surface, inside the weld material or at heat affected zones.	• Multiple peak reflectors, usually of high amplitude, are dependent on the type of crack and size. The echo has a fir tree appearance. • They completely reflect the sound energy in a particular direction.
Lack of Fusion (LOF): Fusion between the welding material and base parts to be welded does not occur properly.	• Echo is large, single and narrow at time base when on sidewall. Poor echo from opposite side. • A planar target is located at the edge of the bead, which can be confirmed by irradiating the weld from the other side.
Cluster Porosity (Po): Group cavities originating from gas bubbles trapped within the welding region, leaked during the solidification process.	• Consists of multiple peak echoes. Low intensity height at time base due to numerous ranges. • Gives rise to numerous tiny echoes, depending on the number and distribution of the pores.
Lack of Penetration (LOP): Weld material penetration does not occur through the entire thickness of the base metal at the root area.	• Similar to corner reflector with large, narrow echo from both sides. • Located at the centre of the weld root in a single V-butt joint. • The echo height will be nearly equal when checking from either side.

Slag Inclusion (SI): Non-metallic component that exists in the weld, on the surface or between layers	• The echo from this flaw can be as high as a crack or lack of fusion, but the shape of the echo is quite different. • Due to its rugged surface, which offers many small targets at different distances, the echo rises like a pine tree from the zero line of the screen.

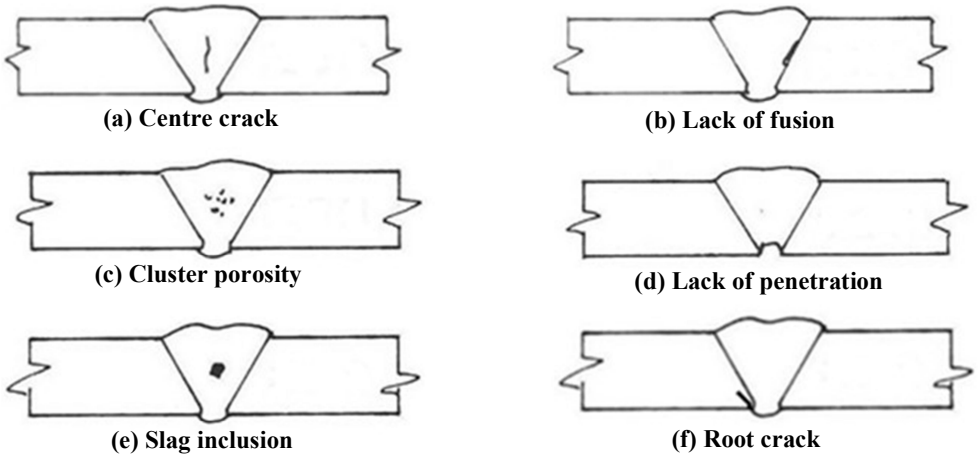


Figure 1: Schematic drawings of the six defect conditions of V-welded steel plates (Sonaspection, 2017).

2. METHODOLOGY

2.1 Specimens and Equipment

This study utilised specimens of six different weld defects in single V-butt weld plates (Sonaspection International Ltd.). Figure 2 presents the specimens that only contain a specific type of defect in each plate. These specimens are composed of carbon steel with a thickness of 10 mm. The defects are about 25 mm in length. Ultrasonic measurements of these defects were conducted using a 60° angle beam ultrasonic probe (GB Inspection Systems Ltd.) and a portable USBUT-350 ultrasonic pulser / receiver with an analogue to digital converter (Figure 3). The frequency employed for all the experiments is 4 MHz (Gang *et al.*, 2002; Seyedtabaii, 2012; Shahriari *et al.*, 2013). An in-house device control and data acquisition system was developed in C# programming language for data acquisition. The system was applied to obtain the pulse-echo signals used to generate the B-scan results at a sampling rate of 50 MHz and scanning sensitivity of 43 dB. Calibration of the sensor sensitivity was conducted using a V2 block prior to the scanning process (Drury, 2004, Sani *et al.*, 2019, Cai *et al*., 2020). The processed signal underwent two types of filters before characterisation was performed. The first filter is the low-pass filter (LPF), which can be appropriately selected from the data acquisition card. The second filter is the moving average filter (MAF), symmetrically used to further smoothen the signal until a single peak appears. Scanning was accomplished from half skip distance (HSD) to full skip distance (FSD) as well as half of the weld cap on sides A and B, as shown in Figure 3. Scanning was also conducted in the manual transverse movement, which kept the probe at 90° to the weld cap. B-scans at five locations on both sides were obtained, resulting in a total of ten B-scans for each specimen. The normal condition (no defects) was used as reference to compare the obtained results.

Figure 2: Defect specimens

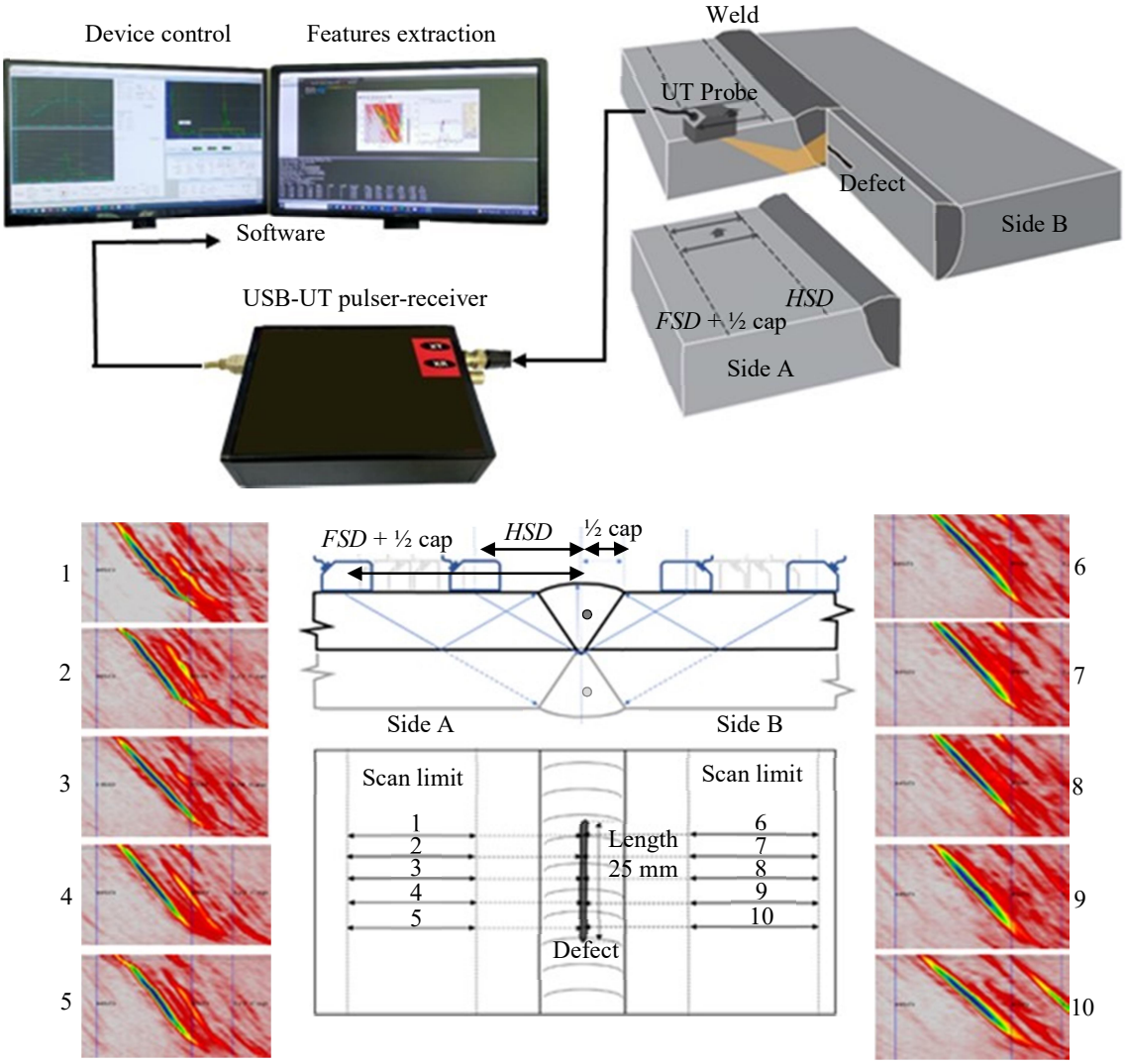


Figure 3: Experimental setup.

2.2 Features Extraction

The FWPE data measurement was processed in Python programming language to extract the features of signals and display the results in the form of data images (B-scans). Figure 4 displays the outcomes of the B-scans. The beam path length and waveform dataset are in the *x*- and *y*-axes, respectively. Each B-scan dataset recorded the A-scan data at 0.02 mm or 0.01 µs intervals for the image displayed, recording more than 500 waveform data points.

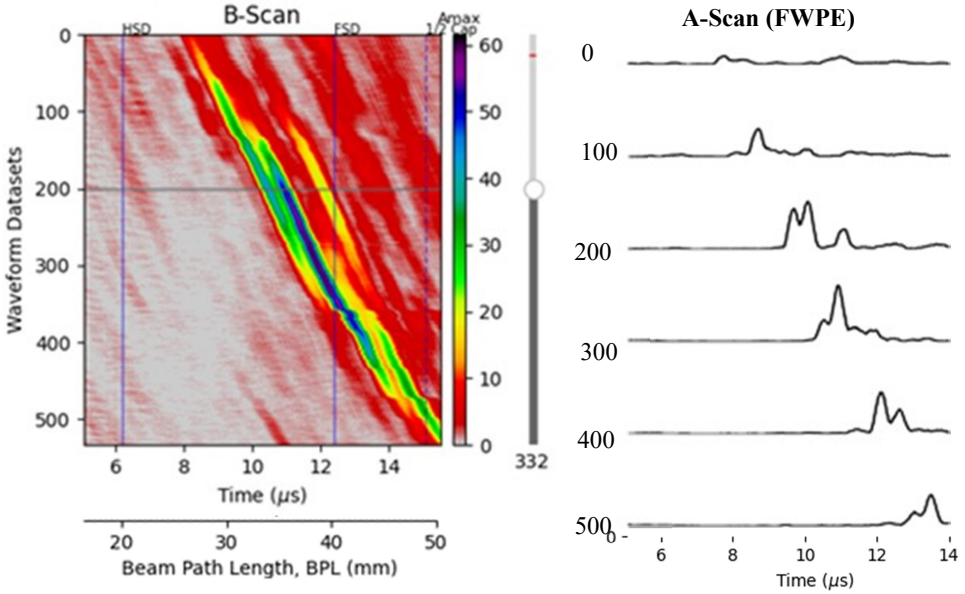


Figure 4*:* B-scan data waveform and A-Scan envelops.

Transverse movement was conducted on both sides of the weld within the range of the beam path length (*BPL*) while observing the defect echoes. All planar defects, including cracks, lack of penetration and lack of side wall fusion, are difficult to distinguish from the information based on the echo height and shape of the measured signals. The nature of the defects was precisely determined from their locations inside the welds according to (IAEA, 2018).

The locations of the defects were obtained from *BPL*, which was determined using a V2 block. It was deduced from the time of flight (t (µs)) of the pulse and the sound velocity (C (mm/µs)), as follows:

$$BPL = {}^{C \times t}/_{2} \tag{1}$$

Figure 5 presents a diagram of the weld specimen with the thickness of sample (T_i) and *BPL*, which is the distance from i to k, where i is the index point of the probe with angle θ and k is the location of defect in the apparent depth.

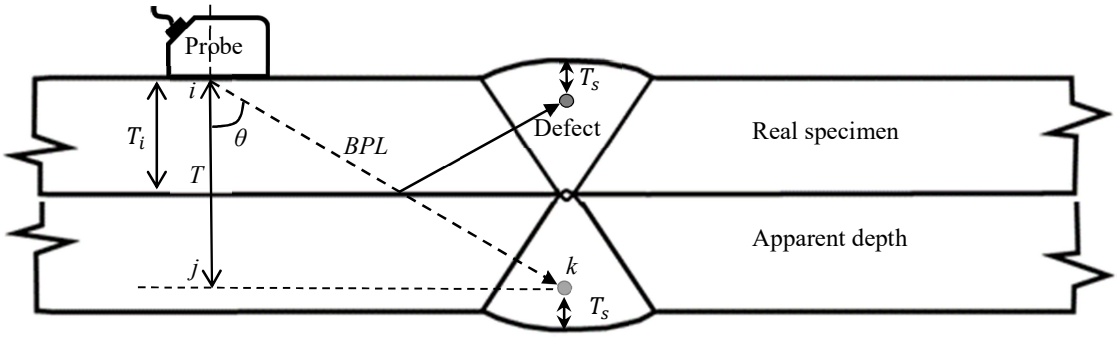


Figure 5: Diagram of the weld specimen from the reflected signal

The skip distance (*SD*) is the horizontal distance between the index point of the probe and centre of the weld cap, indicated as the distance from *j* to *k,* while *T* is the vertical distance between the index point probe to point *j*. Both parameters, *SD* and *T*, are shown in the following relationships:

$$SD = BPL \times \sin\theta \tag{2}$$

$$T = BPL \times \cos\theta \tag{3}$$

Next, the actual depth (T_s) is obtained as follows:

$$T_s = 2T_i - T = 2T_i - (BPL \times \cos\theta) \tag{4}$$

T_s is the vertical distance of the top cap of the weld.

Figure 6 presents the feature extraction in the developed software. The maximum amplitude (A_{max}) was obtained using the maximum peak in the signal processing. Data was extracted from FWPE, equal or above 10% of the full screen height (FSH). The extracted data is represented by the blue line in the figure. The features of rise time (R_t), fall time (F_t) and duration time (D_t) were directly extracted from the FWPE envelop pattern (Lingvall & Stepinski, 1999; Drury, 2004; Droubi *et al.*, 2017).

Peak identification and skewness were computed using Python library by utilising the find peaks method. Two different settings were employed in the peak identification process: 1) for N_{10}, FWPE was extracted equal to or greater than 10% of the FSH, and the dataset was then computed by setting the parameters (distance=none, prominence=2) in the find peaks method to determine the maximum peak; and 2) for N_{90}, FWPE was extracted equal to or greater than 90% of the FSH, and the parameters were changed (distance=1, prominence=none) to identify all peaks for N_{90}. The skewness of N_{10} was calculated using the statistics skew method by setting the parameters (axis=0, bias=True) (Zwillinger & Kokoska, 2000).

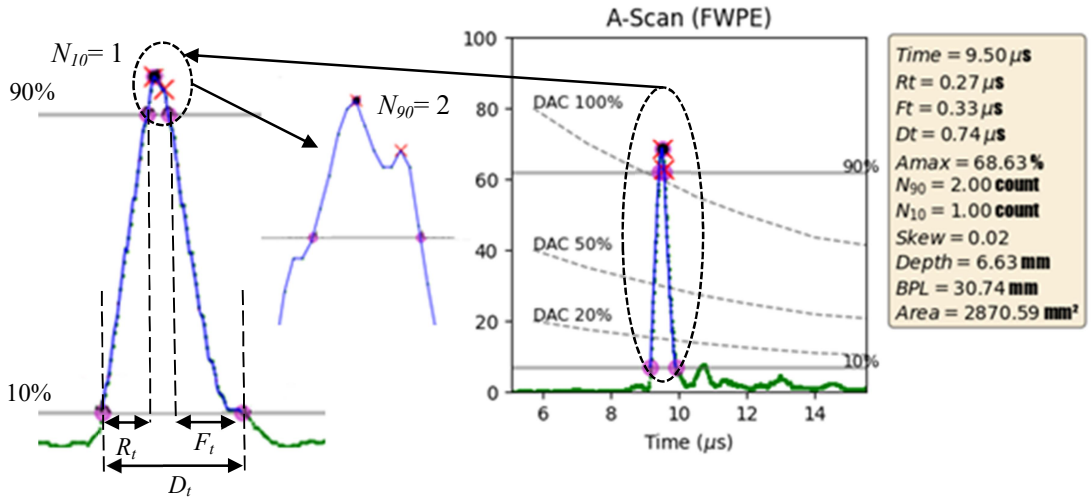


Figure 6*:* Features extraction from FWPE.

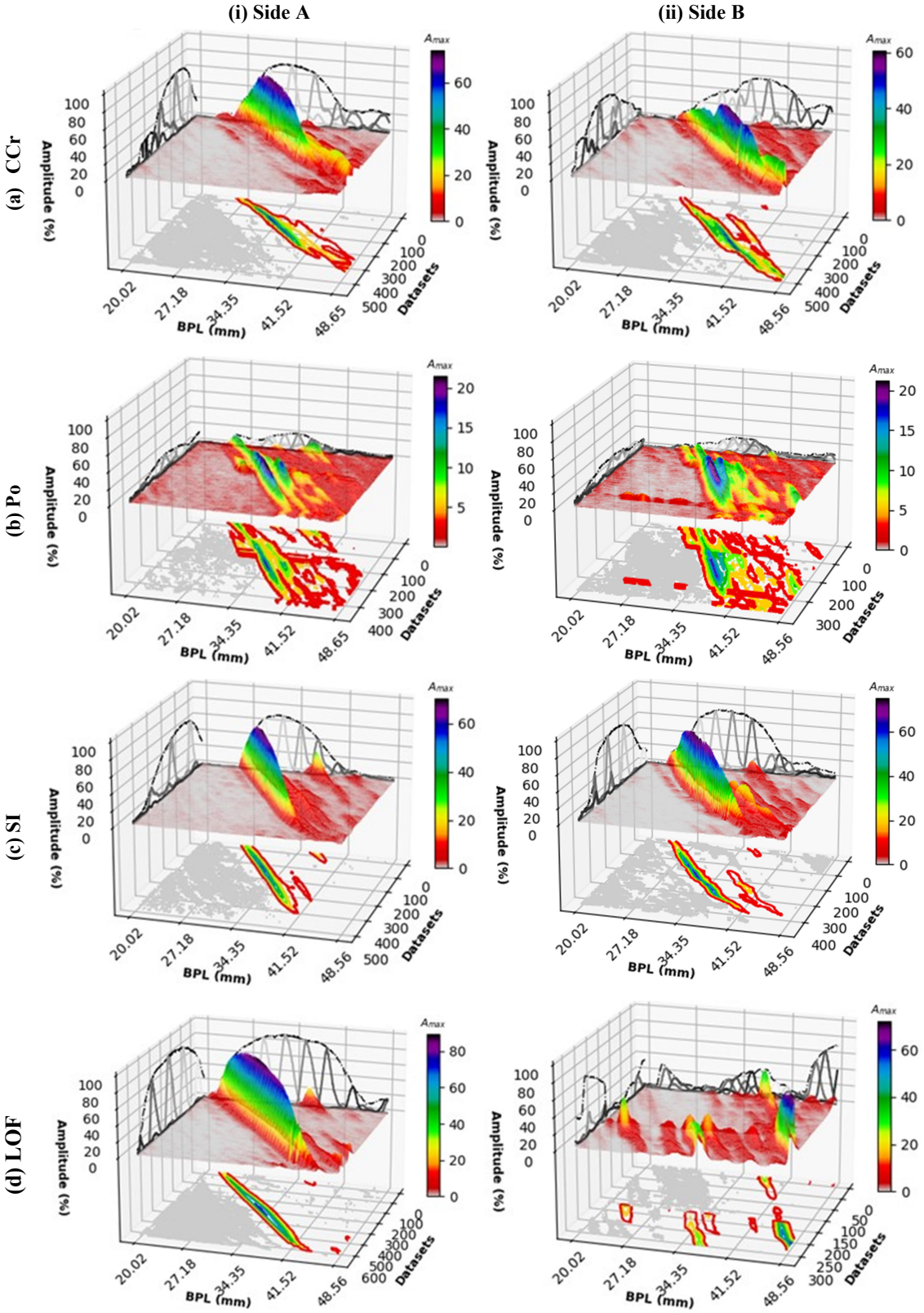
3. RESULTS AND DISCUSSION

All measured FWPE data for the defects were processed, as shown in Figure 7. The results are displayed in 3D view of the B-scan, together with the 2D contour view at the bottom. The amplitude of the 3D view contains a projection of both *BPL* and amplitude of the datasets from the B-scan data.

The comparison of both sides of the scan profiles (sides A and B) indicates that the CCr, Po, SI and LOP (Figures 7(a), 7(b), 7(c) and 7(e)) have similar contour intensity plots. However, both LOF and RCr have different scan profiles (Figures 7(d) and 7(f)). These differences are due to the FWPE reflection signal being influenced by the location, orientation, reflection surface and shape of defects. The characterisation of both sides A and B is essential so that the data obtained can provide more information to distinguish all types of defects rather than only obtaining information of a single side.

Analysis of the amplitudes indicated that the lowest amplitude value is for plates Po and N on both sides as compared to the other defects. LOF presented the maximum amplitude values on only one side of the weld as compared to the other defects, whereas CCr, Po and SI showed almost comparable amplitude distributions on both sides.

Analysis of *BPL* indicated that CCr, Po, SI and LOF presented the highest amplitude distribution in the middle of the weld body at *BPL*, approximately 34 mm. Evaluation of the thickness using Equations 3 and 4 revealed that the actual depth is approximately 3 mm from the surface. On the other hand, LOP and RCr exhibited the highest amplitude distribution at the root position of the weld at *BPL*, approximately 20 to 23 mm, with the actual depth at approximately 10 to 12 mm.


(i) Side A
(ii) Side B
(a) CCr
(b) Po
(c) SI
(d) LOF
Amplitude (%)
BPL (mm)
Datasets
Amax
20.02
27.18
34.35
41.52
48.65
48.56

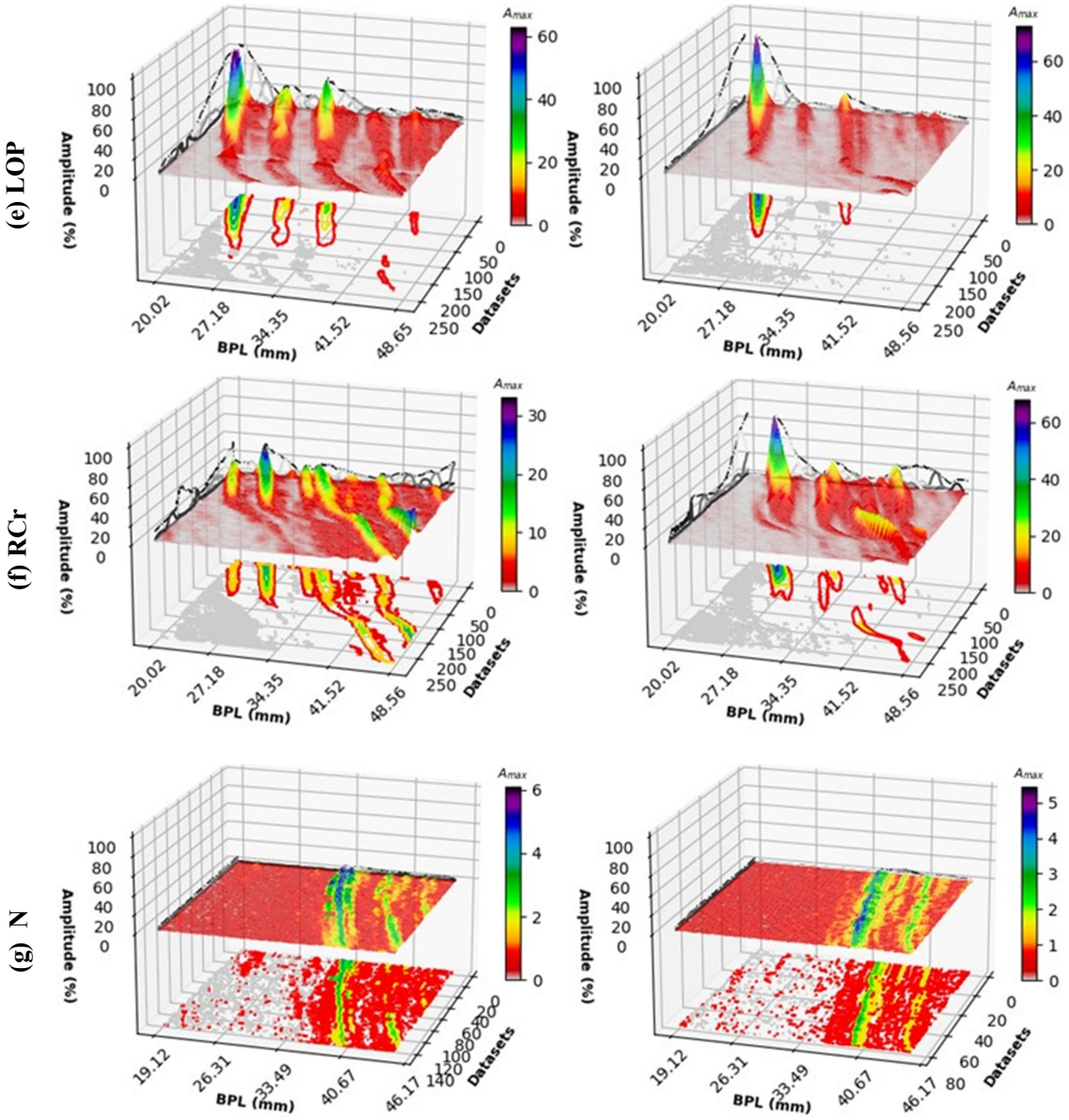


Figure 7: 3D views and B-Scan contours for all tested weld specimens at sides A and B: (a) centre crack (CCr); (b) cluster porosity (Po); (c) slag inclusion (SI); (d) lack of fusion (LOF); (e) lack of penetration (LOP); (f) root crack (RCr); and (g) normal condition (N).

3.1 Features Extraction Analysis

The mean values of R_t are shown in Figure 8(a). Each column represents the mean value of the FWPE comparison as well as the standard error. The rise time is the time interval between the initial threshold crossings, which was set between 10% and 90% of the maximum amplitude. Po has a higher value when compared to the other defects and N, while SI and RCr showed almost the same mean value on both sides.

The mean values of F_t are shown in Figure 8(b). F_t is the time interval between the first threshold crossing, set to 90%, until the signal falls to 10% of its maximum amplitude. Excluding Po and CCr, the F_t of each weld defect and N was generally the lowest. Figure 8(c) illustrates a similar trend by showing the mean D_t value, which is the sum of the rise and fall time for all samples tested. Overall,

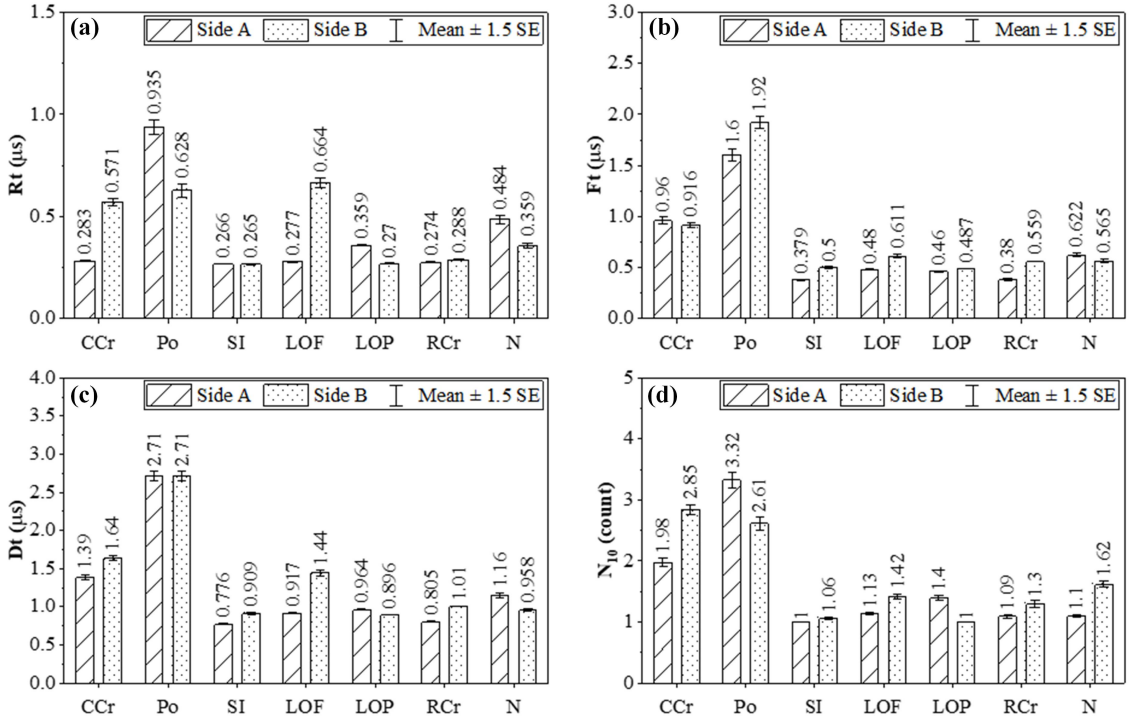
the characterisation of the lowest mean values revealed a narrow sharp pattern, whereas the characterisation of the highest mean values presented a broad pattern for the Po and CCr defects.

The number of peaks recorded for all welding test specimens is shown in Figure 8(d). Po and CCr have a higher peak number than other defects and N. This is due to the fact that the Po specimens used are composed of clusters, whereas Cr has a jagged surface. Both defects will produce multiple reflections on the signal. The number of peaks recorded in Figure 8(e) is the number of peaks calculated at the same level and exceed the 90% amplitude threshold. SI showed a similar number of peaks on both sides A and B, while no significant trend was present in other weld defects and N.

The mean values of A_{max} are shown in Figure 8(f). It is clear that Po and N have the lowest amplitude values on both sides A and B. The mean value of Po is influenced by small pore size, which weakens the reflection signal, while N is reflected from the end corner of the weld cap slightly in contact with the surface. LOF exhibited the highest mean amplitude value on side A when compared to the other defects. This is due to the defect orientation that is perpendicular to the wave reflection, causing the highest reflection of ultrasonic energy. In general, the mean value of the maximum peak amplitude is influenced by the reflection angle, the defect surface, the shape and size of the defect tested.

The mean value of depth is shown in Figure 8(g). Overall, the defect depth in the range of 3.58 to 5.39 mm corresponds to the defect position in the weld body. LOP and RCr ranged from 9.01 to 10.6 mm, close to the thickness of the parent metal that measured at 10 mm and located at the weld root. The depth parameters can be useful in determining the location of a particular type of defect.

Figure 8(h) represents the mean skew values of all tested samples. The mean value of skewness for each weld is the lowest, except for Po, CCr and N at both sides. Overall, all types of defects and controls exhibited a positive mean skew value, i.e., a value greater than zero and heavier on the right tail of distribution.



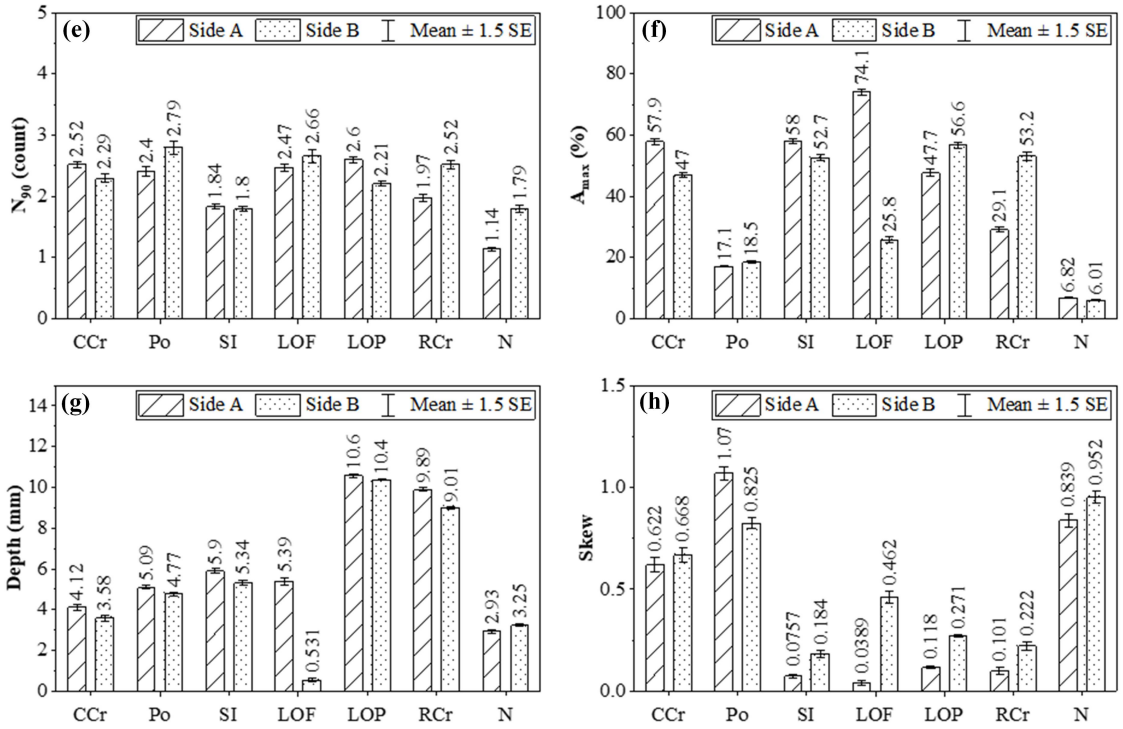


Figure 8: Features extraction analysis for all tested weld specimens at sides A and B: (a) rise time (R_t); (b) fall time (F_t); (c) duration time (D_t); (d) number of peak counts≥ 10% threshold (N_{10}); (e) number of peak counts ≥ 90% threshold (N_{90}); (f) maximum amplitude (A_{max}); (f) depth; and (g) skew.

3.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) was performed to determine the correlation between feature extraction parameters and weld defect types, as shown in Figure 9. Both axes describe a cumulative variance of 80.39% and the symbols represent the types of defects on both sides of the weld. All defects and N are in almost identical clusters, except LOF and RCr, which are isolated between the two sides. A similar trend was previously shown in Figures 7(d) and 7(f). The first component describes 59.34% of variation and separates the extraction of the highest value traits for CCr and Po from the other defects, while N is at the lowest position for all the defects as the reference data. Po shows the extraction of the highest N_{10}, N_{90}, R_t, F_t, D_t values and skew, followed by CCr and N. The second component describes the 12.17% variance and separates the extraction of the highest value features for LOF-A, CCr and LOP from the other defects. The A_{max} extraction exhibited the highest values for LOF-A (on side A). The same trend was previously shown in Figure 8(f). RCr-B (on side B) is with the LOP cluster, indicating that the relationship has similar characteristics, where the signal obtained is almost the same due to the position of the defect orientation.

Based on the direction of feature extraction, PCA differentiates defects into four different quadrant groups. The defects with similar features are grouped in the same quadrant. The top right quadrant shows that the extraction direction of features N_{10}, N_{90}, F_t and D_t is the appropriate characterisation for Po and CCr, while the top left quadrant in the direction of A_{max} is the appropriate characterisation for LOF-A and LOP. The bottom left quadrant reveals that the extraction direction of features of depth is the appropriate characterisation for SI, LOP and RCr-B, while N located in the lowest quadrant with no defects showed suitable characterisation for the comparison of all types of defects.

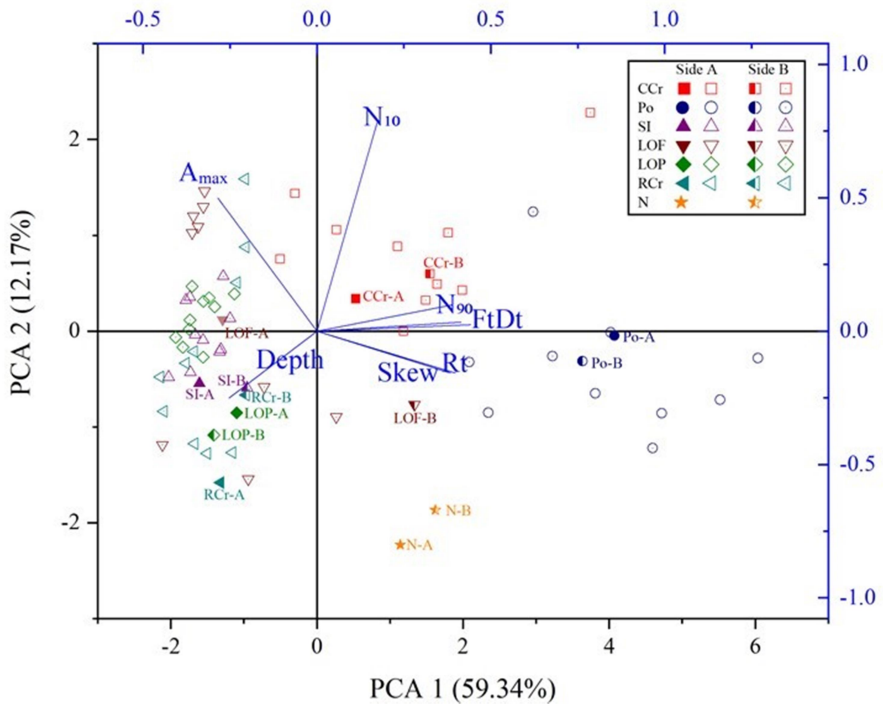


Figure 9: Principal component analysis (PCA) for all tested weld specimens at sides A and B

4. CONCLUSION

In this paper, various types of welding defects were evaluated using the UT technique on FWPE signals. The results of the investigation revealed that the characteristic extraction parameters were influenced by the presence of defects. The findings from these experiments proved the capability of the UT method using the transverse scan technique to detect the presence of different defects in V-butt welded specimens with the following conclusions:

a) UT offers the potential to detect and identify different weld defects and evaluates the entire welded structure on both sides A and B.

b) From the extraction of features tested in this study, the time-domain (R_t, F_t and D_t), number of peaks (N_{10} and N_{90}), depth, skew and maximum amplitude were found to be the main parameters in detecting the presence of weld defects.

c) Analysing the number of peaks was found to be a good detection indicator for detecting Po and CCr defects. Depth characterisation can further distinguish LOP and RCr from other defects.

d) Scanning on both sides (A and B) of the weld as a recommended standard is essential to distinguish the types of LOF and RCr defects since they are influenced by defect orientation.

e) Although the SI defect is quite difficult to identify, its symmetric pattern shows low skew values. The comparison of mean values for $N_{10,}$ N_{90} and A_{max} revealed identical distribution values on both sides A and B.

ACKNOWLEDGEMENT

The authors would like to express our appreciation to the Ministry of Science, Technology & Innovation (MOSTI) and the Malaysian Nuclear Agency for their support for this project.

REFERENCES

Birks, A.S. (1991). Waveform and data analysis technique. In Mclntire P. (Ed.), *Nondestructive Testing Handbook, 2nd Ed., Vol. 7*. American Society for Nondestructive Testing (ASNT), Columbus, Ohio, US, pp. 147–149.

Cai, Z., Jin, Z., Zhu, L., Li, Y., Lei, Y. & Gao, Z. (2020). Optimizing the calibration error of refraction angles in ultrasonic angle beam testing. *Sens.*, **20**: 1-20.

Consonni, M., Howse, D., Wee, C. F. & Schneider, C. (2014). Production of joints welded with realistic defects. *Weld. Int.*, **28**: 535-546.

Droubi, M.G., Faisal, N.H., Orr, F., Steel, J.A. & El-Shaib, M. (2017). Acoustic emission method for defect detection and identification in carbon steel welded joints. *J. Constr. Steel Res.*, **134**: 28-37.

Drury, J. C. (2004). *Ultrasonic Flaw Detection for Technicians,3rd Ed.*, Silverwing Limited, United Kingdom.

Gang, T., Takahashi, Y. & Wu, L. (2002). Intelligent pattern recognition and diagnosis of ultrasonic inspection of welding defects based on neural network and information fusion. *Sci. Tech. Weld. Join.*, **7**: 314-320.

Hernandez, A., Altuzarra, O., Petuya, V., Pinto, C. & Amezua, E. (2018). A robot for non-destructive testing weld inspection of offshore mooring chains. *Int. J. Adv. Robot. Syst.*, **15**: 1-12.

Hoseini, M. R., Zuo, M.J. & Wang, X. (2013). Using ultrasonic pulse-echo B-scan signals for estimation of time of flight. *Meas.: J. Int. Meas. Confed.*, **46**: 3593-3599.

Hu, H., Peng, G., Wang, X. & Zhou, Z. (2018). Weld defect classification using 1-D LBP feature extraction of ultrasonic signals. *Nondestruct. Test. Eval.*, **33**: 92-108.

IAEA (International Atomic Energy Agency) (2018). *Training Guidelines in Non-Destructive Testing Techniques: Manual for Ultrasonic Testing at Level 2 (IAEA-TCS-67)*. International Atomic Energy Agency (IAEA), Vienna.

ISO (International Organization for Standardization) (2012). *ISO 16827:2012(E): Non-destructive testing - Ultrasonic testing - Characterization and sizing of discontinuities.* International Organization for Standardization (ISO), Geneva.

Kim, Y. L., Cho, S. & Park, I.K. (2021). Analysis of flaw detection sensitivity of phased array ultrasonics in austenitic steel welds according to inspection conditions. *Sens.*, **21**:, 1-16.

Lin, Z., Yingjie, Z., Bochao, D., Bo, C., & Yangfan, L. (2019). Welding defect detection based on local image enhancement. *IET Img. Process.*, **13**:, 2647-2658.

Lingvall, F., & Stepinski, T. (1999). *Ultrasonic Characterization of Defects: Part 4. Study of Realistic Flaws in Welded Carbon Steels*. SKI Report 99:25, Department of Signals and Systems, Uppsala University, Sweden.

Ma, M., Cao, H., Jiang, M., Sun, L., Zhang, L., Zhang, F., Sui, Q., Tian, A., Liang, J. & Jia, L. (2020). High precision detection method for delamination defects in carbon fiber composite laminates based on ultrasonic technique and signal correlation algorithm. *Mater.*, **13**:, 1-21.

Qidwai, U. & Bettayeb, M. (2009). Fuzzy time-frequency defect classifier for NDT applications. *IEEE Int. Symp. Signal Process. Inf.*, **1**: 303-309.

Sambath, S., Nagaraj, P. & Selvakumar, N. (2011). Automatic defect classification in ultrasonic NDT using artificial intelligence. *J. Nondestruct. Eval.*, **30**: 20-28.

Sani, S., Saad, M.H., Jamaludin, N., Salim Muhammad, N., Mustapha, I., Mansor, I. & Tengku Amran, T.S. (2019). Design and development of real-time welded metal defect classification automated ultrasonic testing system. *IOP Conf. Ser. Mater. Sci. Eng.*, **555**: 1-9.

Seyedtabaii, S. (2012). Performance evaluation of neural network based pulse-echo weld defect classifiers. *Measurement Science Review*, **12**: 168-174.

Shahriari, D., Zolfaghari, A., Jahazi, M. & Bocher, P. (2013). Development of an Expert System To Characterize Weld Defects. *Proc. ASME 2013 Press. Vsl. Pip. Conf.*, 14-18 July 2013.

Singh, A. K., Singh, G.P. & Sudheera, K. (2015). Weld flaw characterization through mathematical modeling from ultrasonic signal. *IEEE Int. Conf. Commun. Signal Process.*, 2-4 April 2015.

Singh, R. (2012). *Applied Welding Engineering: Processes, Codes, and Standards*. Butterworth-Heinemann, United Kingdom.

Sonaspection. (2017). *Experts in Manufacturing Flawed Specimens and Mock-Ups*. Available online

at: https://sonaspection.com/images/Sonaspection_Brochure_2017_WEB.pdf. (Last access date: 24 April 2022).
Sudhamayee, K. (2019). Pipeline monitoring using ultrasonic sensors. *Int. J. Eng. Adv. Technol.*, **8**: 1937-1940.
Zolfaghari, A., Shahriari, D. & Masoumi, F. (2013). Characterization of welded components flaws using an ultrasonic expert system based on static patterns. *Key Eng. Mater.* **531-532**: 531-532.
Zwillinger, D., & Kokoska, S. (2000). *Standard Probability and Statistics Tables and Formulae*. Chapman & Hall, New York.

ACOUSTIC EMISSION WITH HIGH SENSITIVITY FOR VALVE LEAK DETECTION

Rokhmadi[1,2], Nor Salim Muhammad[1*], Ridzuan Ahmad[3], Rozaimie Daud[4], Calvin Khoo[4], Abd Rahman Dullah[1] & Ruztamreen Jenal[1]

[1]Faculty of Mechanical Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Malaysia
[2]Research Center for Nuclear Reactor Technology, National Research and Innovation Agency (BRIN), Indonesia
[3] Industrial Technology Division, Malaysian Nuclear Agency, Malaysia
[4] Vision One Sdn. Bhd., Malaysia

[*]Email: norsalim@utem.edu.my

ABSTRACT

Acoustic emission (AE) sensors have been used to record AE signals in a straight pipe with a valve. In addition, multiple AE sensors on the valve body were used to increase the sensitivity of the leak detection. Simultaneous data collections from the sensors were performed to observe the changes of AE signals in the pipe sections and the valve. The recorded root mean square (RMS) and peak amplitude data that were used in the data analysis indicated an identical pattern of AE results in the leaked valve. However, the use of the RMS signals on the gas valve indicates more significant leakage signals from the sensors that were used in the test and one of the valve sensors indicated the highest RMS reading, which may be due to the internal leak that was close to the position of the sensor.

Keywords: *Non-destructive testing; acoustic emission; pipeline; valve inspection; internal valve leak.*

1. INTRODUCTION

Recently, the use of acoustic emission (AE) technology has been given attention in many inspections of essential equipment in plants, especially in monitoring of structural defects and leakage valves in electrical power plants and petrochemical industries (Zuluaga-Giraldo *et al.*, 2004; Bakirov *et al.*, 2015; Chacon *et al.*, 2016; Bhuiyan *et al.*, 2018; Ahn *et al.*, 2019; Chiappa *et al.*, 2020). The attention on this technique is due to its ability for defect detection in large structures and its appropriateness in conducting non-intrusive inspection without dismantling the equipment. The technique has also been proven to be helpful in detecting internal leaks in valves (Yan *et al.*, 2015) as well as in identifying cracks (Carlyle, 1989; Bhuiyan *et al.*, 2018; Chiappa *et al.*, 2020) and corrosion activities (Fregonese *et al.*, 2001; Kim *et al.*, 2003;Park *et al.*, 2006; Baran *et al.*, 2012) in fundamental structures such as pipes, storage tanks, and pressure vessels. Based on the similarities found in the structures, it also has feasibility for monitoring of the military facilities such as the marine valves and ballast tanks in the naval ships.

Studies on AE have been extensively conducted by previous researchers in monitoring crack propagation activities. In general, any deformation involving elastic or plastic deformation will produce stress waves that can be used to detect the activities of crack propagation in structures (Lindley *et al.*, 1978; Roberts & Talebzadeh, 2003). These include studies on crack propagation in metal structures (Baran *et al.*, 2012; Mazal *et al.*, 2015; Bhuiyan *et al.*, 2018; Chiappa *et al.*, 2020), cracking signals in concrete structures (Soulioti *et al.*, 2009; Ohno & Ohtsu, 2010; Aggelis, 2011), and corrosion activities (Fregonese *et al.*, 2001; Park *et al.*, 2006; Kasai *et al.*, 2009; 2014; Baran *et al.*, 2012).

There are also studies that utilised AE for inspections of equipment such as gearbox failures on wind turbine and helicopter applications (Elasha *et al.*, 2015; Chacon *et al.*, 2016), detection of internal leakage in closed valves (Yan *et al.*, 2015), estimation of passing rate in valves (Kaewwaewnoi *et al.*, 2010), and monitoring of check valves for nuclear power plants (Lee *et al.*, 2006). These studies found that the detection for internal leaks in valves can be accomplished by listening to the sound of the leak produced when the valve is fully closed. At the same time, the tight valves are predicted to have insignificant level of AE signals.

AE also gives a good prospective in monitoring of the military equipment, such as naval ships and aircrafts. It has been used for monitoring the corrosion on the naval ships, which was focused of the ship hull and the deck (Baran *et al.*, 2012; Emilianowicz *et al.*, 2014). It also a practical method for inspection of crack growth on pressurised structures such as pressure vessels and ballast tanks (Baran *et al.*, 2012; Mazal *et al.*, 2015; Chiappa *et al.*, 2020). At the same time, the technique shows potential in online monitoring of fatigue crack for aircrafts and other critical structures, where the failures can be identified at as early stage, such as the implementation of damage control for warships or aircrafts (Carlyle, 1989; Park *et al.*, 2016).

This study discusses an application of AE for leak detection in a gas valve that is used to purge flammable and harmful gases in a confined space. The leak detection used multiple AE sensors that are fixed around the valve for the acquisition of AE signals on the valve body. Comparisons with the background noise are obtained from the AE sensors that are placed close to the valve. Analysis of high root mean square (RMS) and peak amplitude on the valve body are expected to indicate the presence of internal leakage in the valve.

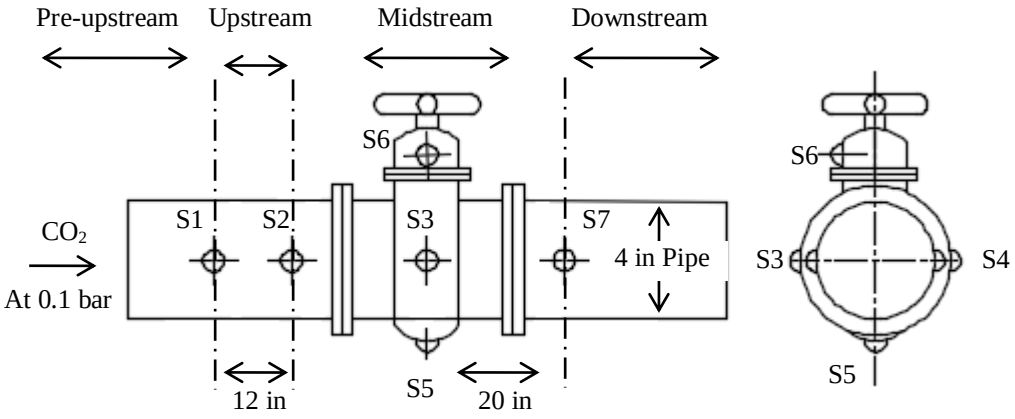
2. MEASUREMENT AND SIGNAL PROCESSING

2.1 Measurement of Acoustic Noise

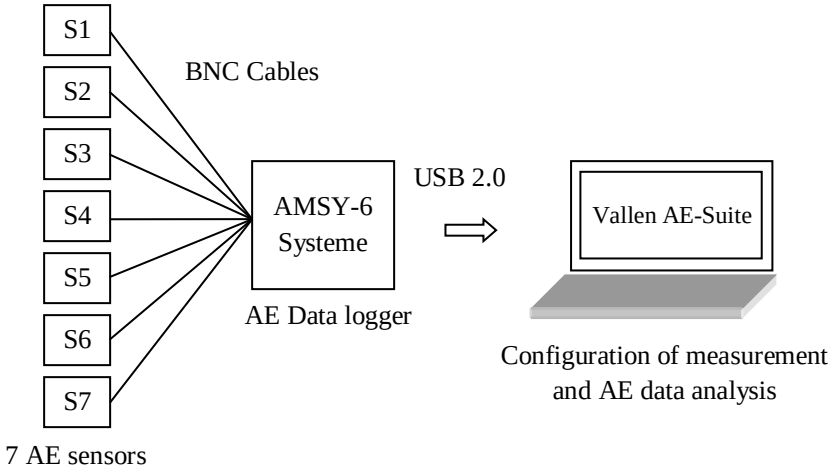
An internal leakage inspection for a gas valve was carried out using the AE technique, as in Figures 1(a) and 1(b). An AMSY-6 system, which is a digital multi-channel AE-measurement system from Vallen, was used to record the generated AE signal in the valve, as shown in Figure 1(c). This equipment was configured at sampling rate of 5 MHz with band pass filter between 95 and 300 kHz. Seven channels were used to acquire the AE data, consisting of RMS and peak amplitude for the valve leak analysis. VS150-RSC piezoelectric sensors of with central frequency of 150 kHz were placed at different locations on the valve and pipe to simultaneously record the AE signals in the open and close conditions of the valve, as shown in Figure 1(a).

2.2 Test Valve and Purging Process

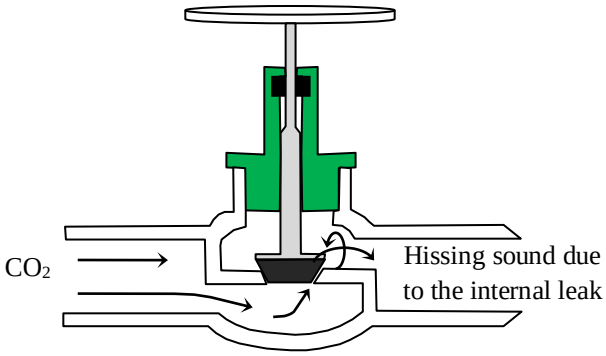
The internal leak test was performed on a globe valve that was used as an isolation valve in Figure 1(c). The valve was used in a 4 in pipe that supplies carbon dioxide into a confined space at pressure of approximately 0.1 bar. Generally, the purging process is performed to reduce the hazard of fires or explosions from the presence of flammable gases, and avoid accidents due to inhalation of harmful or toxic gases in the confined space. Usually, the valve will be closed at the end of the process, where the carbon dioxide or purging gas has purged out the harmful gases and filled the entire confined space. It is a process that needs to be implemented before performing any ventilation process as preparation before any working activity is allowed in the confined space (Finkel, 2000).



(a) Placement of sensors on valve and pipe



(b) Measurement of AE signal and data analysis



(c) The presence of a hissing sound in a leaking valve

Figure 1: Measurement of AE noise for the valve in the pipe.

2.3 Position of Sensors and Data Collection

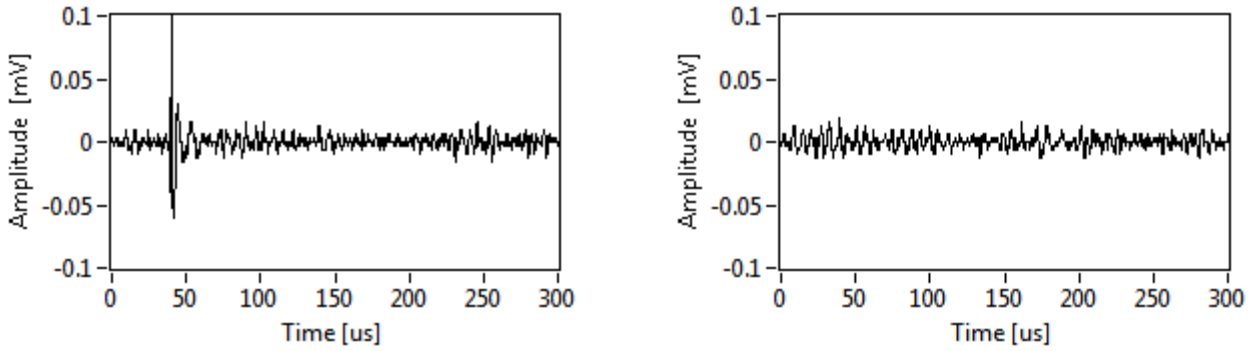
The locations of the sensors are depicted in Figure 1(a), where Sensor S1 was placed on the pipe surface at the pre-upstream location and Sensor S2 at the upstream of the valve. Another two locations were labelled as midstream and downstream locations as in the figure. Four sensors (S3 to S6) were placed at the midstream of the valve to detect the resulting leak signals, while Sensor S7 was placed at the downstream section to collect the background noise that similar to S1 and S2. The sensors at the midstream were placed close to the cross-section of the valve seat and evenly surrounded the valve. All the sensors were fixed using magnetic sensor holders and grease couplant to provide good coupling between the sensors and surface of the materials. The sensors were connected to an AE data logger (AMSY-6 Systeme) that can be operated via a laptop or personal computer for channel configuration and data analysis, as shown in Figure 1(b). The inspection was started by measuring the background noise from a steady-state gas flow at the open valve condition, which was recorded at about 24 dB_{AE}. However, the inspection used a higher measurement threshold at approximately 30 dB_{AE} throughout the inspection for significant measurement of the generated AE signals.

The observation on the valve test started from closing the valve manually, which took place at approximately 100 s. The condition was then left for about 6 min to acquire consistent data in RMS and peak amplitude from the internal leak signals, as shown in Figure 1(c). The valve was then reopened to return the gas flow to its original state. At the end of the inspection, the results from the open to close condition and vice versa were analysed to determine the leakage in the valve.

2.4 Removal of Unwanted Signal

The collected AE data could possibly contain noise from the environment, such as corrosion signals from sacrificial anodes as well as surrounding electromagnetic signals (Fregonese *et al.*, 2001; Emilianowicz *et al.*, 2014; Elasha *et al.*, 2015). Examples of unwanted time waveforms in the AE dataset are shown in Figure 2. One of the signals indicates an amplitude change that is too fast to be identified as an acoustic signal while the other one has very small amplitude.

A proper configuration of the equipment, such as the use of threshold and filter configurations, can prevent the recording the unwanted signals (Carlyle *et al.*, 1989; Elasha *et al.*, 2015; Mousmoulis, 2019). The configuration is necessary in order to avoid excessive use of storage space and limits the recorded data based on the signals associated with valve leakage signatures. Thus, the AE equipment was configured to record wave signals and AE data when it reached a minimum amplitude of 30 dB_{AE} with number of wave cycles of at least five exceeding the threshold.



(a) Possible non-mechanical signal **(b) Weak signal**

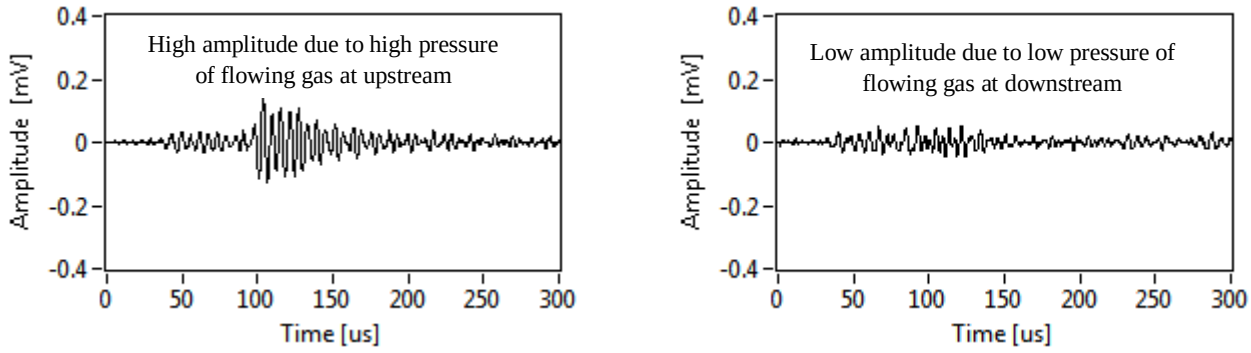
Figure 2: Examples of the unwanted signals in the measurement.

3. RESULTS AND DISCUSSION

3.1 Time Waveform and Content of Frequency Component

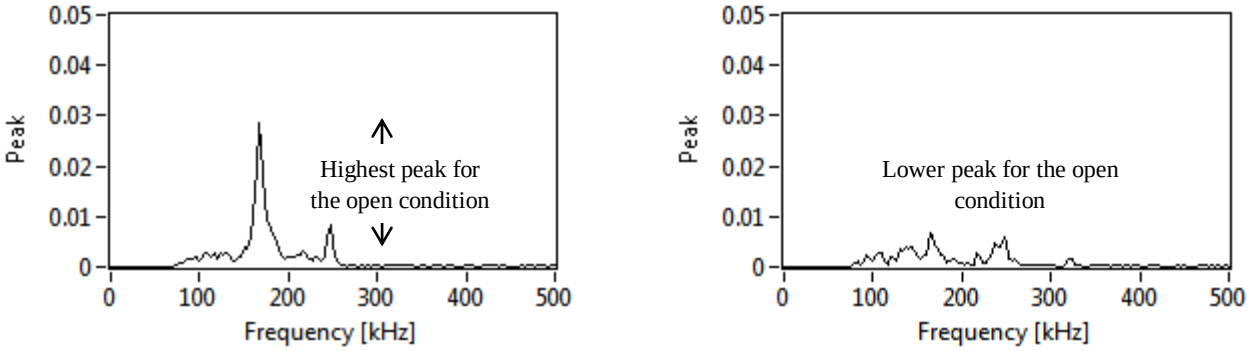
Sample waveforms and their fast Fourier transform (FFT) spectra from the monitored AE signals are shown in Figures 3 to 10. Figures 3, 5, 7, and 9 show the amplitude of the AE signals at the upstream, midstream, and downstream locations when the valve was opened and closed. The waveforms were then used to obtain the FFT spectra to identify their significant frequency content as shown in Figures 4, 6, 8, and 10. The results in Figures 3 and 5 show that when the valve was opened, the AE signal at the upstream location recorded a higher amplitude level as compared to the signals at the midstream and downstream sections of the valve. The difference in the AE signal is probably related to the direction of the gas flow that was flowing from the upstream to the midstream and then reached the downstream. It also indicates that the upstream section has more energy than the downstream as the gas flows through the valve (Lee *et al.*, 2006; Jafari *et al.*, 2014). The computed FFT spectra for the frequency content from 0 to 500 kHz also show dominant peaks at about 160 and 240 kHz.

Figures 7 and 9 show the AE signals when the valve was closed, which indicate higher amplitude of AE signal at the midstream location as compared to the signals at the upstream and downstream sections of the valve. The higher amplitude level at midstream of the valve is probably caused by the hissing sound of the internal leak in the valve. Besides that, the sensors that were placed at the midstream of the valve recorded different amplitude levels, as shown in Figure 9, where the sensor with the highest amplitude is considered as the closest sensing element to the leakage point in the valve. In line with the waveform analysis, FFT spectra can also be used to identify internal leaks (Lee *et al.*, 2006), where valves with higher signals at the midstream than at the upstream and downstream tend to have internal leaks that cannot effectively stop the gas flow.



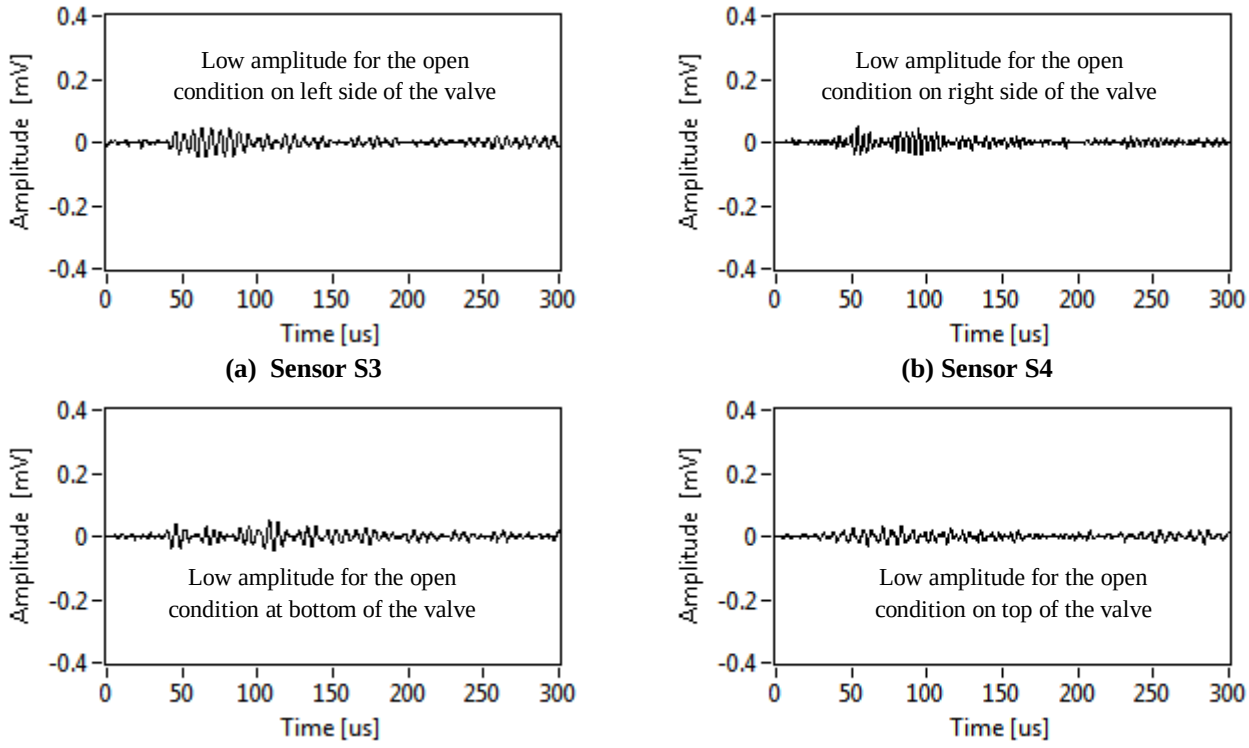
(a) Upstream (Sensor S2) **(b) Downstream (Sensor S7)**

Figure 3: Signals of the open condition at upstream and downstream.



(a) Upstream (Sensor S2) (**b) Downstream (Sensor S7)**

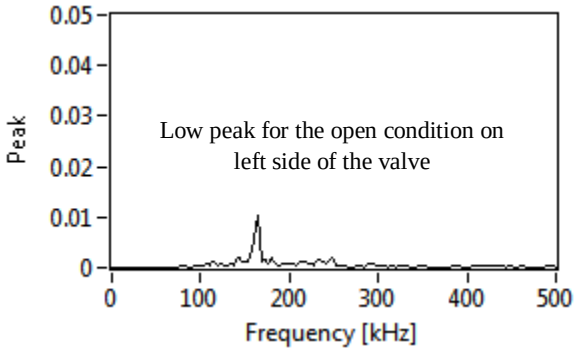
Figure 4: FFT spectra of the open condition for the upstream and downstream signals.



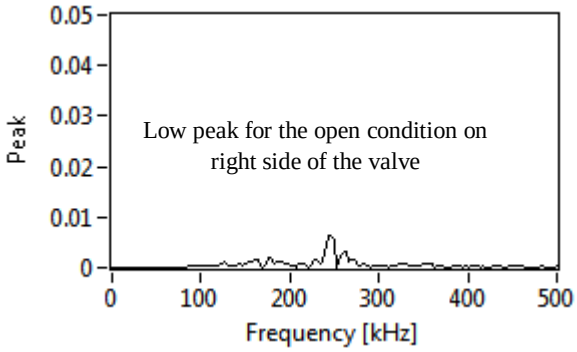
(a) Sensor S3 **(b) Sensor S4**

(c) Sensor S5 **(d) Sensor S6**

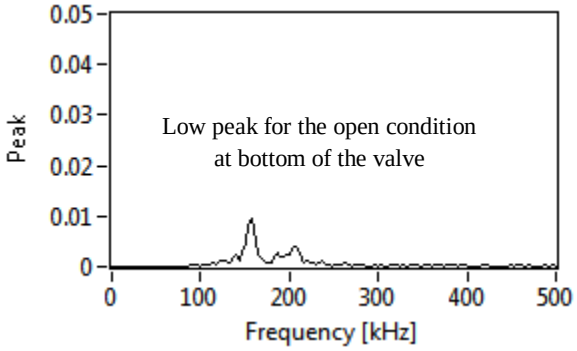
Figure 5: Signals of the open condition at the valve (midstream).



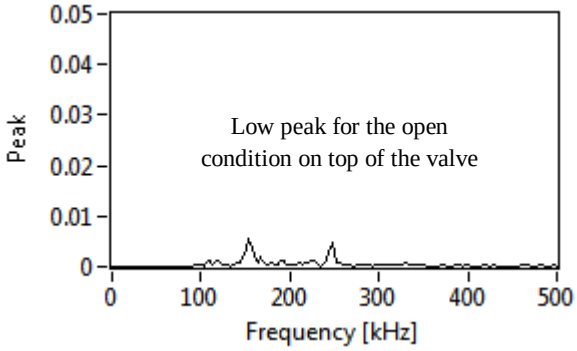
(a) Sensor S3



(b) Ssensor S4

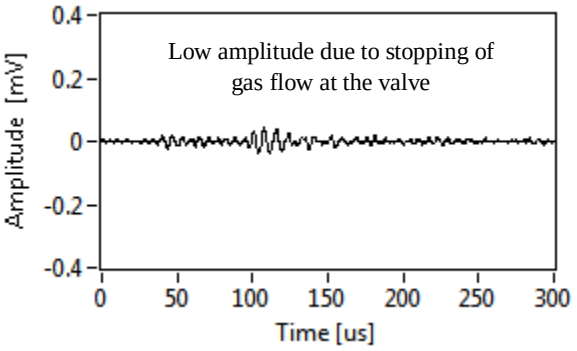


(b) Sensor S5

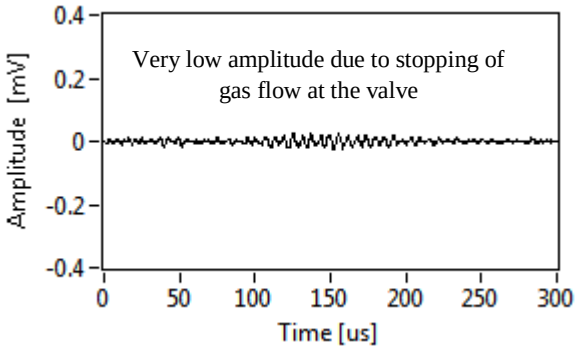


(d) Sensor S6

Figure 6: FFT spectra of the open condition at the valve (midstream).

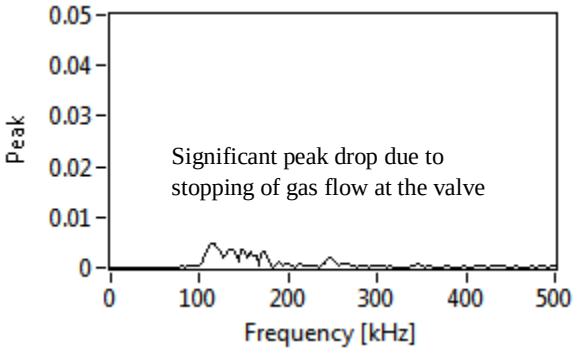


(a) Upstream (Sensor S2)

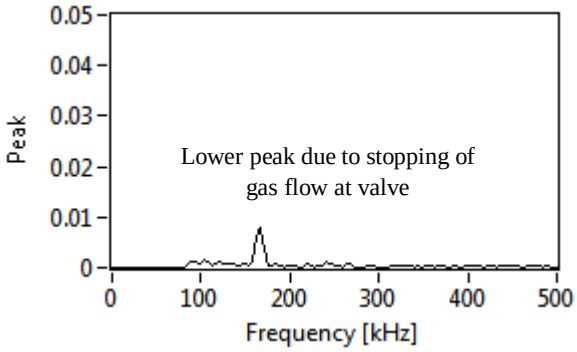


(b) Downstream (Sensor S7)

Figure 7 Signals of the close condition at upstream and downstream.

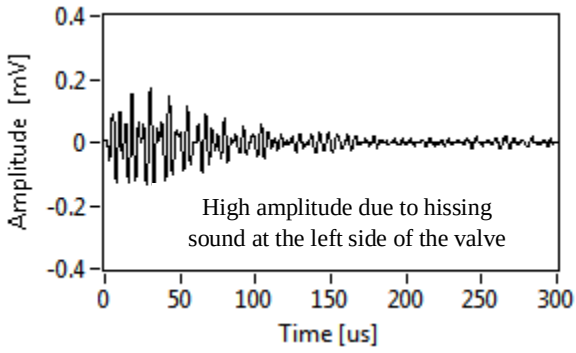


(a) Upstream (Sensor S2)

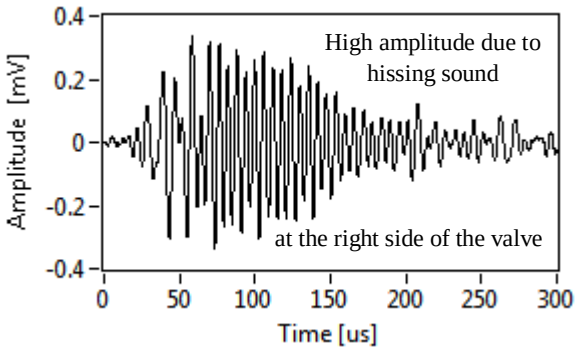


(b) Downstream (Sensor S7)

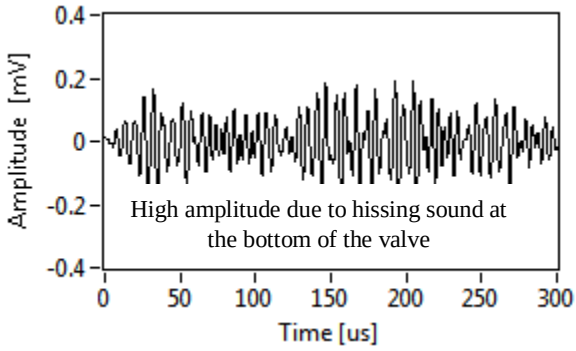
Figure 8: FFT spectra of the close condition at upstream and downstream.



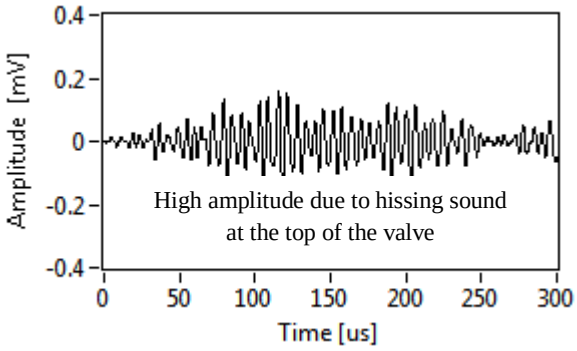
(a) Sensor S3



(b) Sensor S4

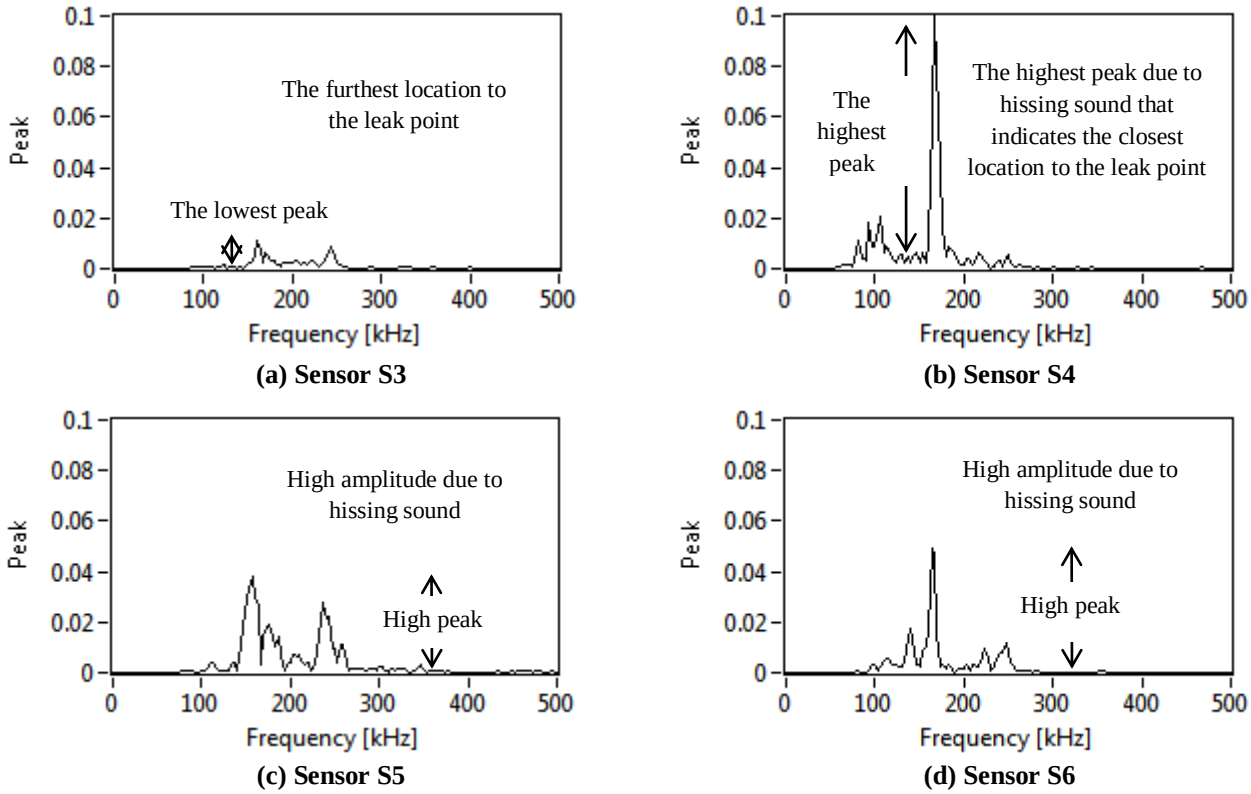


(c) Sensor S5



(d) Sensor S6

Figure 9: Signals of the close condition at the valve (midstream).



(a) Sensor S3 **(b) Sensor S4**

(c) Sensor S5 **(d) Sensor S6**

Figure 10: FFT spectra of the close condition at the valve (midstream).

3.2 Collected RMS and Amplitude Signals

The use of time waveforms in valve monitoring is useful in identifying the condition of a valve as they can be processed into FFT or wavelet transforms, but it requires large data storage space. Recording the waveform features in the form of AE parameters is an alternative in order to identify the conditions of a valve. The use of the AE parameters can also allow longer duration of the inspections and greater number of valve inspections (Ohno & Ohtsu, 2010; Aggelis, 2011).

The recorded RMS values during the inspection are as shown in Figures 11, which were taken at four positions, consisting of pre-upstream, upstream, midstream and downstream of the valve respectively. Monitoring of AE signal at a position before the upstream was performed in order to observe the condition of the AE signal before approaching the valve. Most of the previous studies on the valve leak used test rigs with valves at fully close condition during the AE measurement (Lee *et al.*, 2006; Jafari *et al.*, 2014; Yan *et al.*, 2015). In our study, the AE dataset was recorded from the closing process until the valve was reopened as indicated in the figure. The closing process took place approximately 100 s and the close condition was kept for up to approximately 470 s of the time frame. The closure for almost 6 min was intended to get the reliable signals that represent the conditions of the valve. The reopening process of the valve took approximately 90 s of the time frame and the valve was kept at the open condition after the process.

The RMS noise at the pre-upstream and downstream of the valve recorded almost constant noise during the close condition. On the contrary, the midstream and upstream sections indicated excessive noise signal at the close condition, which is due to hissing signals that were produced by the internal leak (Lee *et al.*, 2006; Jafari *et al.*, 2014; Kaewwaewnoi *et al.*, 2010; Yan *et al.*, 2015).

Figure 12 shows the recorded peak amplitude that was simultaneously recorded to evaluate the valve condition as well as the RMS value. Although the peak amplitude distributions that were obtained from the gas valve are slightly complicated as compared to the RMS parameter at the close condition, the recorded parameters at the upstream and midstream of the valve show higher data density of high amplitude level as compared to the other two sections. The results obtained are almost identical to that obtained through the RMS distribution. However, further analysis needs to be done to identify the valve condition more clearly. For this purpose, comparisons of the two valve conditions were performed to identify the different characteristics using the RMS and peak parameters.

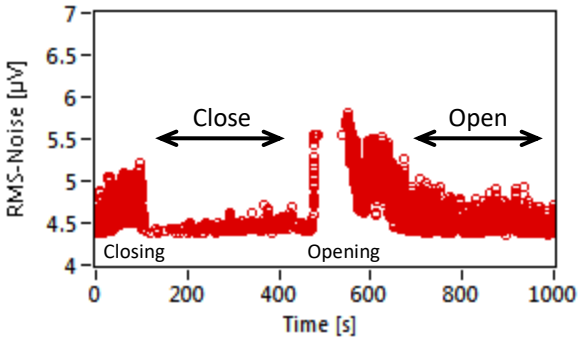
The dataset in the last 300 s during the valve inspection in Figure 11 shows the characteristics of AE noise when the valve was reopened. Meanwhile, the dataset in the last 300 s before the valve reopening process shows the characteristics of AE noise when the valve was in the close condition. The plots of the RMS and peak parameters within 300 s of the time frame for the open and close conditions are shown in Figures 13 to 16.

Figure 13 shows the RMS results for the open condition of the valve. It shows higher density of high RMS levels at the pre-upstream and upstream sections as compared to the middle and downstream sections. The four sensors (S3 to S6) that were placed on the valve at the midstream in Figure 13(c) show lower readings of RMS at approximately 4.5 µV with a relatively low number of AE hits as compared to results recorded in the previous sections that were measured by one sensor at each section. Lower RMS readings and lower AE hit distribution were also observed at the downstream section for the open condition, as shown in the Figure 13(d).

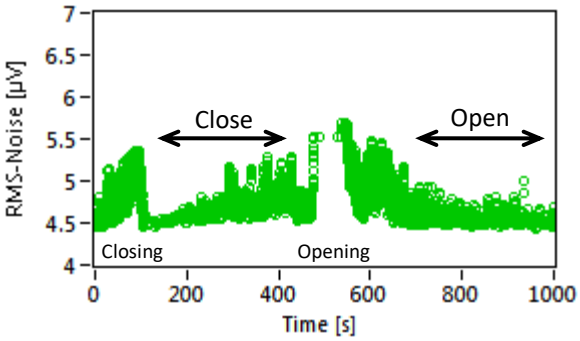
On the other hand, the close condition indicated significant increase in RMS level from two of the midstream sensors that were placed in the middle of the valve. One of the sensors showed an increase of RMS level up to approximately 6.0 µV, as in Figure 14(c). Simultaneously, the upstream sensor also showed a change after 100 s of the valve closure, as shown in Figure 14(b). The significant changes in RMS levels at the pre-upstream and midstream sections for the two conditions can be used to determine the leakage in the valve.

Attenuation of the AE signal from the pre-upstream to downstream of the valve in Figure 13 shows normal behaviour of AE signal attenuation when there is no interruption to the flowing gas in the pipeline. The slightly higher signal at the pre-upstream is likely due to the highest gas pressure location as compared to the other three locations that produced the highest level of AE signal. High distribution of AE hits at higher RMS level at the midstream as compared to the pre-upstream and downstream sections during the close condition indicates a valve leak in the pipe. The gas leakage is likely to be close to the sensor with the highest reading (Lee *et al.*, 2006; Jafari *et al.*, 2014; Kaewwaewnoi *et al.*, 2010; Yan *et al.*, 2015). Therefore, it is not necessary to have high RMS reading from all of the midstream sensors, as a significant reading is sufficient to predict the conditions of the valve.

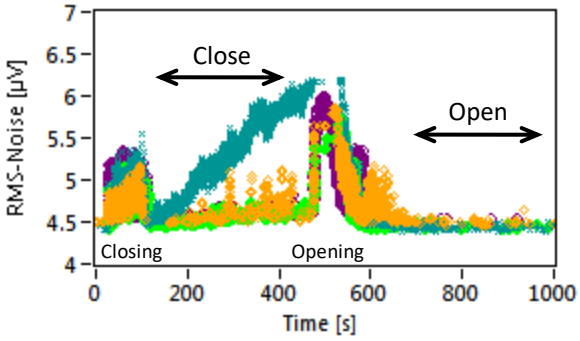
Leakages in the valve can also be detected using the recorded peak amplitude. This is due to the recorded peak amplitude that exhibits identical behaviour to the RMS parameter when the gas is flowing through the valve in the open condition, as shown in Figure 15. The amplitude indicated the highest distribution of AE hits at higher peak amplitude in pre-upstream followed by the sensor at the upstream location, while the hits and signals decrease for the sensors from midstream to downstream.



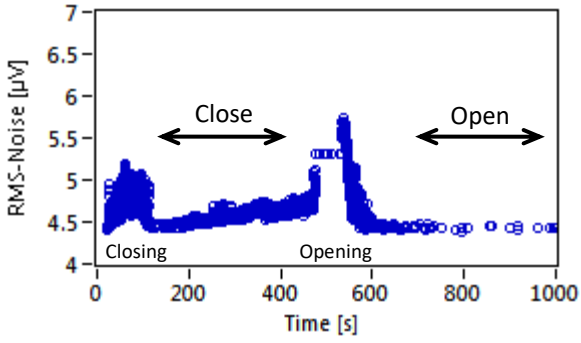
(a) Pre-Upstream (Sensor S1)



(b) Upstream (Sensor S2)

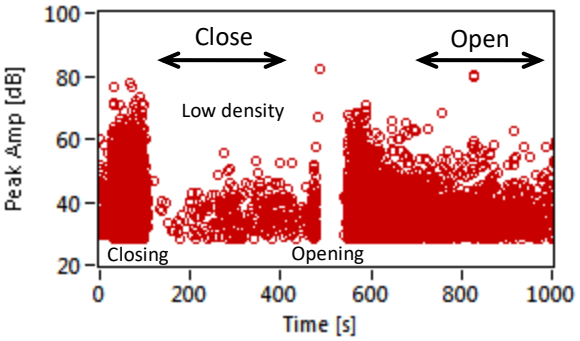


(c) Midstream (Sensors S3 to S6)

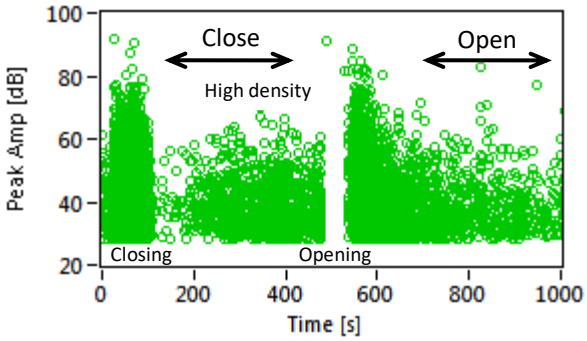


(d) Downstream (Sensor S7)

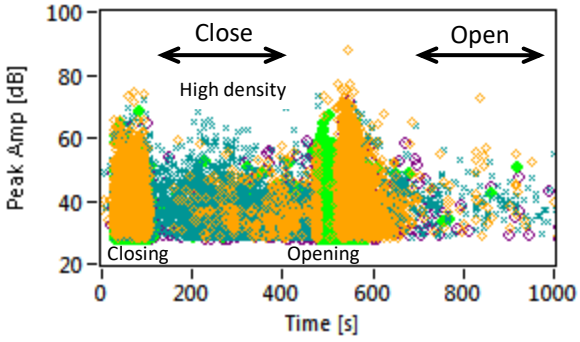
Figure 11: The filtered RMS of the AE signal of the inspected valve.



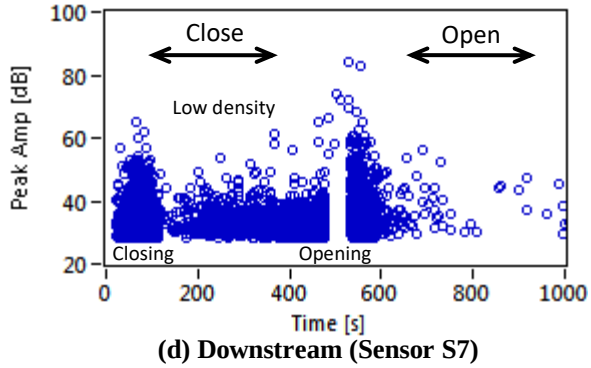
(a) Pre-Upstream (sensor S1)



(b) Upstream (Sensor S2)

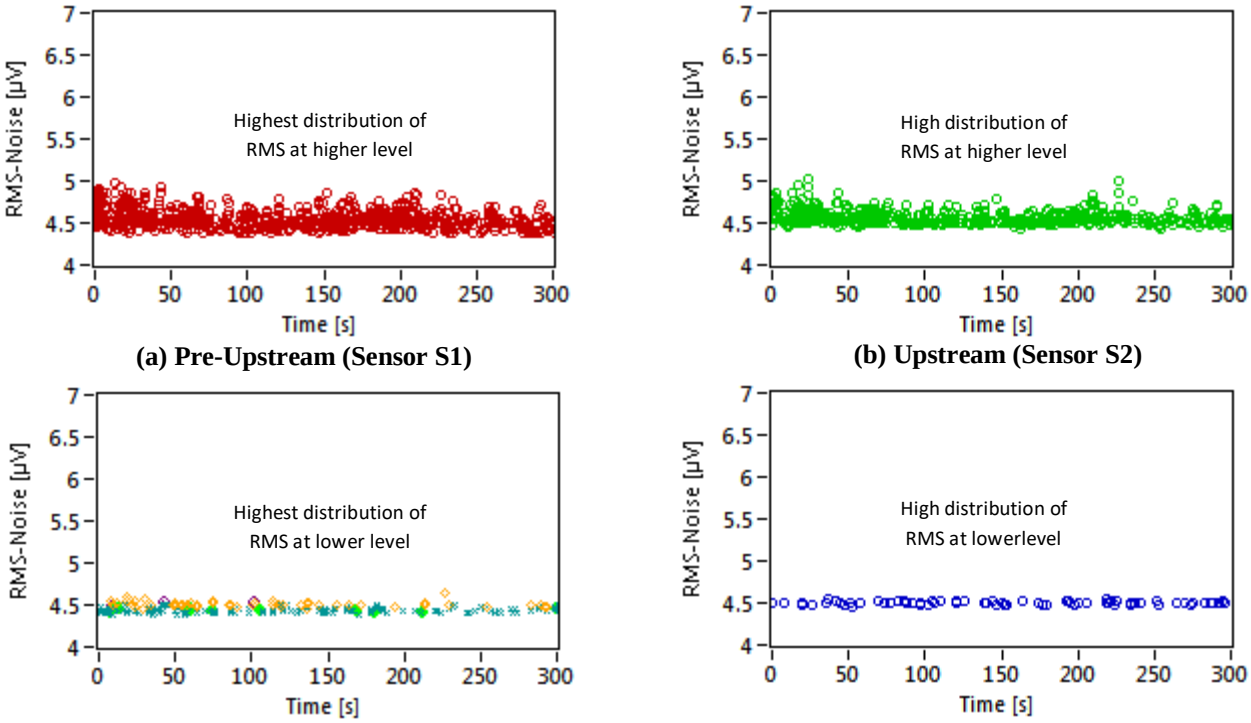


(c) Midstream (Sensors S3 to S6)



(d) Downstream (Sensor S7)

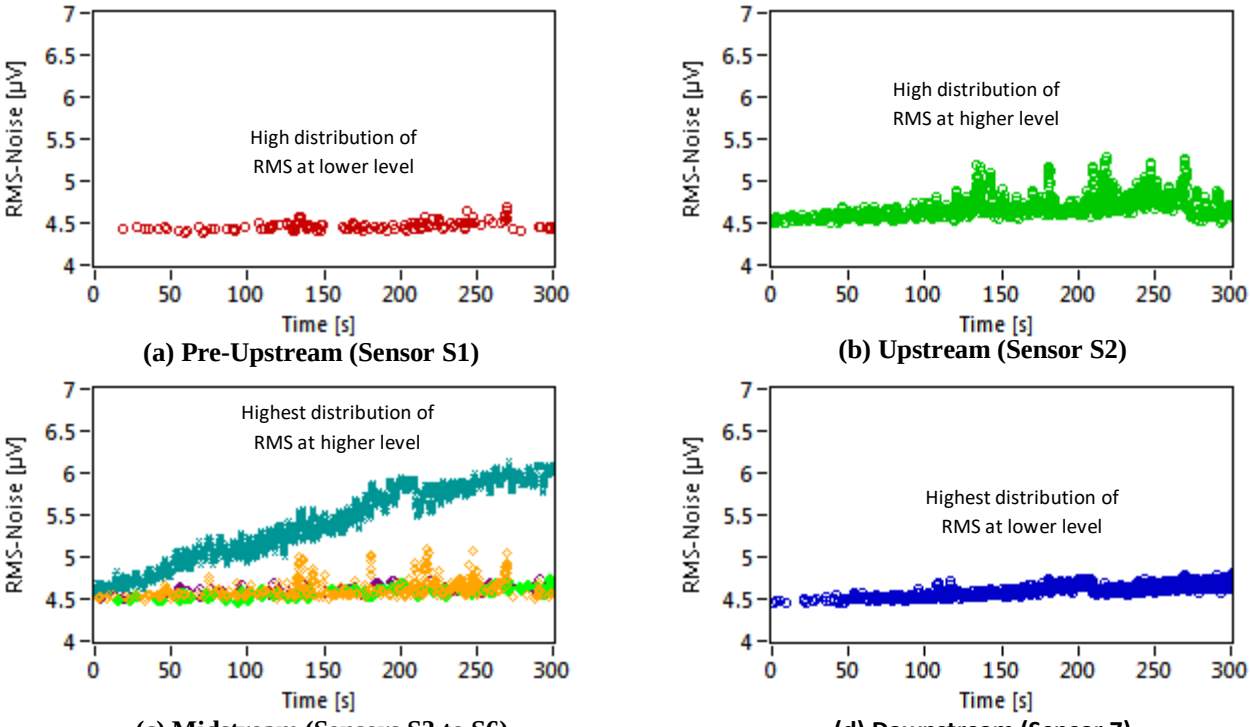
Figure 12: The filtered amplitude of the AE signal of the inspected valve.



(a) Pre-Upstream (Sensor S1) **(b) Upstream (Sensor S2)**

(c) Midstream (Sensors S3 to S6) **(d) Downstream (Sensor 7)**

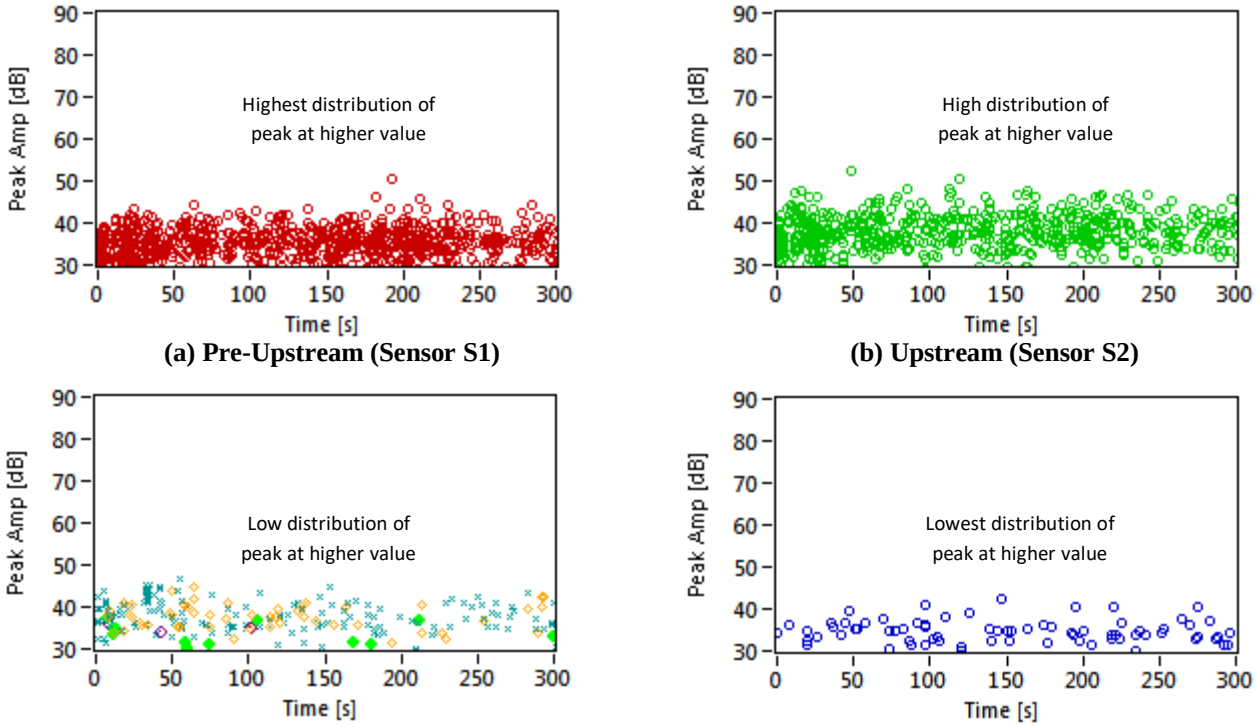
Figure 13: RMS of the AE signal of the inspected valve (open valve).



(a) Pre-Upstream (Sensor S1) **(b) Upstream (Sensor S2)**

(c) Midstream (Sensors S3 to S6) **(d) Downstream (Sensor 7)**

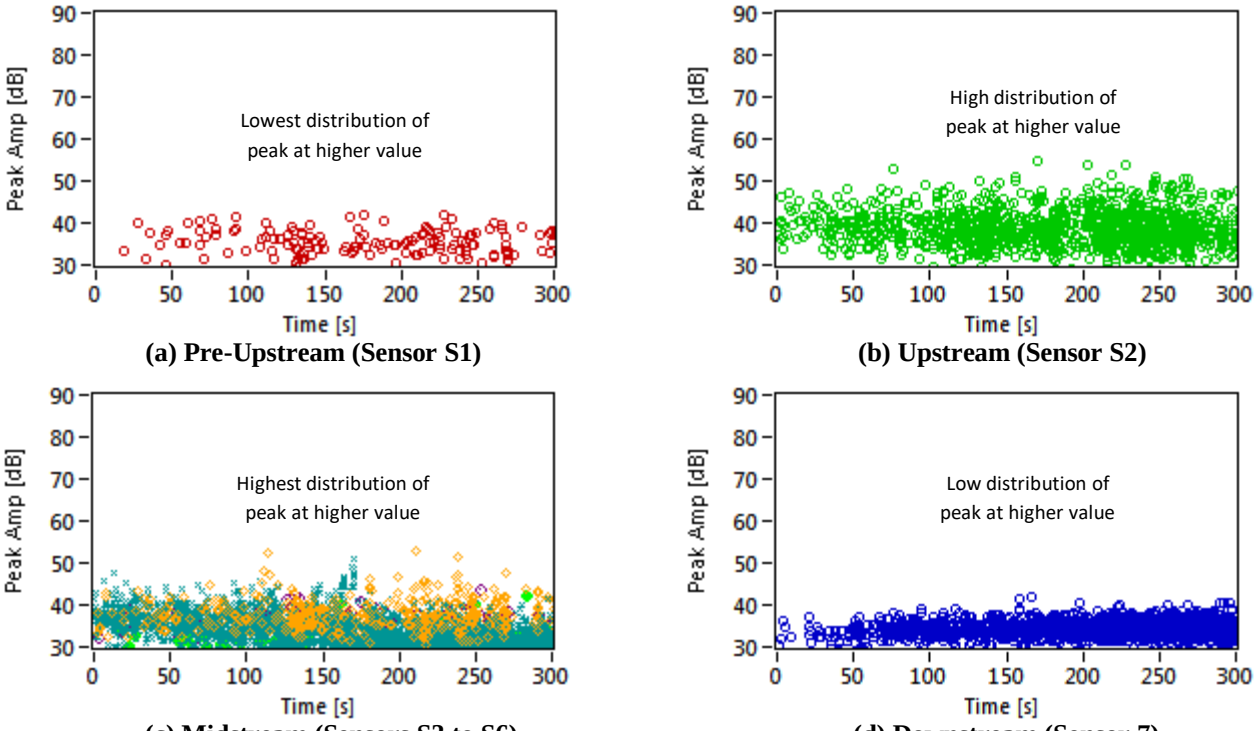
Figure 14: RMS of the AE signal of the inspected valve (closed valve).



(a) Pre-Upstream (Sensor S1) **(b) Upstream (Sensor S2)**

(c) Midstream (Sensors S3 to S6) **(d) Downstream (Sensor 7)**

Figure 15: Amplitude of the AE signals of the inspected valve (open valve).



(a) Pre-Upstream (Sensor S1) **(b) Upstream (Sensor S2)**

(c) Midstream (Sensors S3 to S6) **(d) Downstream (Sensor 7)**

Figure 16: Amplitude of the AE signals of the inspected valve (closed valve).

Once the valve was closed by turning the hand steer completely, the highest density at high amplitude has shifted to the midstream and followed by a sensor mounted at the upstream as shown in Figure 16. At the same time, the pre-upstream sensor recorded the lowest AE hits than the sensors at the other locations.

The use of four sensors at midstream during the open and close conditions recorded different hit distributions, as shown in Figures 15(c) and 16(c) respectively. The midstream sensors at the close condition indicated higher distribution of AE hits at higher amplitude as compared to the open condition. The higher AE hits of high amplitude at midstream as compared to the pre-upstream and downstream sections seem to indicate a leak in the inspected valve.

The findings of this study demonstrate that the use of statistical analysis of peak amplitude data may be helpful in improving the detection of leaks in the gas valve. At the same time, the use of RMS is seen to have higher sensitivity in detecting leaks in the gas valve, whereby the closest sensor to the leak point has the highest level of AE signal.

4. CONCLUSION

A study on internal leakage in a gas valve using AE technique was performed. RMS and peak amplitude parameters were recorded for both open and closed valve conditions. The results indicate that the use of RMS parameter is more sensitive in monitoring leakage signals for a gas valve. Meanwhile, the use of statistical analysis on the peak amplitude data is possible to improve the detection of internal leaks.

ACKNOWLEDGMENT

The authors would like to express their deepest gratitude to Vision One Sdn. Bhd. for giving the opportunity to carry out this work. This work was also supported by Universiti Teknikal Malaysia Melaka (UTeM) and Malaysian Ministry of Higher Education (MOHE) under research grants FRGS/1/2014/TK01/FKM/02/1/F0211, FRGS/2/2013/TK01/UTEM/02/F0171 and ERGS/1/2013/FKM/ TK01/UTEM/02/01/E00014.

REFERENCES

Aggelis, D. (2011). Classification of cracking mode in concrete by acoustic emission parameters. *Mech. Res. Commun.*, **38**: 153-157.

Ahn, B., Kim, J. & Choi, B. (2019). Artificial intelligence-based machine learning considering flow and temperature of the pipeline for leak early detection using acoustic emission. *Eng. Fract. Mech.*, **210**: 381-392.

Bakirov, M.B., Povarov, V.P., Gromov, A.F. & Levchuk, V.I. (2015). Development of a technology for continuous acoustic emission monitoring of in-service damageability of metal in safety-related NPP equipment. *Nucl. Energy Technol.*, **1**: 32-36.

Baran, I., Nowak, M. & Buglacki, H. (2012). Acoustic Emission Monitoring of Structural Elements of a Ship for Detection of Fatigue and Corrosion Damages. *29th Conf. Eur. Group Acoust. Emission.* 12-15 September 2012, Granada, Spain.

Bhuiyan, M. Y., Lin, B. & Giurgiutiu V. (2018). Acoustic emission sensor effect and waveform evolution during fatigue crack growth in thin metallic plate. *J. Intell. Mater. Syst. Struct.*, **29**: 1275-1284.

Carlyle, J. M. (1989). Acoustic emission testing the F-111. *NDT Int.*, **22**: 67-73.

Chacon, J. L. F., Andicoberry, E. A., Kappatos, V., Papaelias, M., Selcuk, C. & Gan, T. H. (2016). An experimental study on the applicability of acoustic emission for wind turbine gearbox health diagnosis. *J. Low Freq. Noise Vibr. Act. Control*, **35**: 64-76.

Chiappa, A., Augugliaro, G., Brutti, C., Brini, F., Groth, C., Mennuti, C., Porziani, S., Quaresima, P., Salvini, P. & Biancolini, M. E. (2020). AE fatigue experiments on tanks test samples with artificial pre-cracking. *Procedia Struct. Integrity*, **25**: 128-135.

Elasha, F., Greaves, M., Mba, D. & Addali, A. (2015). Application of Acoustic Emission in Diagnostic of Bearing Faults within a Helicopter Gearbox. *Procedia Struct. Integrity,* **38**: 128-135.

Emilianowicz, K. (2014). Monitoring of underdeck corrosion by using acoustic emission method. *Polish Marit. Res.*, **21**: 54-61.

Finkel, M. H. (2000). *Guidelines for Hot Work in Confined Spaces: Recommended Practices for Industrial Hygienists and Safety Professionals*, ASSE, USA.

Fregonese, M., Idrissi, H., Mazille, H., Renaud, L. & Cetre, Y. (2001). Initiation and propagation steps in pitting corrosion of austenitic stainless steels: monitoring by acoustic emission. *Corros. Sci.*, **43**: 627-641.

Jafari, S. M., Mehdigholi, H. & Behzad, M. (2014). Valve fault diagnosis in internal combustion engines using acoustic emission and artificial neural network. *Shock Vib.*, **2014**: 1-9.

Jirarungsatian, C. & Prateepasen, A. (2010). Pitting and uniform corrosion source recognition using acoustic emission parameters. *Corros. Sci.*, ***52***: 187-197.

Kaewwaewnoi, W., Prateepasen, A. & Kaewtrakulpong, P. (2010). Investigation of the relationship between internal fluid leakage through a valve and the acoustic emission generated from the leakage. *Measurement,* **43**: 274-282.

Kasai, N., Utatsu, K., Park, S., Kitsukawa, S., & Sekine, K. (2009). Correlation between corrosion rate and AE signal in an acidic environment for mild steel. *Corros. Sci.*, **51**: 1679-1684.

Kim, Y. P., Fregonese, M., Mazille, H., Féron, D. & Santarini, G. (2003). Ability of acoustic emission technique for detection and monitoring of crevice corrosion on 304L austenitic stainless steel. *NDT E. Int.*, **36**: 553-562.

Lee, J. H., Lee, M. R., Kim, J. T., Luk, V. & Jung, Y. H. (2006). A study of the characteristics of the acoustic emission signals for condition monitoring of check valves in nuclear power plants. *Nucl. Eng. Des.*, **236**: 1411-1421.

Lindley, T. C., Palmer, I. G. & Richards, C. E. (1978). Acoustic emission monitoring of fatigue crack growth. *Mater. Sci. Eng.*, **32**: 1-15.

Mazal, P., Vlasic, F. & Koula, V. (2015). Use of acoustic emission method for identification of fatigue micro-cracks creation. *Procedia Eng.*, **133**: 379-388.

Mousmoulis, G., Davies, N. K., Aggidis, G., Anagnostopoulos, I. & Papantonis, D. (2019). Experimental analysis of cavitation in a centrifugal pump using acoustic emission, vibration measurements and flow visualization. *Eur. J. Mech. B/Fluids,* **75**: 300-311.

Ohno, K. & Ohtsu, M. (2010). Crack classification in concrete based on acoustic emission. *Constr. Build. Mater.*, ***24***: 2339-2346.

Park, D. K., Shin, Y. H., Chung, J. H. & Jung, E. S. (2016). Development of damage control training scenarios of naval ships based on simplified vulnerability analysis results. *Int. J. Nav. Archit. Ocean Eng.*, **8**: 386-397.

Park, S., Kitsukawa, S., Katoh, K., Yuyama, S., Maruyama, H. & Sekine, K.(2006). Development of AE monitoring method for corrosion damage of the bottom plate in oil storage tank on the neutral sand under loading. *Mater. Trans.*, **47**: 1240-1246.

Roberts, T. M. & Talebzadeh, M. (2003). Acoustic emission monitoring of fatigue crack propagation. *J. Constr. Steel Res.*, **59**: 695-712.

Soulioti, D., Barkoula, N. M., Paipetis, A., Matikas, T. E., Shiotani, T. & Aggelis, D. G. (2009). Acoustic emission behavior of steel fibre reinforced concrete under bending. *Constr. Build. Mater.*, **23**: 3532-3536.

Yan, J., Heng-hu, Y., Hong, Y., Feng, Z., Zhen, L., Ping, W & Yan, Y. (2015). Nondestructive detection of valves using acoustic emission technique. *Adv. Mater. Sci. Eng.*, **2015**: 1-9.

Zuluaga-Giraldo, C., Mbaa, D. & Smart, M. (2004). Acoustic emission during run-up and run-down of a power generation turbine. *Tribol. Int.*, **37**: 415-422.

REVIEW OF RECENT PHOSPHORUS-BASED FLAME RETARDANTS FOR TEXTILES

Faris Rudi*, Ridwan Yahaya, Noreen Farzuhana, Hidayah Aziz, Haryaty Zahari & Khairunnajwa Md Said

Science and Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*Email: faris.rudi@stride.gov.my

ABSTRACT

Flame retardant textiles have been used extensively in the last few decades in apparels, military, automotive and aerospace industries. Brominated flame retardant (BFR) is a chemical additive that can help to prevent and / or slow down combustion by blocking branching reaction between free radicals from the flame and atmospheric oxygen. However, due to its toxicity towards the environment, scientists have found an eco-friendly alternative to replace BFR, which is phosphorus-based flame retardant. Traditionally, the retardant will exhibit its function in condensed phase at elevated temperature, but with the newly proposed phosphorus-based flame retardant, the combustion inhibition effect can be activated in both condensed and vapour / gas phase. This paper reviews various methodologies used by researchers to synthesise phosphorus-based flame retardants on different textile materials, such as cotton, polyethylene terephthalate (PET), polyamide 6 (PA6), nylon and wool. Characterisations such as limiting oxygen index (LOI) value, vertical flame test and tensile strength for each material will also be further discussed in this paper.

Keywords: *Phosphorus-based flame retardant; textile materials; combustion cycle; flammability tests; toxicity.*

1. INTRODUCTION

The term "textile" originated from Latin adjective *textilis*, which means "woven". It is a flexible material consisting of an interlacing network of fibres commonly used in cloth (Chen *et al*., 2021). Textiles play a significant role in daily life as it used for clothing, household and workplace applications as well as other uses such as parachute, tents and flags (Gordon, 2010). One of the major drawbacks of textiles is that they are mainly made of organic polymers that have low thickness and large specific surface area, as well as flammable. These features lead to ignition of flames in a short period of time with a large number of toxic smokes released (Zheng & Zhang, 2010). A large number of people have died worldwide every year in fire accidents due to easy ignition of textiles (Zhang *et al*., 2019). According to the Fire and Rescue Department of Malaysia (JBPM), based on statistics of investigated structural fires by source of ignition for 2018 revealed that 66% of such incidents comes from residential areas (JBPM, 2019). This statistic is similar with annual UK fire statistics, which clearly demonstrates that most fire incidents occur in residential areas, specifically involving fibres and fabrics such as cloth, upholstering furniture and bedding (Khalifah *et al*., 2016).

Nowadays the issue of flammability of textile materials has raised an area of safety concern, whereby textiles and upholstered furniture are the first thing to be ignited by small fires, such as from cigarettes and candles. These small fires could turns into large fires quite quickly with catastrophic consequences

(Brushlinsky *et al*., 2006). In addition, a large amount of toxic gas released by these textile materials contribute to difficulties for casualties to evacuate from the building, with smoke being the first cause of death during fires (Gann, 2004). Hence, the flame retardant textiles have become an important feature of materials under the field of technical textile. Technical textile can be defined as textile materials manufactured specifically for its performance and technical characteristics as compared to normal textiles that are primarily made for design and ornamental features (Byrne, 2000; Madhav *et al*., 2018). Flame retardant textiles not only focus on household applications, but also have been utilised in many different fields, such as regular apparels, military, automotive and aerospace applications, protective clothing, as well as first responder garments (Erdem *et al*., 2009; Chanchal *et al*., 2020).

In the last decade, a number of studies have been conducted on flame retardant textiles, with numerous approaches being reported. These studies focused on designing and synthesising suitable flame retardants by means of additives that are capable to suppress or delay the appearance of flames and / or slowing down flame-speed rate (flame retardants). Furthermore, the additives are able to delay the ignition or reduce the rate of combustion (Khalifah *et al*., 2016). Brominated flame retardant (BFR) is one of flame retardant chemical additives that are commonly used in a variety of commercial products, such as textiles, automation and electronic applications. It is currently the largest market group share worldwide of flame retardants with approximately 5,000,000 metric tonnes of bromine produced each year because of its superior performance efficiency and low cost (Linda & Daniele, 2004; Ying *et al*., 2019). However, there are concerns about the consumption of BFR as it has been identified as persistent, non-biodegradable, bioacumulative and toxic (Waaijers & Parsons, 2016). Studies of the toxicity of BFR have been conducted by researchers, who found quantifiable levels both in wildlife and humans (de Wit *et al*., 2010; Feiteiro *et al*., 2021).

The need for textile product such as clothing to have efficient flame retardant with high efficiency but low toxicity is crucial (Kimmey *et al*., 2020). Phosphorus-based flame retardant is an alternative to replace BFR additives due to its high efficiency, non-toxicity and environmentally friendly characteristics (Lu & Hamerton, 2002; Armarnath *et al*., 2018). In addition, the efficiency of phosphorus-based flame retardant additives is based on its good interaction with the materials to which it binds to, synergistic reagents and the structure of material itself (Schartel, 2010; Markwart *et al*., 2019). In this context, the objective of this paper is to review recent designs of phosphorus-based flame retardants for different textiles, focusing on development of additives and its impact towards the product.

2. COMBUSTION BEHAVIOURS OF TEXTILES

This section discusses on recent methods of flammability tests for assessing the reaction of textiles to exposure to a flame. In general, combustion of typical textiles triggers in the presence of flames and presence of air or oxygen. Prior to combustion, textile materials go through thermal degradation, where some of the degraded materials turn into combustible volatile products and subsequently, combined with oxygen, they fuel the flame. In a way, if the heat generation surpasses the threshold to sustain the combustion process, the heat can easily be transferred to the textile material, where it can accelerate the degradation process and developed a self-sustaining combustion cycle, as shown in Figure 1 (Chanchal *et al*., 2020).

It is known that the combustion is a complex physical and chemical process where several uncontrolled activities happen simultaneously. Due to this situation, it is difficult to consider which is are the most prominent mechanism for this process. However, gas-phase and condensed-phase actions have often been utilised by researcher as primary mechanisms of combustion in preparing flame retardants treatments for textile materials (Dasari *et al*., 2013). The process that occurs in activities of flame retardant in gas phase demand interference with the combustion process, resulting in reduction of flame propagation, heat

returned to the materials and generation of free radicals. These free radicals will combine with atmospheric oxygen via branching reaction, as shown in Equations 1 and 2, where it will quench the total combustion process (Khalifah *et al*., 2015).

$$H^{\cdot} + O_2 \longrightarrow OH^{\cdot} + O^{\cdot} \quad (1)$$

$$O^{\cdot} + H^{\cdot} \longrightarrow OH^{\cdot} + H^{\cdot} \quad (2)$$

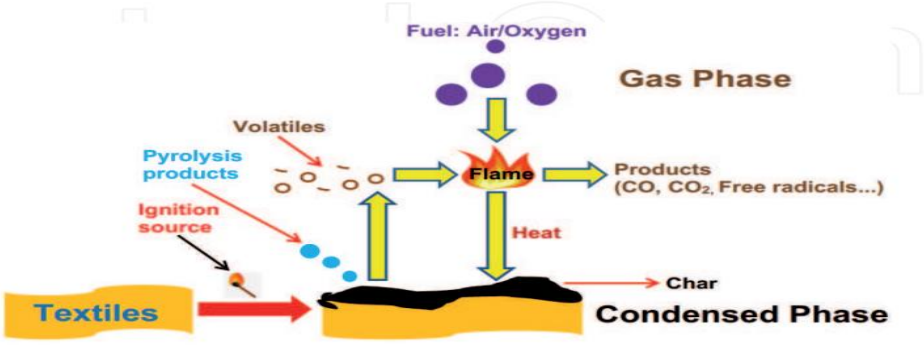


Figure 1: Combustion cycle of a typical textile material (Chanchal *et al*., 2020).

Normally in the gas phase mechanism, the flame retardants function to obstruct this branching reaction so that the combustion can be slowed down or completely stopped. Halogenated compounds such as bromine and chlorine are one of most used flame retardants in earlier times to stop the combustion process in the gas phase mechanism (Laoutid *et al*., 2009). Thus, halogen-containing additives denoted as RX will react with key combustion radicals ($OH^{\cdot}$ and $H^{\cdot}$) to form less reactive halogen atoms. The equation for this reaction is shown in Equations 3 to 6 (Mngomezulu *et al*., 2014). Different halogens provide different effectiveness as a flame retardant, where the order of halogen effectiveness is F < Cl < Br = I (Babushok *et al*., 2014).

$$RX \longrightarrow R^{\cdot} + X^{\cdot} \quad (3)$$

$$X^{\cdot} + R'X \longrightarrow R'^{\cdot} + HX \quad (4)$$

$$HX + H^{\cdot} \longrightarrow H_2 + X^{\cdot} \quad (5)$$

$$HX + {}^{\cdot}OH \longrightarrow H_2O + X^{\cdot} \quad (6)$$

However, as discussed earlier, halogen compounds are toxic and cause a threat to humans and the environment. Thus, phosphorus compound is another option for flame retardants, whereby it releases reactive phosphorus species such as $PO^{\cdot}$, $PO_2^{\cdot}$ and $OHPO^{\cdot}$, where it can react with combustion intermediates in gas phase, as shown in Equations 7 to 11 (Salmeia *et al*., 2015). Frequently, this interaction results in the recombination of $OH^{\cdot}$ and $H^{\cdot}$ radicals, and prevent oxidation from occurring. This condition directly stifles the combustion process.

$$PO^{\cdot} + H^{\cdot} \longrightarrow HPO \quad (7)$$

$$PO^{\cdot} + {}^{\cdot}OH \longrightarrow HPO \quad (8)$$

$$HPO_2^{\cdot} + H^{\cdot} \longrightarrow H_2O + PO \quad (9)$$

$$HPO_2^{\cdot} + H^{\cdot} \longrightarrow H_2 + PO_2 \quad (10)$$

$$HPO_2^{\cdot} + HO^{\cdot} \longrightarrow PO_2 + H_2O \quad (11)$$

Meanwhile, for the condensed phase mechanism, the combustion process leads to structural change in the polymer substrate, whereby carbonaceous / protective char are formed onto material surfaces. Char-forming flame retardants work by preventing fuel release through binding up fuel as nonpyrolysable

carbon (char) and also supplying thermal insulation for underlying materials by the formation of char protective layers (Schartel, 2010). In the end, the presence of these protective char layers provides shielding effects that slow down the degradation rate of polymer, thus directly reducing the intensity of combustion as shown in Figure 1 (Chanchal *et al*., 2020).

3. LABORATORY FLAME TESTING

This section summarises the current methods for assessing the reaction of a textile to exposure to a flame or heat flux. There are several characteristics that need to be measured, such as the rate and extent of flame spread, duration of flame propagation, ease of ignition, ignition resistance, heat release, and smoke (Pettigrew, 1993). Hence, it is impossible to utilise only one instrument for all these parameters. The commonly used flammability tests for polymers is shown in Table 1. Limiting oxygen index (LOI), also known as oxygen index (OI), is one of the most common and popular scientific method used for flame testing. LOI is the minimum concentration (vol%) of oxygen in a mixture of oxygen and nitrogen that is required to sustain flaming combustion of a material. It is used to indicate the relative flammability materials, with the calculation for LOI shown in Equation 12 (Kandola, 2012). LOI values for up to 21 vol% indicates that the textile material burns rapidly, while values in the range of 21 to 25 vol % indicate that it burns slowly. LOI values exceeding 26 vol% indicates that the textile material has some flame retardant features in it (Horrocks *et al*., 1989; Xu *et al*., 2022).

$$LOI = 100 \times [O_2] / ([O_2] + [N_2]) \quad (12)$$

Table 1: Flammability tests for polymers (Pettigrew, 1993).

ASTM Standard	Description	Characteristic Measured
D2863-87	Limited oxygen indices (LOI)	Ease of ignition
UL94	Vertical burn	Ignition resistance
E1354-90	Cone calorimeter	Heat release and smoke

Flame spread (UL94) is a bench-scale testing procedure that measures the rate of flame spread, which is calculated as the ratio of the distance to the time taken of the advancing flame front to reach defined distances marked on the specimen. UL94 also measures ignitability of vertical bulk material exposed to small flames (Patel *et al*., 2012; Khalifah *et al*., 2016). Meanwhile, cone calorimeter was not originally designed for textiles. However, cone calorimetry tests (ISO 5660) have been used as a standard scale model for early flaming (Tata *et al*., 2011). It works based on measurement of decreasing oxygen concentration in the combustion gases of the sample subjected to a given heat flux density (10-100 W/m^2) (Guillaume *et al*., 2014). Cone calorimeter can evaluate several types of parameters, such as time to ignition (TTI), total heat release (THR), heat release rate (pkHRR) and corresponding peak, effective heat of combustion (EHC), mass loss (ML), and mass loss rate (MLR). In addition, cone calorimeter can also be used to assess smoke generation (ISO 5660, Part 2), where it determines carbon monoxide and carbon dioxide concentration, together with the smoke density (specific extinction area-SEA, total smoke production-TSP).

4. PHOSPHORUS-BASED FLAME RETARDANT FOR TEXTILE APPLICATIONS

Phosphorus-based flame retardants are versatile in their flame retardant action as they often exhibit both condensed phase and gas phase activities (Granzow, 1978; Nazir & Gaan, 2020). The mode of action of phosphorus-based flame retardants in simplified scheme is shown in Figure 2. A flame retardant additive is normally active in condensed phase at elevated temperature, where some of chemical reaction will occur depending on the substrate and their chemistry with phosphorus compound. The main chemical reactions that takes place in condensed activity are dehydration, hydrolysis, chain scission or de-polymerisation. At high temperatures, the degradation of phosphorus compounds occurs, which causes changes in the decomposition pathway of the polymer. This also leads to possible formation of carbonaceous protective char residue onto the surfaces of decomposing polymers, and hence, further oxidation can be prevented (Gaan & San, 2009; Chanchal *et al*., 2020).

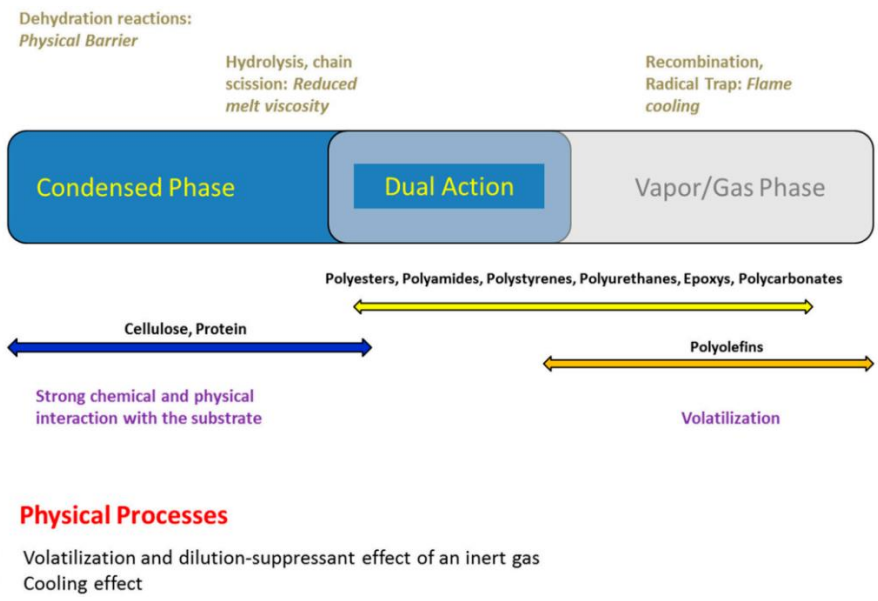


Figure 2: Mode of action of phosphorus-based flame retardants (Khalifah *et al*., 2016).

In other circumstances, the phosphorus compound and some of their decomposed products ideally volatilise during heating and will eventually release reactive phosphorus species. These species then interact with the combustion intermediates coming from the polymer substrates in the gas phase and behave as inhibitors to slow down the combustion rate. Generally, this interaction leads to recombination of the H and OH radicals to prevent further oxidation. Hence, the condensed phase and gas phase activities of phosphorus compounds directly depend on their structure and polymer substrate. For instance, for natural polymers such as cellulose and wool, the phosphorus compounds primarily reveal condensed phase activity, where it starts with dehydration of the polymer, resulting in formation of char onto surfaces of substrates. Meanwhile for synthetic polymers containing oxygen and nitrogen atoms in their structure (for example, polyamides), catalytic hydrolysis of the ester or amide group by the phosphorus acid stimulate enhanced melt dripping and fast shrinkage from flame (Khalifah *et al*., 2016).

Table 2 summarises the recent findings of phosphorus-based flame retardants towards different types of textile materials. The findings will demonstrate whether this type of flame retardant has significant impact to flame retardant properties.

Table 2: Main results achieved by treating textile materials with phosphorus-based flame retardants

Phosphorus-Based Flame Retardants	Textile Material	Highlight	Reference
Phosphoramidate siloxane polymer (PDSTP)	Cotton	• PSDTP was synthesised using sol-gel technology and flame-retardant cotton fabrics were prepared with a multistep coating process. • PSDTP shows good condensed phase and gas phase activities. • Air permeability, whiteness and tensile strength of cotton fabric after coating with PSDTP shows slight decrease in performance. • After 30 washing cycles, the sample did not pass the vertical flammability test. However, it still had good char-forming properties.	Xu *et al.* (2020)
Polythiourea-phosphoric acid	Cotton	• Polythiourea-phosphoric acid contains high amount of phosphorus, nitrogen and sulphur. • 30% concentration of this flame retardant gives excellent flame retardancy with LOI of 48.2%. • The flame retardant has little effect on the mechanical properties of cotton.	Wan *et al.* (2020)
Hyperbranched poly phosphate ammonium salt (HBPOPN)	Cotton	• Formaldehyde-free and halogen-free HBPOPN flame retardant was synthesised under solvent free condition. • The LOI of HBPOPN treated cotton was 42.0%. The LOI value was 29.3% after 50 times washing. • The tensile strength and bending length of the HBPOPN treated cotton fabric shows sustainable mechanical properties.	Ling & Guo (2020)
Poly N,N dimethylene phosphate aminopropyl siloxane (PDPSI)	PET	• PDPI and ferric oxide (Fe_2O_3) were synthesised for flame retardant properties of PET. • The addition of 2% PDPSI/ Fe_2O_3 (1:2) provided significant increase of LOI value from 21.0% to 27.9%. • The char residue image from Fourier-transform infrared spectroscopy (FTIR) and scanning electron microscope (SEM) showed synergistic effect between PDPI and Fe_2O_3, and hence, improved flame retardancy and smoke properties of PET.	Peng *et al.* (2020)
9,10-dihydro-9-oxa-10-phosphaphenanthrene-10-oxide (DOPO)	PET	• DOPO was used as flame retardant for PET and was compared with establish halogen containing flame retardant (DFR). • DOPO could achieved similar excellent flame retardancy and laundering durability with DFR. • The LOI value for both DOPO and DFR was 32%. • The results from the research proved DOPO as an eco-friendly flame retardant with the same characteristics as DFR.	Fang *et al.* (2021)

Melt polycondensation of caprolactam & DDP	Polyamide 6 (PA6)	• Flame retardant polyamide 6 (FRPA6) was prepared by melt spinning. • FRPA6 containing 5 wt% DDP achieved LOI value of 33.7%. • The tenacity at break of FRPA6 fibres met the requirements of textiles.	Liu *et al.* (2018)
Polyphosphoric acid	Nylon	• A sulfur-containing FR (SFR) was synthesised from polyphosphoric acid, epoxy chloropropane & thiourea. • This SFR and water-soluble isocyanate-terminated (WIT) cross-linker was applied as finishing on nylon fabric for flame retardant properties. • The SFR-WIT-treated fabric showed good flame retardancy with LOI value of 29.4%. • After 10 laundry cycles, the LOI value was 26.8% indicating the flame retardancy properties were still intact.	Chen *et al.* (2016)
Ditrimethylolpropane di-N-hydroxyethyl phosphoramide (DDP)	Nylon, cotton and PET	• DDP was successfully synthesised as flame retardant for various fabric. • The vertical flame test showed better flame retardancy of treated nylon fibres and moderate flame retardancy for cotton and PET.	Jiang *et al.* (2015)
Vinyl phosphonic acid (VPA)	Wool	• Graft co-polymerisation of the VPA onto wool fabric was successfully intercalated. • The wool sample with grafting yield (GF) of 8.1% VPA gave LOI of 35.89%, indicating good flame retardancy.	Mohaddes *et al*. (2021)
2-phosphonobutane-1,2,4-tricarboxylic acid (PBTCA)	Wool	• PBTCA was employed as finishing agent to improve flame retardancy of the wool fabrics by pad-dry-cure technique. • The wool treated with PBTCA showed great LOI value of 44%. • PBTCA provides a two-stage flame retardant mechanism that can enhance the crosslinking between the peptide chain in low temperature range and the char formation ability in higher temperature range.	Wang *et al.* (2021)

5. SYNTHESIS AND PREPARATION OF PHOSPHORUS-BASED FLAME RETARDANT

5.1 Synthesis and Preparation of Phosphorus-Based Flame Retardant for Cotton

Xu *et al*. (2020) synthesised water soluble phosphoramidate siloxane polymer (PDTSP) using sol-gel technology, with flame retardant cotton fabric prepared using a multistep coating process. The process to obtain PDTSP started with dissolved 0.5 mol of dimethyl-3-triethoxysilanepropylphosphoramidate (DTSP) in ethanol/water (1:2), and the pH was adjusted to 4.0 with 0.1M hydrochloride solution. The siloxane polymer PDTSP was achieved by removing the solvent under reduced pressure after refluxing at 90 °C for 6 h. The synthesised scheme of PDTSP is shown in Figure 3, where siloxane bond (Si-

OCH_2CH_3) in DTSP was hydrolysed to silanol bond (Si-OH). It then copolymerised themselves to form siloxane polymer, which can be reacted with the hydroxyl group (–OH) and bound onto cellulose. The chemical structure of PDTSP was proved by ^{1}H-NMR, ^{13}C-NMR and P-NMR signals.

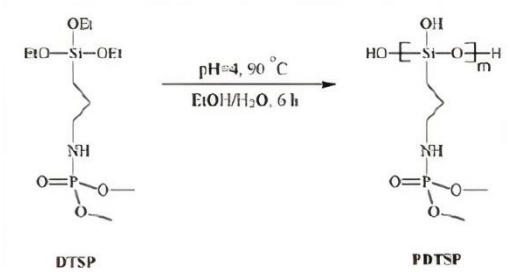


Figure 3: Synthesis route of DTSP to PDTSP (Xu *et al*., 2020).

Wan *et al*. (2020) successfully synthesised a polythiourea-phosphoric acid-based flame retardant having high amounts of phosphorus, nitrogen and sulphur elements. The route of preparation of this fire retardant is shown in Figure 4. The process started with 15.22 g of thiourea that was dissolved in 150 mL of dimethylformamide (DMF). Then, 85% of phosphoric acid was added dropwise while magnetically stirring. The mixture was refluxed at 80 °C for 4 h and then cooled to room temperature. The product obtained was washed three times with ethanol and dried at 80 °C, whereby a white solid (Thiourea-phosphate polymer, PTP) was formed. Next, PTP was dissolved in DMF solvent, and then 0.1 ml of phosphorus pentachloride (PCl_5) was slowly added under magnetic stirring. After that, 0.2 mol of ethanolamine was slowly added, with the temperature increased to 140 °C and the solution reacted for 3 h to generate the intermediate. Eventually, 0.2 mol of urea was added and reacted at 150 °C for 5 h to obtain a flame retardant water-soluble yellow solid, which was then purified three times with absolute ethanol.

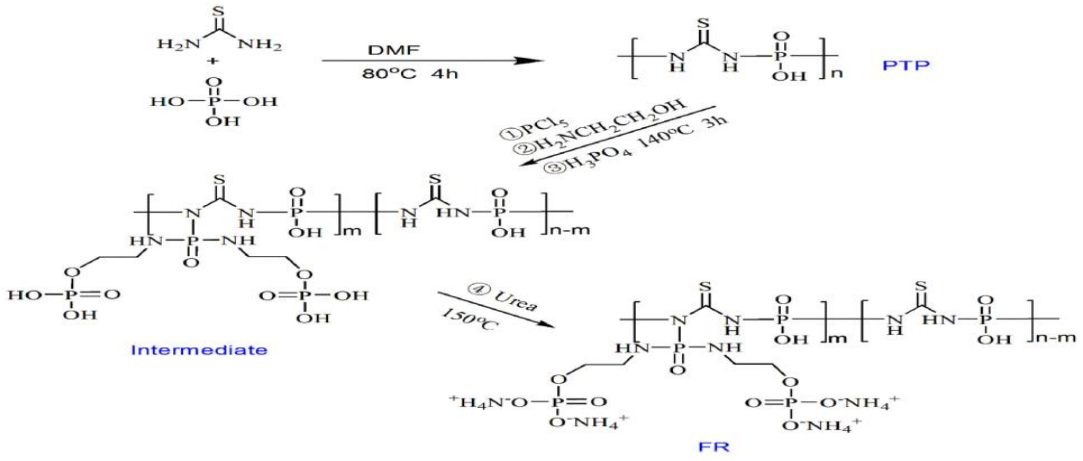


Figure 4: Synthesis route of polythiourea-phosphoric acid (Wan *et al*., 2020).

Ling & Guo (2020) synthesised a formaldehyde-free and halogen-free flame retardant, which was hyperbranched poly phosphate ammonium salt (HBPOPN) on cotton fabric to improve its durability. The process to get HBPOPN started with 8.75 g of HBP dissolved in 0.08 mol phosphoric acid (H_3PO_4) in A 500 mL flask equipped with a thermometer and magnetic stirrer. The mixture was then incubated at 130 °C for 3 h, where a viscous and transparent liquid (HBPOP) was formed. This 15.15 g of HBPOP and 4.8 g of urea were mixed and incubated at 100 °C for 1 h to obtain an off-white and viscous liquid (HBPOPN). The HBPOPN obtained was purified using ethanol and then dried at 100 °C. The synthesis scheme is shown in Figure 5. The HBPOPN FR chemical structure was proved using ^{1}H-NMR, ^{13}C-NMR and P-NMR signals.

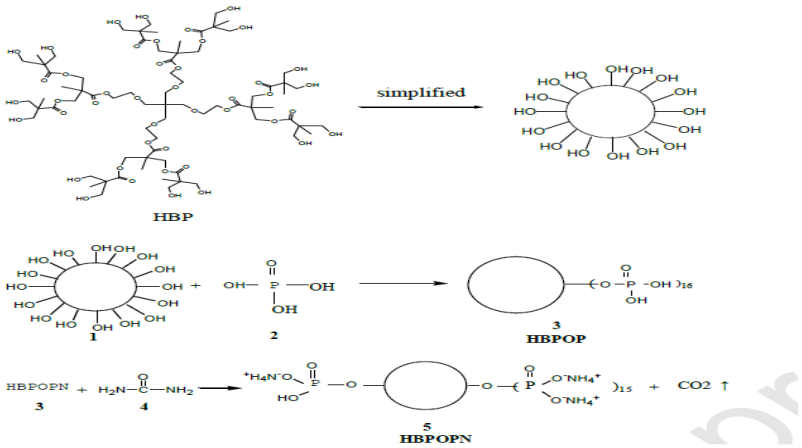


Figure 5: Synthesis route of HBPOPN (Ling & Guo, 2020).

5.2 Synthesis and Preparation of Phosphorus-Based Flame Retardant for Polyethylene Terephthalate (PET)

Peng *et al*. (2020) prepared a halogen-free phosphorus-containing silicone flame retardant poly N,N dimethylene phosphate aminopropyl siloxane (PDPSI) following the Mannich reaction specifically to reduce environmental hazards from flame retardants. This PDPSI and ferric oxide (Fe_2O_3) were utilised in the preparation of flame retardant for PET. Figure 6 shows how PDPSI was synthesised according to the Mannich reaction. The process started with 5 g of phosphorous acid dissolved in a 10 mL beaker containing 10 mL of distilled water, followed by addition of 4.5 g of (3-Aminopropyl) triethoxysilane (APTES). The mixture was stirred until a homogeneous solution was formed, before been put into a three-neck round bottomed flask equipped with a stirrer, thermometer and reflux condenser. The solution was then heated to 80 °C, with 5 mL of concentrated hydrochloric acid added to flask. The solution was stirred for 1 h to activate the catalyst. Next, 5 mL of 37% formaldehyde solution was added slowly and the reaction was allowed to continue for 3.5 h. After that, the mixture was poured into a beaker, precipitated with excess of absolute ethanol, and rapidly stirred to obtain a white powder, which was suction filtered. The filter cake was repeatedly washed with absolute ethanol, producing a organosilicon after drying at 100°C for 10 h.

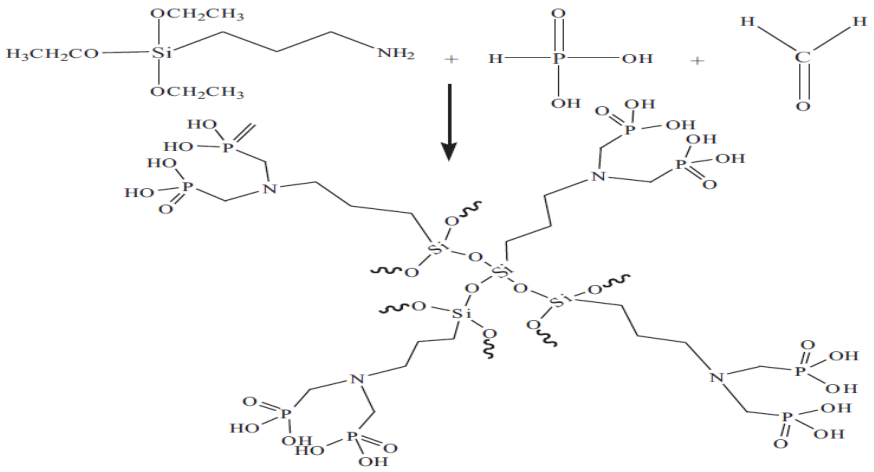


Figure 6: Synthesis route of PDPSI (Peng *et al*., 2020).

Fang *et al*. (2021) conducted a comparative research between 9,10-dihydro-9-oxa-10-phosphaphenanthrene-10-oxide (DOPO) and commercial halogen containing flame retardant as flame retardant finishing agents for PET fabrics. The chemical structure of DOPO is shown in Figure 7. They found that PET fabrics treated with DOPO yielded the same excellent flame retardancy and laundering durability with halogen containing flame retardant.

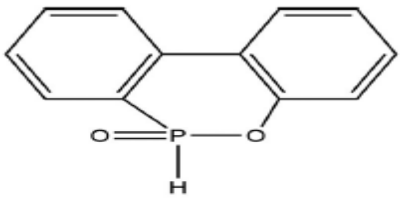


Figure 7: Chemical structure of DOPO (Fang *et al*., 2021).

5.3 Synthesis and Preparation of Phosphorus-Based Flame Retardant for Polyamide 6 (PA6)

Liu *et al*., (2018) successfully prepared flame retardant polyamide 6 (FRPA6), which was synthesised using melt polycondensation of caprolactam and 9,10-dihydro-10-[2,3-di(hydroxycarbonyl)propyl]-10-phosphaphenanthrene-10-oxide (DDP). The scheme of preparation process of FRPA6 is shown in Figure 8. The DDP salt solution was formed with addition a molar ratio of DDP to decamethylene diamine of 1 : 1. Meanwhile for FRPA6, it was prepared with addition of caprolactam, DDP salt solution, adipic acid and deionised water into a polymerisation autoclave. The air in the reactor was purged completely by nitrogen before the reaction. Then, the autoclave was heated to 250 °C and the mixture was maintained at this temperature for 3 h between 0.6 and 0.8 MPa. The temperature then was decreased to 240 °C while the pressure was dropped to atmospheric pressure, which was set steady at 240 °C under vacuum for 30 min. The product obtained was taken out from the reactor, cooled in cold water and cut into slices. The slices were then extracted with boiling water for 24 h and dried in a vacuum oven at 105 °C for 24 h. Finally, the synthesised product (FRPA6) was purified by firstly dissolving in sulfuric acid and precipitation in deionised water, then extracted with dimethylsulfoxide for 6 h and washed with deionised water. This step was repeated three times, with the final product dried in a vacuum oven at 105 °C for 24 h.

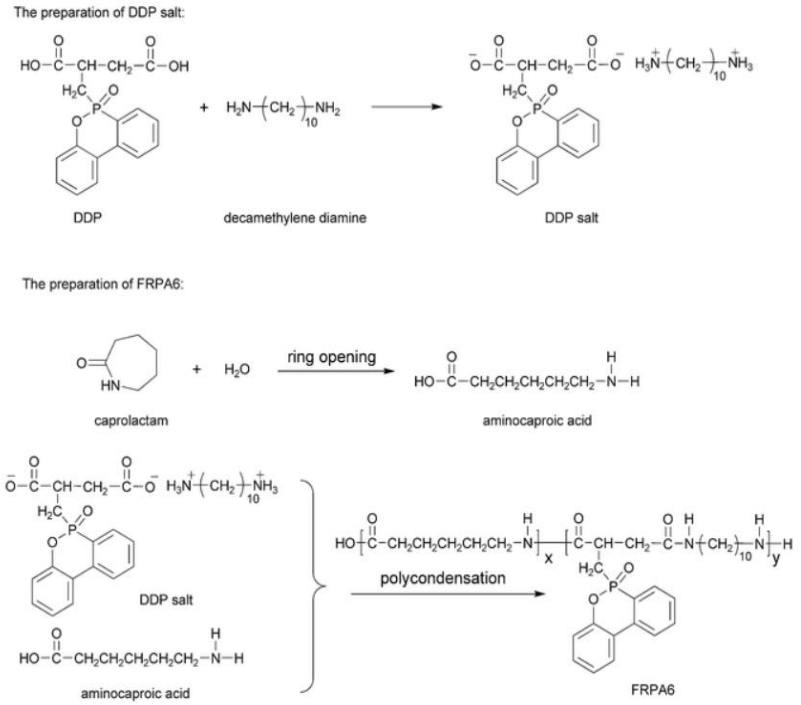


Figure 8: Preparation process for FRPA6 (Liu *et al.*, 2018).

5.4 Synthesis and Preparation of Phosphorus-Based Flame Retardant for Nylon

Chen *et al*. (2016) successfully synthesised a sulphur containing flame retardant (SFR) from polyphosphoric acid, epoxy chloropropane, and thiourea. The author used a water-soluble isocyanate-terminated (WIT) that acted as cross-linker between SFR and nylon fabric. The process to obtain SFR started with addition of polyphosphoric acid and anhydrous aluminium trichloride into a 250 mL four-necked flask. The flask was equipped with agitator and thermometer, and was heated in a water bath. An epoxy chloropropane was then added and the temperature of reaction mixture held at about 40 °C. Next, the temperature of the reaction mixture was set up to 70 °C and kept for 3 h before the intermediate product was successfully formed. After that, the intermediate product and thiourea were added to the four-necked flask. The temperature of the reaction mixture was set up to 70 °C and kept for 2-3 h while the agitator was stirring. Then, the final product was obtained by adjusting to required solid content using deionised water while the pH was adjusted to 7-8 using ammonium hydroxide. The synthesis route of SFR is shown in Figure 9.

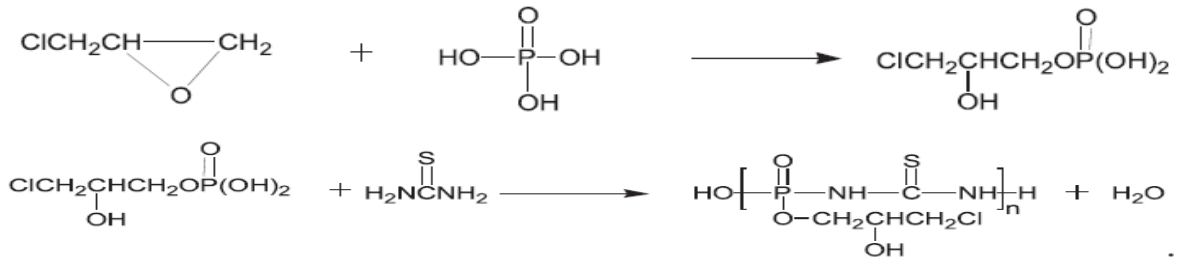


Figure 9: Synthesis route of SFR (Chen *et al*., 2016)

Jiang *et al*. (2015) successfully synthesised ditrimethylolpropane di-N-hydroxyethyl phosphoramide (DDP), which is a phosphorus-nitrogen-containing intumescent flame retardant (IFR) for various fabrics, such as nylon, cotton and PET. Figure 10 shows the schematic outline for synthesis of DDP, where (a) refers to esterification reaction, (b) hydrolysis reaction, and (c) amination reaction. For esterification reaction, ditrimethylpropane (Di-TMP), phosphoryl chloride ($POCl_3$) and 1, 4 dioxane were added into a 250 mL flask. The mixture was heated to 50 °C for 5 h. After the reaction, untreated $POCl_3$ and 1, 4 dioxane were removed, and ditrimethylpropane diphosphorus chloride (DDC) was formed. For hydrolysis reaction, DDC was dissolved in 1, 4 dioxane and distilled water was added. The mixture was heated to 72 °C for 1.5 h. Subsequently, ditrimethylpropane diphosphate eter (DDE) was formed. For amination reaction, DDE was dissolved in 1, 4 dioxane, 2-aminoethanol and boric acid. The mixture was heated to 70 °C for 4 h. After the reaction, unreacted 2-aminoethanol and 1, 4 dioxane were removed, and di-N-hydroxyethyl phosphoramide (DDP) was formed.

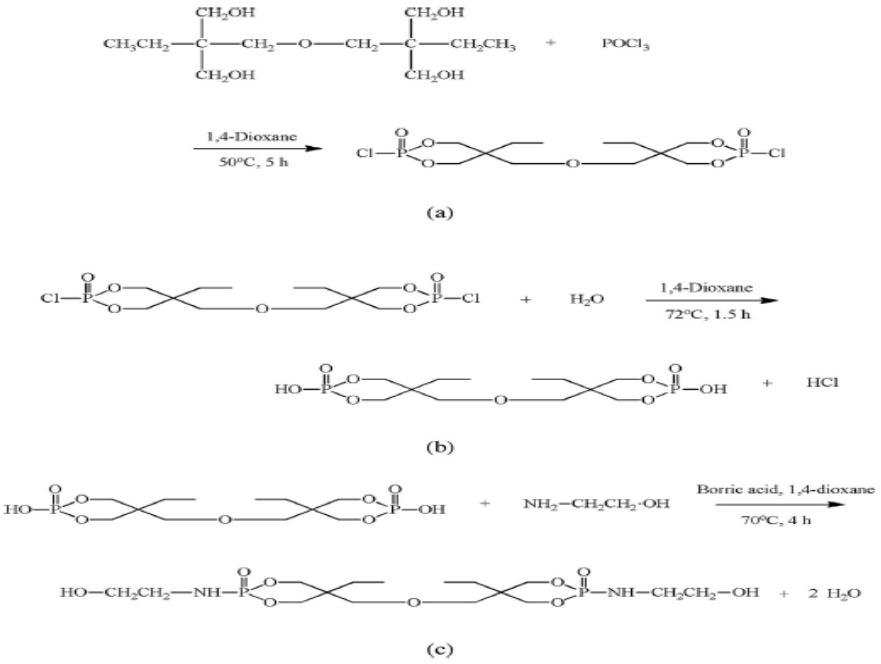


Figure 10: Schematic outline for synthesis of DDP: (a) esterification reaction, (b) hydrolysis reaction, (c) amination reaction (Jiang *et al*., 2015)

5.5 Synthesis and Preparation of Phosphorus-Based Flame Retardant for Wool

Mohaddes *et al*., (2021) successfully utilised vinyl phosphonic acid (VPA) (Figure 11) as durable and environmentally friendly flame retardant for wool via co-polymerisation. The formation of new covalent bonds between the VPA and wools resulted in superior flame retardant, durability of applied finish and, most importantly, it is environmentally friendly.

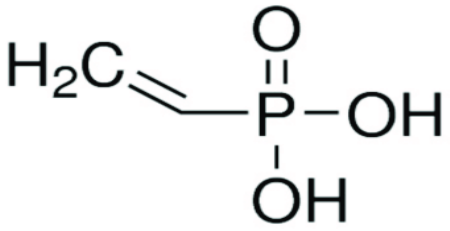


Figure 11: Chemical structure of vinyl phosphonic acid (VPA).

Wang *et al*. (2021) successfully employed an eco-friendly 2-phosphonobutane-1,2,4-tricarboxylic acid (PBTCA) as a finishing agent to improve the flame retardancy of wool fabrics using the pad-dry-cure technique. The chemical structure of PBTCA is shown in Figure 12. The flame retardancy of the treated wool fabrics showed impressive enhancement with incorporation of PBTCA.

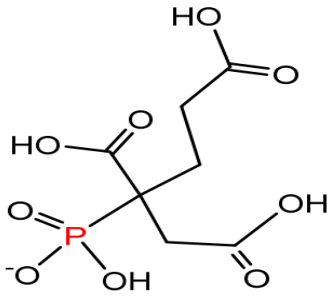


Figure 12: Chemical structure of 2-phosphonobutane-1,2,4-tricarboxylic acid (PBTCA).

6. CONCLUSION

This paper reviewed recent designs of phosphorus-based flame retardants for different types of textiles, where the impact especially on flame retardant properties was studied. From the review, the phosphorus-based flame retardants successfully enhanced the capability of treated textiles with the excellent LOI values. Furthermore, the use of phosphorus-based flame retardant successfully replaced the conventional halogen containing flame retardant that has high toxicity to humans and the environment. Phosphorus-based flame retardants turns out to have same excellent flame retardant characteristics with halogen containing flame retardants. However, the effects on application of phosphorus-based flame retardant towards the mechanical strength of the textiles has been rarely studied. Furthermore, the durability of treated textiles after many times of laundry shows decreases in performance. The suggestions for the future studies should include the improvement of durability after laundry without jeopardising the excellent flame retardancy of phosphorus-based flame retardants and focused study for mechanical strength. In conclusion, phosphorus-based flame retardants proved to have high potential to be utilised as flame retardant additives in future.

ACKNOWLEDGEMENT

This review was conducted as part of Twelfth Malaysia Plan (RMK12) project entitled *Development of Smart Fabric*.

REFERENCES

Armarnath, N., Appavoo, D. & Lochab, B. (2019). Eco-friendly halogen-free flame retardant cardanol polyphosphazene polybenzoxazine networks. *ACS Suistainable Chem Eng*., **6**: 389-402.

Babushok, V.I., Linteris, G.T., Meier, O.C. & Pagliaro, J.L. (2014). Flame inhibition by CF3CHCl2 (HCFC-123). Combust. Sci. Technol., **111**: 149-182.

Brushlinsky, N.N., Sokolov, S.V., Wagner, P. & Hall, J.R. (2006). World fire statistics, Report No 10.

Center of Fire Statistics. *International Association of Fire and Rescue Service*, Ljubljana, Slovenia

Byrne, C. (2000). Technical textiles market- An overview. Horrocks, A.R. & Anand, S.C. (Eds), Handbook

of technical textiles. Woodhead Publishing, United States of America, pp. 1-23.

Chanchal, K.K., Zhiwei, L., Lei, S. & Yuan. H. (2020). An overview of fire retardant treatment for synthetic textiles: From traditional approaches to recent applications. *Eur. Polym*. J., **137**: 109-171.

Chen, J., Kan, W., Yi, L., Chuck, Z. & Ben, W. (2021). Application of textile technology in tissue engineering: A review. *Acta Biomater*., **128**: 60-76.

Chen, Y., Sun, B., Zhang, H. & Zhou, X. (2016). Synthesis and application of a sulfur-containing phosphoric amide flame retardant for nylon fabric. *Fire Mater*., **7**:959-972.

Dasari,A., Yu, Z.Z., Cai, G. & Mai, Y.M. (2013). Recent developments in fire retardancy of polymeric materials. *Prog. Polym. Sci*., **38**:1357-1387.

Erdem, N., Cireli, A.A. & Erdogan, U.H. (2009). Flame retardancy behaviors and structural properties of polypropylene/nano-SiO_2 composite textile filaments. *J. Appl. Polym. Sci*., **111**: 2085-2091.

Fang, Y., Liu, X. & Wu, Y. (2021). High efficient flame retardant finishing of PET fabric using eco-friendly DOPO. *J. Text. Inst*., **112**: 1-8.

Gann, R.G. (2004). Estimating data for incapacitation of people by fire smoke. *Fire Technol*., **40**: 201-207.

Gaan, S. & Sun, G. (2009). Effect of nitrogen additives on thermal decomposition of cotton. *J. Anal. Appl.*
Pyrolysis., **84**: 108-115.

Gordon, B. (2010). The fiber of our lives: A conceptual framework for looking at textile meaning. *Textile Society of America-12th Biennial Symp.*, **18**: 6-9.

Granzow, A. (1978). Flame retardation by phosphorus compounds. *Acc. Chem. Res.*, **11**: 177-183.

Guillaume, E., Marquis, D. & Saragoza, L. (2014). Calibration of flow rate in cone calorimetry test. *Flame Mater*., **38**: 194-203.

Horrocks, A.R., Tunc, M. & Price, D. (1989). The burning behaviour of textiles and its assessment by oxygen-index methods. *Text. Prog*., **18**: 1-186.

JBPM (Fire and Rescue Department of Malaysia) (2019). *Statistik Kebakaran Struktur Yang Disiasat Mengikut Sumber Nyalaan Bagi Tahun 2018*. Fire and Rescue Department of Malaysia (JBPM), Malaysia.

Jiang, W., Jin, F.L. & Park, S.J. (2015). Synthesis of a novel phosphorus-nitrogen-containing intumescent flame retardant and its application to fabrics. *J. Ind. Eng. Chem*., **27**: 40-43.

Kandola, B.K (2012). Flame retardant characteristics of natural fiber composite. *Polym. Compos*., **1**: 86-117.

Khalifah, A. S., Julien, F., Shuyu, L. & Sabyasachi, G. (2015). An overview of mode action and analytical methods for evaluation of gas phase activities of flame retardants. *Polymers*., **7**: 504-526.

Khalifah, A. S., Sabyasachi, G. & Giulio, M. (2016). Recent advances for flame retardancy of textiles based on phosphorus chemistry. *Polymers*., **8**: 319-355.

Kimmey, C., Sargunas, P., Iaccarino, P. & Benz, S. (2020). Analysis of a novel phosphorus-nitrogen based flame retardant for cotton. *J. Undergrad. Chem. Res*, 12-18.

Laoutid., F., Bonnaud, L., Alexandre, M., Lopez-Cuesta, J.L & Dubois, P. (2009). New prospects in flame retardant polymer materials from fundamentals to nanocomposites. *Mater. Sci. Eng. R*., **63**: 100-125.

F. Laoutid., L. Bonnaud., M. Alexandre., J.M. Lopez-Cuesta. & P. Dubois. (2009). New prospects in flame retardant polymer materials from fundamentals to nanocomposites. *Mater. Sci. Eng. R*., **63**: 100-125.

Linda, S. & Daiele, F.S. (2004). Brominated flame retardants: Cause for concern?. *Environ. Health Perspect*., **112**: 9-17.

Ling, C. & Guo, L.M. (2020). Preparation of a flame-retardant coating based on solvent-free synthesis with high efficiency and durability on cotton fabric. *Carbohydr. Polym*., **230**.

Liu, K., Li, Y., Tao, L. & Xiao, R. (2018). Preparation and characterization of polyamide 6 fibre based on a phosphorus-containing flame retardant. *RSC Adv*., **8**: 9261-9271.

Lu, S.Y. & Hamerton, I. (2002). Recent developments in the chemistry of halogen free flame retardant polymers. *Prog Polym Sci*., **27**: 1661-1712.

Madhav, S., Ahamad, A., Singh, P. & Mishra, P.K. A review of textile industry; wet processing, environmental impacts, and effluent treatment methods. *Environ. Qual. Manag*., **27**: 31-41.

Markwart, J.C., Battig, A., Zimmermann, L., Wagner, M., Fischer, J., Schartel, B. & Wurm, F.R. (2019). Systematically controlled decomposition mechanism in phosphorus flame retardants by precise molecular
architecture: P-O bs P-N. *Appl. Polym. Mater*., **1**: 1118-1128.

Mngomezulu, M.E., John, M.J., Jacobs, V. & Luyt, A.S. (2014). Review on flammability of biofibres and biocomposites. *Carbohydr. Polym*., **111**: 149-182.

Mohaddes, F., Islam, S., Padhye, R. & Wang, L. (2021). Durable and environmentally friendly flame-retardant finish on wool via graft copolymerisation of vinyl phosphonic acid. *Biointerface Res. Appl. Chem*., **12**: 3647-3663.

Nazir, R. & Gaan, S. (2020). Recent developments in P(O/S)-N containing flame retardants. *J. Appl. Polym.*
Sci., **137**: 47910-47937.

Patel, P., Hull, T.R. & Moffatt, C. (2012). Peek polymer flammability and the inadequacy of the UL-94 classification. *Fire Mater*., **36**: 185-201.

Peng, Y., Niu, M., Qin, R., Xue, B. & Shao, M. (2020). Study of flame retardancy and smoke suppression of PET by the synergy between Fe_2O_3 and new phosphorus containing silicone flame retardant. *High Perform. Polymer*., **32**.

Pettigrew, A. (1993). Halogenated flame retardants. In: Kirk-Othmer Encyclopaedia of chemical technology, Wiley. Choudry, A.K.R (Eds), Flame retardants for textile materials. CRC Press Taylor & Francis Group, pp. 954-976.

Salmeia, K.A., Fage, J., Liang, S. & Gaan, S. (2015). An overview of mode of action and analytical methods for evaluation of gas phase activities of flame retardants. *Polymers*., **7**: 504-526.

Schartel, B. (2010). Phosphorus-based flame retardancy mechanism solid hat or a starting point for future development. *Materials*., **3**: 4710-4745.

Tata, J., Alongi, J., Carosio, F. Frache, A. (2011). Optimization of the procedure to burn textile fabrics by cone calorimeter: Part I. Combustion behavior of polyester. *Fire Mater*, **35**: 397-409.

Waaijers, S.L. & Parsons, J.R. (2016). Biodegradation of brominated and organophosphorus flame retardants. *Curr. Opin. Biotechnol*., **38**: 14-23.

Wan, C., Liu, M., He, P., Zhang, G. & Zhang, F. (2020). A novel reactive flame retardant for cotton fabric based on a thiourea-phosphoric acid polymer. *Ind Crops Prod*., **154**: 1-10.

Wang, H., Guo, S., Zhang, C., Qi, Z., Li, L. & Zhu, P. (2021). Flame retardancy and thermal behaviour of wool fabric treated with a phosphorus-containing polycarboxylic acid. *Polymers*., **13**.

Xu, D., Wang, S., Wang, Y., Liu, Y., Dong, C., Jiang, Z. & Zhu, P. (2020). Preparation and mechanism of flame-retardant cotton fabric with phosphoramidate siloxane polymer through multistep coating. *Polymers*., **12**: 1-14.

Xu, F., Zhang, G., Wang, P. Dai, F. (2022). Durable and high-efficiency casein-derived phosphorus nitrogen-rich flame retardants for cotton fabrics. *Cellulose*., **29**: 2681-2697.

Ying, L., Qimin, C., Huabo, D., Yicheng, L., Juan, Z. & Jinhui, L. (2019). Occurrence, level and profiles of brominated flame retardants in daily-use consumer products on the Chinese market. *Environ. Sci. Process Impact*., **21**:446-455.

Zhang, Z., Ma, Z., Leng, Q. & Wang, Y. (2019). Eco-friendly flame retardant coating deposited on cotton fabrics from bio-based chitosan, phytic acid and divalent metal ions. *Int. J. Biol. Macromol*., **140**: 303-310.

Zheng, M. & Zhang, P. (2010). On the flammability properties of interior decorative fabric. *J. Chin. People's Armed Police Acad*., **2**: 15-18.

MICROMECHANICAL STUDY ON HYBRID CARBON AND GLASS FIBRE REINFORCED POLYMER PROPERTIES

Ahmad Fuad Ab Ghani[1*], Ridhwan Jumaidin[1], Mohamed Saiful Firdaus Hussin[1], Mohd Fariduddin Mukhtar[1], Sivakumar Dharlingam[2] & Rahifa Ranom[3]

[1]Faculty of Engineering Technology Mechanical and Manufacturing (FTKMP)
[2]Faculty of Mechanical Engineering (FKM)
[3]Faculty of Electrical Engineering (FKE)
Universiti Teknikal Malaysia Melaka (UTeM), Malaysia

[*]Email: ahmadfuad@utem.edu.my

ABSTRACT

In order to achieve both design adaptability and reduction of cost, there is a need to develop carbon / glass hybrid composites and evaluate their mechanical properties. This is a major challenge that can only be met through an understanding of the relationships between material architecture and mechanical response, as well observing microstructure formation. This research work assesses deformation, behaviour and failure prediction of real scale hybrid composite carbon and glass fibre reinforced polymer (hybrid C / GFRP) using microscale representative volume element (RVE). RVE for hybrid C / GFRP is assumed to have isotropic behaviour for carbon fibre, glass fibre and epoxy resin matrix, as well as assumed to be a perfectly bonded interface between the fibre and matrix regions, i.e., strain compatibility at the interface. Multiscale modelling of hybrid C / GFRP via RVE results is presented, which proves to be a practical tool for modulus of elasticity prediction. The results of computation of modulus of elasticity from RVE are then compared to real scale computation from experimental results. The failure mode observed from experimental results is also compared at the microlevel from RVE deformation perspective.

Keywords: *Representative volume element (RVE); micromechanical; composite; hybrid composite; finite element modelling (FEM).*

1. INTRODUCTION

Damage accumulates in a widespread fashion in composites, with many individual processes occurring at the microstructural level (Cai *et al*., 2017; Zhou *et al*., 2017). Micromechanical is the analysis on the level of individual constituents that represents the whole material of composite materials. It is also defined as heterogeneous materials that are composed of diverse parts that occupy the same volume (Zhang *et al*., 2012). Representative volume element (RVE) finite element modelling (FEM) provides answers on fibre / matrix deformation and homogenous study of hybrid composites at the microscale level. The use of FEM enables a thorough study and understanding of multiscale modelling, which comprises of the concept of RVE at the microscale level and layup at the macroscale level (Qingping *et al*., 2018)

Tan *et al.* (2000) studied the behaviour of 3D orthogonal woven carbon fibre reinforced polymer (CFRP) composites using FEM and analytical modelling approaches. FEM was employed to predict the stiffness constants of 3D orthogonal woven CFRP composites. In order to validate the models, a number of tensile tests were conducted. The method was developed based on the isostrain assumption, which refers to the local strain compatibility between fibre /

matrix and maximum stress failure criterion**.** In micromechanics, the effective properties and response of composites are computed based on the properties and response of the individual constituents (Babu *et al*., 2008; Silva *et al*., 2012).

The objective of this study is to develop an equivalent continuum that simulates the average mechanical behaviour and deformation as the real homogenous material. One of the key assumptions in the micromodelling of fibre / matrix is the perfect bonding (i.e., local strain compatibility) between the fibre reinforced polymer (FRP) and matrix. This assumption has also been used in the rule of mixture computation and Halphin Tsai equation to estimate the elastic of modulus of lamina (Harris, 1999). Homogenisation is a theory that allows us to obtain the significant dependence of macroscale on microscale. The selected RVE should be comparable and should be sufficient to represent the feature of the material, as well as be as small as possible to reduce the computation cost (Qingping *et al*. 2018). More importantly, the RVE should have the same volume fraction with the composite (Jiang *et al*., 2014). The conventional method of doing it is by performing experiments that are time consuming and costly. Therefore, numerical simulation is an alternative in the prediction of the overall property of the unidirectional composite (Karkkainen & Sankar, 2006; Wang *et al*, 2016; Qingping *et al*, 2018).

2. THEORETICAL BACKGROUND

Macroscale is usually referred to a homogenised size and microscale is usually related to a representative volume element, as shown in Figure 1. Both Figures 1(a) and 1(b) can be chosen as unit cell representation of RVE. The unit cell on the square packed array has been chosen to represent the homogenisation of FRP for this study as shown in Figure 1(a). There are also other RVE shapes such as the hexagonal packed array but normally square packed array is selected (Alfaro *et al.*, 2010; Abbassi *et al.*, 2011). In the generation of RVE for hybrid composite, it is necessary for the RVE or unit cell to be isolated for simplicity. The RVE model generated should possess an equal value of elastic constant and fibre volume fraction as the composite as shown in Figure 2.

Figure 1: Example of RVE to represent homogenous of material unit cell on: (a) Square packed array (b) Hexagonal packed array (Aboudi *et al*, 2013).

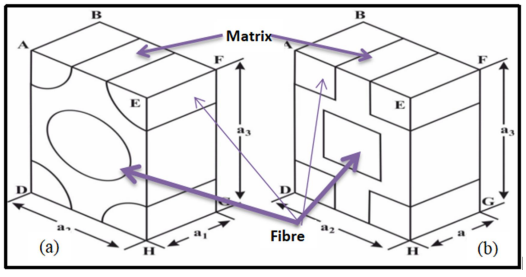


Figure 2: Hexagonal shaped RVE with: (a) Circular fibre (b) Square fibre (Bhaskara *et al*., 2014).

The computation of modulus of elasticity, which depends on the volume fraction of the composite, is also performed on the basis of micromechanical theory (Qingping *et al.*, 2018). Banerjee & Sankar (2014) investigated the micromechanical analysis of the RVE of a unidirectional hybrid composite, which was performed using finite element method. The fibre location inside an RVE for every volume fraction ratio of fibres is determined to be at a random position. The radius of fibres was assumed to be equal while the number of glass and carbon fibres varies within the RVE in order to change the volume fraction. It is assumed that the fibres have a circular cross-sectional area and arranged hexagonally across the RVE (Banerjee & Sankar, 2014). In modelling the micromechanics, the simplest method to estimate the stiffness of a composite is where the direction of the fibre is parallel to its applied load, which is made with an assumption the structure is a simple beam, and both of components are perfectly bonded and in turn equally deformed. The sum of the volume fraction of the matrix and the fibre is equated through volume fraction for fibre and volume fraction for a matrix as shown in Figure 3.

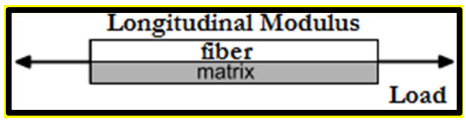


Figure 3: Simplified parallel model of a unidirectional composite (Harris, 1999).

The load of composite (P_c) is shared between two phase where P_c (load borne by composite) equals the summation of P_f (load borne by carbon fibre) and P_m (load borne by matrix), and the strain in the two phases is as same as in the composite, i.e.,, ε_c equals ε_f and ε_m respectively (isostrain). Since stress equals load divided by area:

$$\boldsymbol{\sigma_c A_c = \sigma_f A_f + \sigma_m A_m} \tag{1}$$

where σ_c is the stress borne by the composite, A_c is the area covered by the composite, σ_f is the stress beared by fiber, A_f is the cross section area of the fiber, σ_m is the stress held by the matrix, and A_m is the area covered by the matrix.

From isostrain (Voigt estimate):

$$\boldsymbol{E_c = E_f V_f + E_m (1 - V_f)} \tag{2}$$

where E_c, E_f and E_m are the modulus of elasticity for the composite, fibre and matrix respectively, and V_f is the volume fraction.

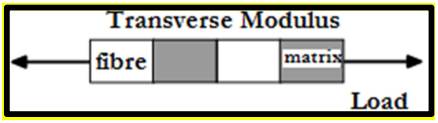


Figure 4: Simple series model of composite under transverse loading (Harris, 1999).

In estimating the transverse modulus (E_t), a similar approach is utilized with the same constraints as longitudinal modulus, such as well-bonded components with similar Poisson ratios as shown in Figure 4. It is an isostress model with stress of composite that is assumed to be equal to the stresses held by the fibre and matrix. The total elongation is the sum of both components of elongation:

$$\varepsilon_c L_c = \varepsilon_f L_f + \varepsilon_m L_m \tag{3}$$

where ε_c, ε_f and ε_m represent the strains of the composite, fibre and matrix respectively. Meanwhile L_c, L_f and L_m are the original lengths of the composite, fibre and matrix respectively.

$$L_c = \varepsilon_f L_f + \varepsilon_m L_m$$

If L equals V, so dividing through by the stress produces relation as below:

$$\frac{\mathbf{1}}{\boldsymbol{E_t}} = \frac{\boldsymbol{V_f}}{\boldsymbol{E_f}} + \frac{\boldsymbol{V_m}}{\boldsymbol{E_m}} \tag{4}$$

Therefore, the transverse modulus is (Reuss estimate):

$$\boldsymbol{E_t} = \frac{\boldsymbol{E_f E_m}}{\boldsymbol{E_m V_f} + \boldsymbol{E_f}(\mathbf{1} - \boldsymbol{V_f})} \tag{5}$$

where E_t, E_f and E_m are the modulus of elasticity transverse, modulus of elasticity of the fibre and modulus of elasticity of the matrix respectively.

A RVE should contain sufficient information about the microstructure and be a good representation of continuum mechanics (Drathi & Ghosh, 2015). Another purpose of the micromechanical study is the distribution of stresses and strains within the micro-regions of the composite under loading (Babu *et al*., 2008). Throughout the studies, the geometry of the RVEs are subjected to a simple form of loading in order to evaluate the average stiffness, which is also known as modulus of elasticty of the composites, using the RVE method. The results of simulation and experimental approaches of micromechanics can assist in understanding the sharing of load among the constituents of the composites and microscopic structure (arrangement of fibres) within the composites. Microstructural parameters that influence the composite behaviour are fibre diameter, length, volume fraction, packing and orientation of fibre (Babu *et al*, 2008). This supports the hypotheses that RVE could be a tool for the design of unidirectional (UD) hybrid composite CFRP / GFRP for optimum mechanical properties amongst others.

Wang *et al.* (2016) performed a study on FEM with random RVEs and the results demonstrated that the randomness of fibre distribution has insignificant effect on the effective elastic properties, resulting in very small deviation values for all predicted elastic constants. Valavala *et al.* (2009) discovered that for each set of Young's modulus predictions, as the RVE sizes increase, the predicted values of Young's modulus match closely with the experimental value for both polymer systems and force fields. Deviation generally decreases as the RVE size increases for the simple averaging approach. The Voigt model assumes that the strains in all the phases of a composite material are the same for a given macro scale deformation (Valavala *et al.*, 2009).

3. METHODOLOGY

The FEM of composite material deformation is conducted using ABAQUS 6.14 with built-in feature for validation of composites. The scope of the study on the micromechanical modelling via RVE is limited to only unidirectional composites. Multiscale modelling using RVE is also introduced in this study as another practical approach in predicting the modulus of elasticity of C / GFRP hybrid composites. The chosen model and geometry of RVE are discussed with respect to this hybrid composite.

FEM is an approximate numerical method where the basic theory is the physical discretisation of a continuum (Prabu *et al*., 2004). This implies dividing an accounted domain on a finite number of small dimensions and simple shapes, which represent the basis for all considerations. Mesh generation is the process of dividing a certain area into nodes and finite elements. This type of element is used throughout the research to capture the behaviour in tensor stresses in the x, y and z directions. The microstructure-based approach describes the local material microstructure as two individual phases (fibre and epoxy), each with unique material properties. The concept of RVE is implemented to determine the composite material mechanical properties.

3.1 Material Properties for RVE

The constituent material properties can be seen in Table 1, where the properties listed have been used in the simulation. The properties were obtained from the literature (DOD, 2002; Kanit *et al,* 2003; Kutz, 2005) for material engineering databases.

Table 1: The material properties of constituents for 3D RVE modelling (DOD, 2002; Kanit *et al,* 2003; Kutz, 2005)

Mechanical Properties	Carbon Fibre	E-Glass Fibre	Epoxy Resin
Young's Modulus, E (GPa), Longitudinal	230	35.4	3.4
Young's Modulus, E (GPa), Transverse	22		
Poisson's Ratio, v	0.3	0.23	0.37
Tensile Strength (MPa)	4900	2000	85
Fibre Diameter (μm)	7	12	-

2.2 3D RVE Model of CFRP and GFRP

The purpose of micromechanical simulation is to investigate the distribution of stresses and strains within the micro-regions of the composite under loading. Throughout the study, loading and geometry are subjected to a simple form in order to evaluate the average of composite stiffness. Figure 5 depicts the 3D geometry of RVE of CFRP and GFRP modelled using Abaqus 6.14. The diameter of carbon fibre is approximately 7 μm, based on observation at 5,000 times magnification using a scanning electron microscope (SEM), while the diameter of glass fibre is observed as 12 μm.

The square element arrangement is used in modelling the RVE of composites as it is the most relevant and practical representation of RVE. The size for RVE is determined by using the value of fibre diameter and volume fraction, which assumed to be 60%. The size of RVE is calculated using the Equation 6, where V_C and Vf are the composite and fibre volume fractions respectively.

$$V_F = \frac{v_f}{v_C} = 0.6 \tag{6}$$

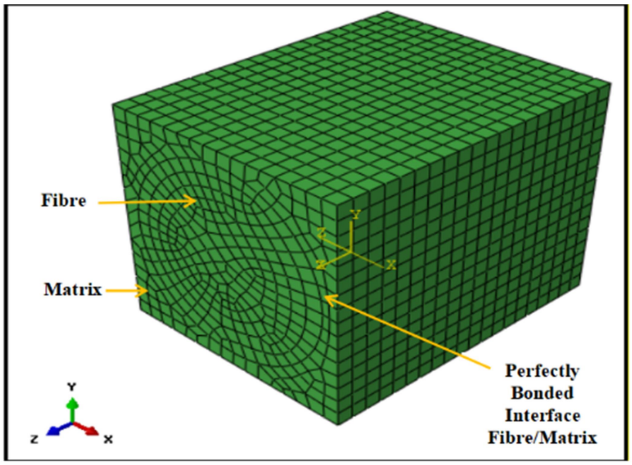


Figure 5: Geometry for 3D RVE of CFRP and GFRP.

The calculation for the 3D CFRP RVE model is as follows:

$$0.6 = \frac{v_f}{v_C} = \frac{A_f}{A_C} \tag{7}$$

By inserting the value of cross section (A_f), the following equation is obtained:

$$A_C = 64.149\mu m^2$$
$$L_C = \sqrt{64.149} = 8\mu m \tag{8}$$

where V_f is the volume for the fibre, Vc is the volume for the composite, Ac is the cross-sectional area for the composite, and L_c is the length of the RVE dimension for the cross section. The model used for 3D CFRP assumes that the fibre is a perfect cylinder with a length of 15 µm and diameter of 7 µm, as observed using a SEM, with the calculation shown above obtaining the length of the RVE for CFRP. The fibre is filled in a cube of 8 x 8 x 15 µm^3 of matrix. It is assumed that the geometry, material and loading of the RVE are symmetrical with respect to the *x-y-z* coordinate system.

The calculation for 3D GFRP RVE model is as follows:

$$0.6 = \frac{v_F}{v_C} = \frac{A_F}{A_C} \tag{9}$$

Dividing the area of the fibre (A_f) with volume fraction of 0.6 yields:

$$A_C = 189\mu m^2$$

$$L_C = \sqrt{189} = 13.7\mu m \tag{10}$$

Meanwhile, the model used for 3D GFRP is based on the assumption that the fibre is a perfect cylinder with a length of 25 µm and diameter of 12 µm, with the calculation shown above used to obtain the length of the RVE. The fibre was filled in a cube of 13.7 x 13.7 x 25 µm^3 of matrix. It is also assumed that the geometry, material and loading of the RVE are symmetrical with respect to *x-y-z* coordinate system as shown in Figure 5.

2.3 Boundary Conditions

The boundary conditions for the RVE models were based on tensile test experiments. The displacement and rotational displacement for fixed boundary condition were all constrained in the x (1), y (2) and z (3)-directions. The boundary conditions were set in accordance to the aims of the computation from the RVE, which are longitudinal, transverse and shear modulus of elasticity as tabulated in Table 2. Figure 6 illustrates boundary conditions of the 3D RVE model, showing the region where the displacement load is applied and fixed region being located at the back.

Table 2: Boundary conditions for 3D model RVE for each loading type.

Analysis for Modulus of Elasticity	Boundary Conditions	Description / Surface
Longitudinal	Fixed	(U1=U2=U3=UR1=UR2=UR3 =0) BACK
	Displacement control	U3 ≠ 0
Transverse	Fixed	(U1=U2=U3=UR1=UR2=UR3 =0) WEST
	Displacement control	U1 = 1 µm EAST
Shear	Fixed	(U1=U2=U3=UR1=UR2=UR3 =0) BACK
	Displacement control	U1 = 1 µm FRONT

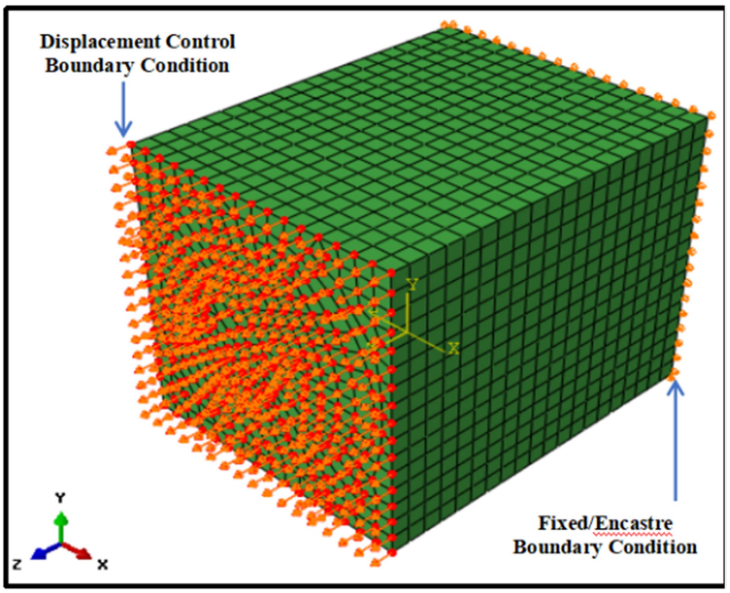


Figure 6: Boundary conditions applied on tensile test simulation of the RVE, with the displacement load at the front showing the fixed at the back.

2.4 RVE With Multiple Fibres

In forming hybrid composite C / GFRP, which consists of several fibres at different sizes, an attempt was made in modelling RVE with nine units of CFRP in order to validate the model before simulation of hybrid composites C / GFRP RVE. The size of the cross section of RVE with nine fibres was determined as 22.74×22.74 according to its volume fraction. Figure 7 shows the geometry and finite element meshing of RVE with nine units of CFRP

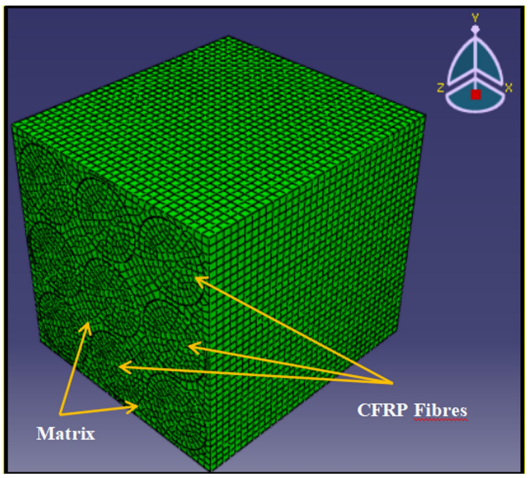


Figure 7: Finite element meshing for RVE with nine fibres of CFRP.

Figure 8 shows the geometry and boundary conditions of RVE for the purpose of computing the modulus of elasticity in longitudinal direction (E_1).

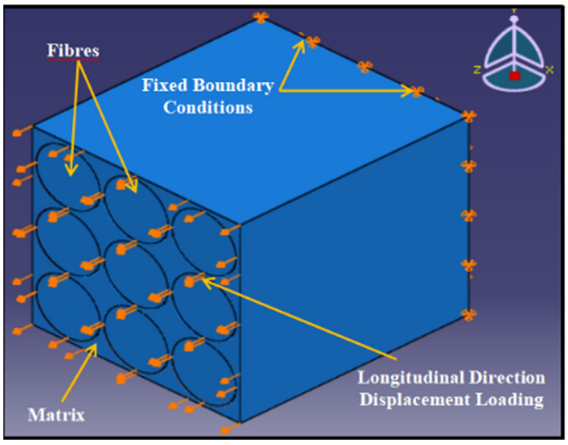


Figure 8: RVE with longitudinal direction tension loading.

2.5 Prediction of Modulus of Elasticity of Hybrid Composite C / GFRP Using RVE

With the aim to predict deformation and material stiffness of several unidirectional hybrid composite C / GFRP, the RVE geometry has been extended to model and simulate three types of hybrid composite C / GFRP, which are hybrids #W, #R and #S. Considering the fact that these hybrid composites consist of at least five layups of CFRP and GFRP, an attempt to model three layers of CFRP, which comprises of nine units of fibres, was performed. This will validate the formation of multi-layup and fibres of hybrid composite C / GFRP to compute modulus of elasticity. Figure 9 depicts the configuration of composite layup of hybrid #R, which is represented as $[0^oC/0^o6G/0^oC]_T$, hybrid #W is represented as $[0^oC/90^oG/0^oC/90^oG/0^oC]_T$, while hybrid #S is represented as $[0^o3G/0^o2C/0^o3G]_T$. Figure 10 depicts boundary conditions set for hybrid #R under tensile loading.

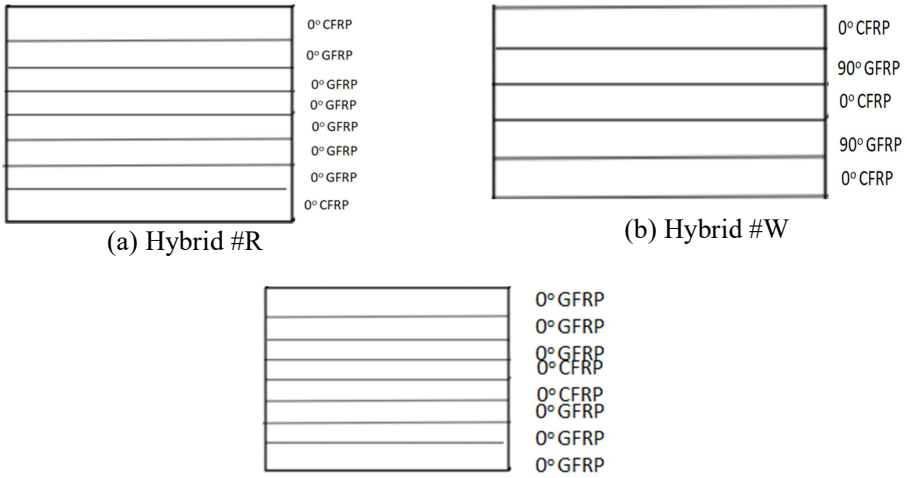


Figure 9: Composite configurations of hybrids #R, #W and #S.

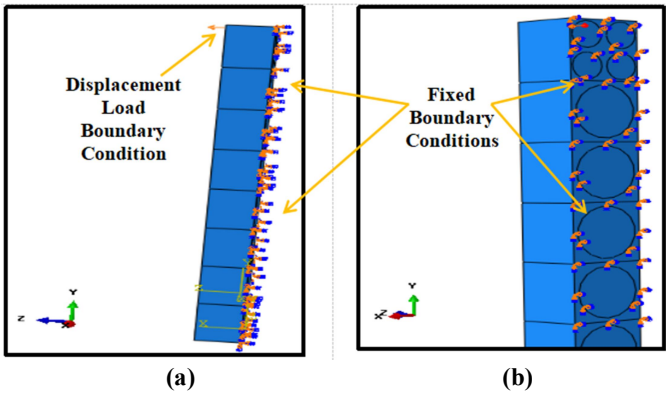


Figure 10: Boundary conditions for hybrid #R: (a) Side view showing the displacement control load. (b) Fixed surface at the back of the RVE.

3. RESULTS AND DISCUSSION

For comparison of microscale RVE and macroscale FEM, the results in terms of modulus of elasticity agreed with each other with a deviation of within 7 to 9%, as shown in Table 3. Meanwhile, Table 4 presents the deviation of modulus of elasticity obtained from the RVEs for hybrids #R, #S and #W with experimental values. The range of deviations between 3 to 9.82% is considered to be within close proximity.

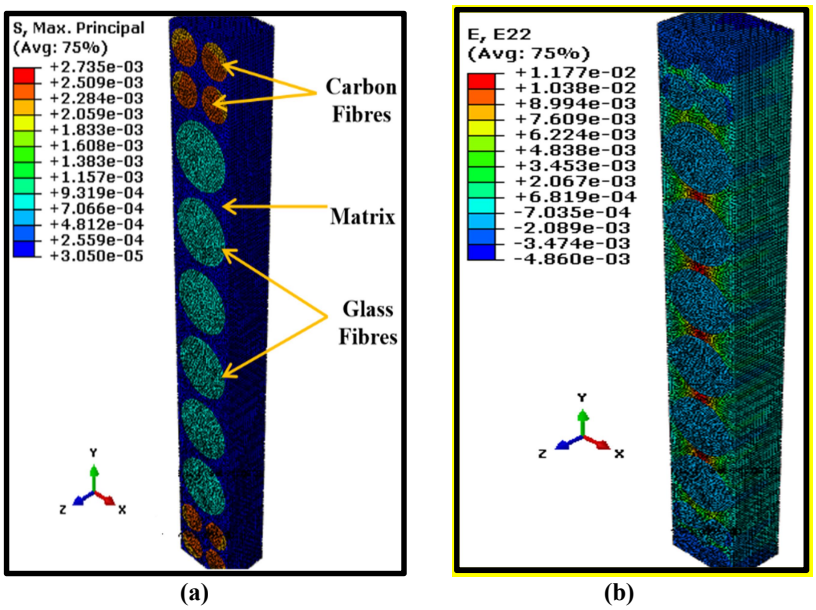
Table 3: Results modulus of elasticity for hybrids #R, #S and #W using microscale RVE and macroscale FEM.

Hybrid Composites	Modulus of Elasticity (E_{11}) obtained from microscale RVE	Modulus of Elasticity(E_{11}) obtained from macroscale FEM	Deviation (%)
Hybrid #R	68.531	74.779	8.35
Hybrid #S	68.521	74.776	8.36
Hybrid #W	89.42	96.854	7.67

Table 4: Results of modulus of elasticity for hybrids #R, #S and #W using RVE and experimental technique (digital image correlation).

Hybrid Composites	Modulus of elasticity (E_{11}) obtained from microscale RVE	Modulus of elasticity(E_{11}) obtained from experimental technique	Deviation (%)
Hybrid #R	68.531	70.72	3.09
Hybrid #S	68.521	75.99	9.82
Hybrid #W	89.42	85.59	4.47

On the whole, the effect of hybridisation was found to affect the modulus of elasticity. The sequence of CFRP at the outer layers and GFRP layups in the interior layers for hybrid #R as compared to hybrid #S is insignificant, whereby hybrid #S recorded slightly lower modulus of elasticity. This is due to the fact that GFRP, which has less modulus of elasticity than CFRP, plays dominant role in deformation of the hybrid composite. This subsequently brings slightly lower elasticity overall. Carbon fibre holds higher stress, as reflected from the peak contour as compared to glass fibre. CFRP fibre recorded higher stress when held under tensile as compared to lower modulus of elasticity of GFRP for the RVE of hybrid #R subjected to tensile loading, as shown in Figure 11 (a). On the other hand, Figure 11 (b) exhibits contours of strain $\mathcal{E}_{22}$ in the Y global direction for hybrid # R, where high strain was noticed at the interface of the GFRP which was induced to delamination in the real scenario. This is further supported by the outcome of the SEM view on the fractured sample of hybrid #R, as shown in Figure 12, which indicates that the intralayer (GFRP layups) delamination originated from matrix cracking and propagation in transverse direction during tensile loading.



(a) (b)

Figure 11: Contours of (a) maximum principal stress and (b) strain ($\mathcal{E}_{22}$) in the *Y* global direction for hybrid #R under tensile loading.

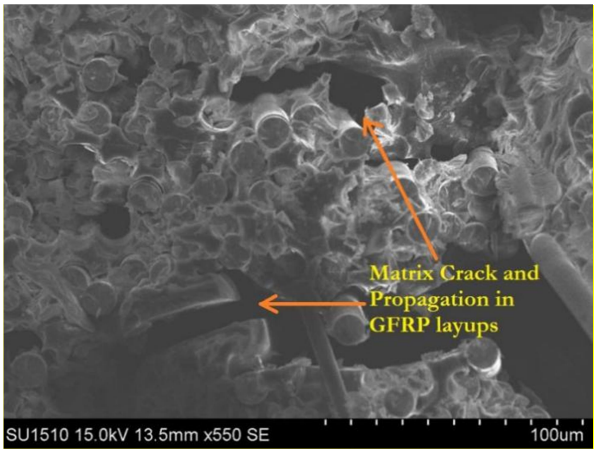
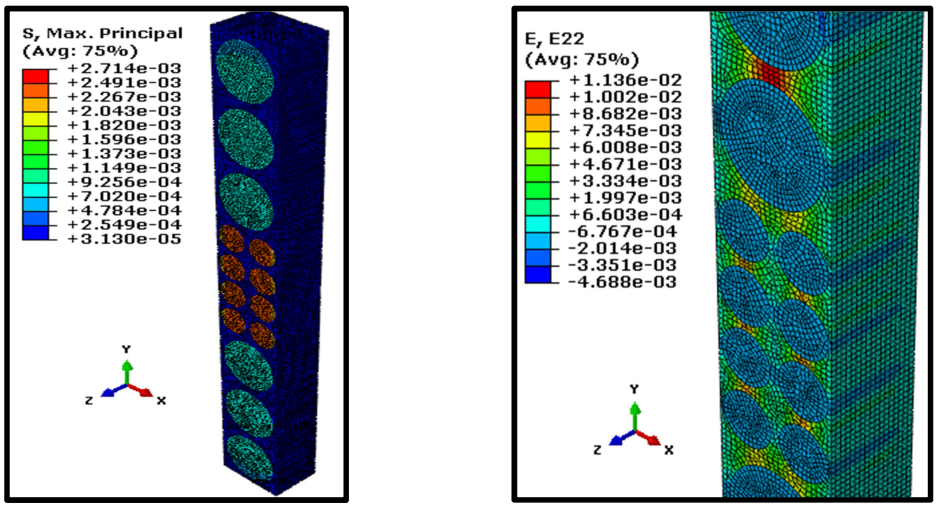


Figure 12: SEM image of hybrid #R depicting matrix cracking and propagation under tensile loading.

Figure 13(a) shows the contours of maximum principal stress for hybrid #S under tensile loading, where the carbon fibre held more stress than its counterpart at the outer GFRP layup. Figure 13(b) depicts significant magnitude of strain in the *Y* global direction at intralaminar level of CFRP and GFRP, which in practical perspective will induce the occurrence of delamination in experimental findings. This is observed clearly in the SEM image shown in Figure 14, whereby discontinuities and delaminations are observed within GFRP layup as result of stress concentration at intralayer of GFRP.



(a) **(b)**

Figure 13: Contours of (a) maximum principal stress and (b) strain ($\mathcal{E}_{22}$) in the *Y* global direction for hybrid #S under tensile loading.

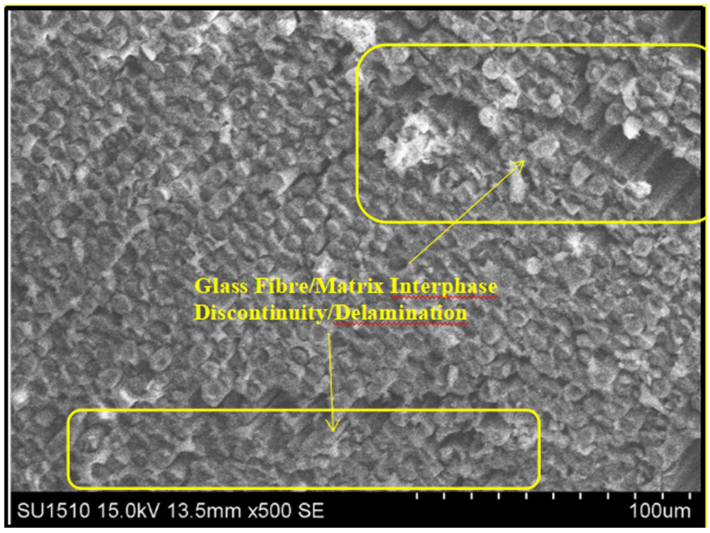
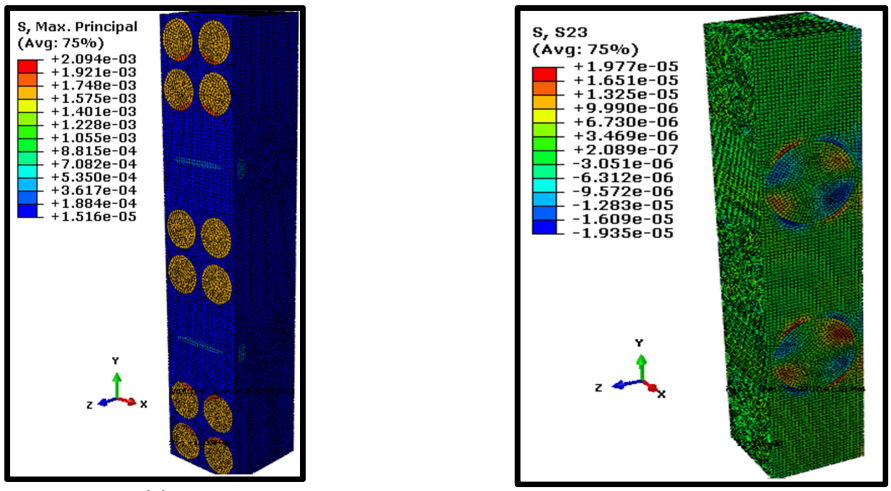


Figure 14: Illustration of hybrid #S with interphase of glass fibre / matrix discontinuities and delaminations.

Figure 15(a) shows the contours of maximum principal stress for hybrid #W under tensile loading, where the carbon fibre holds more stress than 90° glass fibre. Figure 15(b) exhibits the contours of shear stress (σ_{23}) in the *Y* global direction under tensile loading. It displays a high indication of shear stress intensity at the interface of GFRP 90º, which is then observed to experience interlaminar delamination in experimental testing.



(a) (b)

Figure 15: Contours of (a) maximum principal stress and (b) shear stress (σ_{23}) in the *Y* global direction for hybrid #W under tensile loading.

Figures 16, 17 and 18 plot the stress against strain for hybrids #R, #S and #W respectively at the RVE and macroscale levels. The values of modulus of elasticity correspond to the gradient of the plot, which depicts the macroscale recording higher modulus of elasticity (E_{11}) than RVE. Discrepancy in the computed value for modulus of elasticity for microscale RVE and macroscale is the smallest for hybrid #W, as compared to hybrids #R and #S, whereby it is observed that orientation of 90º GFRP, allowed computation of modulus of elasticity to be influenced predominantly by 0º CFRP at the direction of loading (0º principal direction).

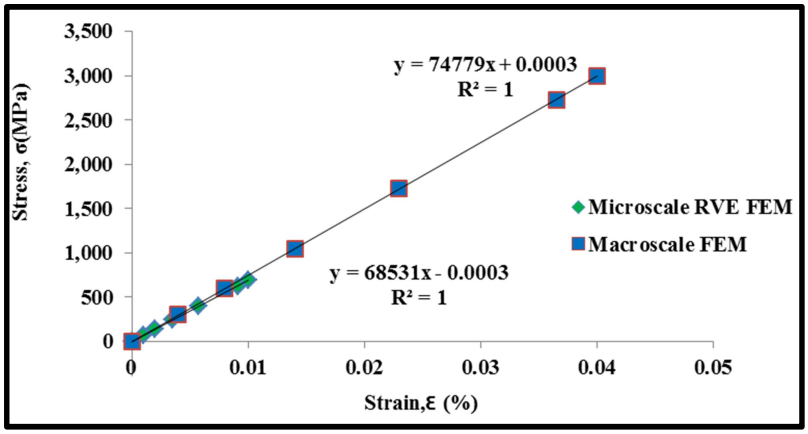


Figure 16: Stress against strain for hybrid #R at RVE and macroscale level under tensile loading.

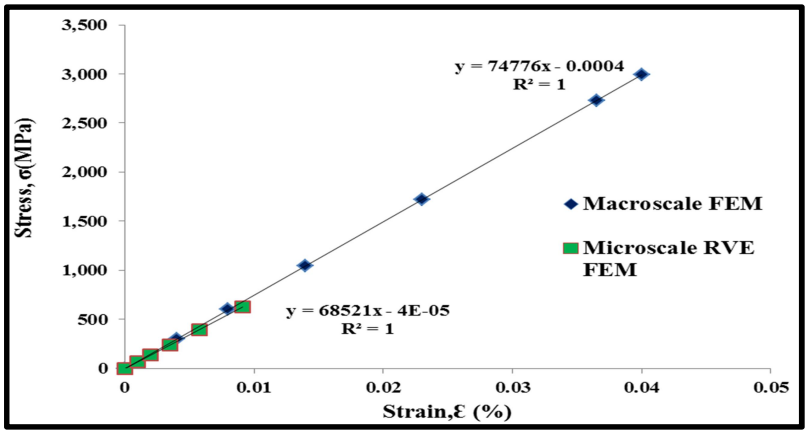


Figure 17: Stress against strain for hybrid #S at RVE and macroscale level under tensile loading.

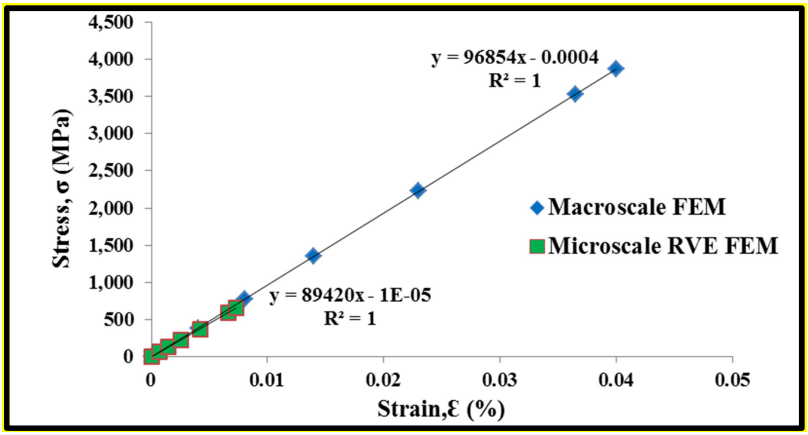


Figure 18: Stress against strain for hybrid #W at RVE and macroscale level under tensile loading.

4. CONCLUSION

In this study, RVE was simulated using FEM and has been proven to be a useful tool in predicting the effective modulus of elasticity of composite / hybrid composites obtained from experimental values. The simulation in static at microlevel, which represented in the RVE squares, shows that RVE can be utilised to represent homogenisation of hybrid FRP to characterise loading behaviour. Hybrid #S simulated under RVE depicts significant principle stress in between the CFRP and GFRP layers, and this is proven to be true with the experimental finding showing delamination between the CFRP / GFRP layers. For hybrid #W, the contours of shear stress were seen to be very significant at the interface of GFRP 90º, which in parallel proves to be similar during experimental testing, where interlaminar delamination is observed. As for hybrid #R, longitudinal strain was found to be very significantly high in the matrix region between the fibres under tensile loading. This was justified from experimental perspective, where matrix cracking and propagation were observed during the failure after tensile loading. Further research could further explore many areas in regards to the application of RVE in predicting strength of macroscale composites as well as an optimisation tool. Failure prediction is an area of study that could be exploited using RVE, whereby prediction of failure or reliability can be determined by incorporating possible failure criteria for each constituent of materials. The contours of shear strain as observed in RVE at the interface of fibre / matrix, as well as normal stress at the interface of fibre / matrix and normal stress at matrix rich region between the layers of CFRP and GFRP provide further assessment and research on the region especially in the study of delamination.

REFERENCES

Abbassi, F. Gherissi, A., Zghal, A., Mistou, S. & Alexis, J. (2011). Micro-scale modeling of carbon-fiber reinforced thermoplastic materials. *Appl. Mech. Mater.*, **146**: 1–11.

Aboudi, J., Arnold, S.M. & Bednarcyk, B.A. (2013). *Micromechanics of Composite Materials: A Generalized Multiscale Analysis Approach*. Elsevier, Amsterdam.

Alfaro Cid, M.V, Suiker, A.S.J & Borst, R. de (2010). Transverse failure behavior of fiber epoxy systems. *J. Compos. Mater,* **44:** 1493–1515.

Babu K.S., Rao, K.M, Raju, V.R.C, Murthy, V.B.K & Kumar, M.S.R.N. (2008). Micromechanical analysis of FRP hybrid composite lamina for in-plane transverse loading. *Indian J. Eng. Mater. Sci*, **15:** 382–390.

Banerjee, S. & Sankar, B.V. (2014). Mechanical properties of hybrid composites using finite element method based micromechanics. *Compos. Part B Eng*, **58:**318–327.

Bhaskara, S., Devireddy, R. & Biswas, S. (2014). Effect of fiber geometry and representative volume element on elastic and thermal properties of unidirectional fiber-reinforced composites. *J. Compos*, **201:** 1-12.

Cai, D., Zhou, G., Wang, X., Li., C. & Deng, J. (2017). Experimental investigation on mechanical properties of unidirectional and woven fabric glass/epoxy composites under off-axis tensile loading. *Polym. Test*., **58**: 142–152.

DOD (U.S. Department of Defense) (2002), *Composite Material Handbook, Volume 3:. Polymer Matrix Composites: Material Usage, Design and Analysis*, US Department of Defence (DOD), Virginia, US.

Drathi M.R & Ghosh A (2015). Multiscale modeling of polymer-matrix composites. *Comput. Mater. Sci*, **99:** 62–66.

Harris, B. (1999). *Engineering Composite Materials, Second Edition*. IOM Communications, London.

Jiang, W.G, Zhong, R.Z, Qin, Q.H. & Tong, Y.G. (2014). Homogenized finite element analysis on effective elastoplastic mechanical behaviors of composite with imperfect interfaces. *Int. J. Mol. Sci*, **15:** 23389–23407.

Kanit, T, Forest, S., Galliet, I. , Mounoury, V. & Jeulin, D. (2003). Determination of the size of the representative volume element for random composites: Statistical and numerical approach. *Int. J. Solids Struct.,* **40:** 3647–3679

Karkkainen, R.L. & Sankar, B.V. (2006). A direct micromechanics method for analysis of failure initiation of plain weave textile composites. *Comp. Sci. Tech.*, **66:**137-150.

Kutz, M. (2005). *Handbook of Materials Selection*. Wiley, New York.

Omairey, S.L, Dunning, P.D. & Sriramula, S. (2018). Development of an ABAQUS plugin tool for periodic RVE homogenisation. *Eng. With Comput.*, **35:** 567–577.

Prabu, S.B, Karunamoorthy, L., & Kandasami, G.S. (2004).A finite element analysis study of micromechanical interfacial characteristics of metal matrix composites. *J. Mater. Process. Technol.*, **154**: 992–997.

Qingping, S., Zhaoxu, M., Guowei, Z., Shih, P.L., Hongtae, K., Sinan, K., Haidig, G. & Xuming, S. (2018). Multi-scale computational analysis of unidirectional carbon fiber reinforced polymer composites under various loading conditions. *Compos. Struct*., **196:** 30–43.

Silva, L.J., Panzera, T.H., Christoforo, A.L., Rubio, J.C.C. & Scarpa, F. (2012). Micromechanical analysis of hybrid composites reinforced with unidirectional natural fibres, silica microparticles and maleic anhydride. *Mater. Res.*, **15:** 1003-1012.

Tan, P., Tong, L. & Steven, G.P. (2000). Behavior of 3D orthogonal woven CFRP composites . Part II . FEA and analytical modelling approaches. *Compos. Part A: Appl. Sci. Manuf*. **31:**273–281

Valavala, P.K, Odegard, G.M., & Aifantis. E.C. (2009). Influence of representative volume element size on. *Mod. Sim. Mat. Sci. Eng.*, **17:**1–15.

Wang, W., Dai, Y., Zhang, C., Gao, X. & Zhao, M. (2016), "Micromechanical Modeling of Fiber-Reinforced Composites with Statistically Equivalent Random Fiber Distribution," *Materials (Basel)*., **9:** 1–14.

Zhang, J., Chaisombat, K.., He, S. & Wang, C.H. (2012). Hybrid composite laminates reinforced with glass/carbon woven fabrics for lightweight load bearing structures. *Mater. Des.*, **36**: 75–80.

Zhou, H.L., Zhang, W. & Zheng, Y. (2017). Tensile response of carbon-aramid hybrid 3D braided composites. *Mater. Des*, **116:** 246–252.

SHOCKWAVE BOUNDARY LAYER INTERACTION AT VARIOUS MACH NUMBERS AND ANGLES OF ATTACKS

Nurfathin Zahrolayali[1], Mohd Rashdan Saad[1], Azam Che Idris[2] & Mohd Rosdzimin Abdul Rahman[1*]

[1]Department of Mechanical, Faculty of Engineering, National Defence Universiti of Malaysia (UPNM), Malaysia
[2]Faculty of Integrated Technologies, Universiti Brunei Darussalam, Negara Brunei Darussalam

[*]Email: rosdzimin@gmail.com

ABSTRACT

This research examines the shockwave characteristics at a double ramp inlet and its performances at different angles of attacks (AoAs) with various Mach numbers. The k-omega (k–ω) shear stress transport (SST) turbulence model was used over a two-dimensional double ramp inlet at free-stream Mach numbers of 4, 5 and 6 with various AoAs in the range of -10 to 10°. The experimental data was used to validate the numerical results. Analysis of the internal shock structures showed that at a Mach number lower or greater than the intended Mach number, the shock waves' impact on the cowl inlet was considerably influenced. Moreover, the Mach number and AoA had major impact on overall pressure, kinetic energy and compression process efficiencies, whereby kinetic energy and compression process efficiencies improve when AoA is increased. An increase in AoA also shifts the shock wave inside the isolator, while decrease in AoA expels the shockwave from the isolator at the design Mach number. It can be concluded that the interaction of the shockwave with the cowl would be affected by various AoAs.

Keywords: *Shockwave boundary layer interaction (SWBLI); supersonic; hypersonic; angle of attack (AoA); flow control.*

1. INTRODUCTION

Significant progress in the design of high speed vehicles for the defence industry has been achieved over the past several decades. However, one of the critical design issues that remains unresolved is an efficient propulsion system. The most important component in high speed vehicle propulsion systems is a high speed inlet. A function of a high speed inlet is to decelerate the supersonic / hypersonic incoming flows to subsonic flows so as to provide the required air mass flow rate at high total pressure recovery and minimal distortion for the combustion process. A mixed compression inlet is typically selected to overcome the problem. An important phenomenon that impacts on mixed compression is shock wave boundary layer interactions (SWBLI). SWBLI creates boundary layer thickening, detachment and high wall heat flux, which lead to decrease in efficiency of the propusion system (Huang *et al.*, 2020).

Studies on supersonic and hypersonic flow have been conducted to investigate the characteristics of SWBLI. Zhang *et al*. (2015) designed a bump inside the inlet isolator to minimise cowl shock caused by the SWBLI in order to enhance hypersonic inlet performance. They suggested that cowl shocks must consistently impact the bump convex part for optimum control of hypersonic inlet performance. Moreover, Zhang *et al*. (2017) studied a fluidically changing hypersonic inlet with a fixed model and found that the fluidically regulated intake performed well at low operational Mach numbers. This finding improves the hypersonic propulsion system's acceleration performance. In addition, Im & Do (2018) observed that a local unfavourable pressure gradient near the inlet lip was due to flight movements, such as when the vehicle is in an uneven pitching motion. They also found that the impinging shock location in the inlet can be moved by vehicle tilting, resulting in an unanticipated

SWBLI and a wide flow separation zone near the inlet lip. This unwanted phenomenon needs to be eliminated at the supersonic or hypersonic flow inlet because it reduces the inlet performance. A study involved in a free stream of Mach 6 with 20 and 30% obstruction struts was explored experimentally by Saravanan *et al.* (2021). They demonstrated that opening the cowl below 90% speeds up the unstart procedure. Moreover, flow separation caused by SWBLI has been proven to play a critical role in the unstart process.

Studies on hypersonic inlets found that internal compression was increased by adding sidewall compression (Hohn & Gulhan, 2017; Huang *et al*., 2017). However, it is also shown that combining a double ramp intake with external sidewall compression is inadequate for increasing inlet compression capacities as it causes significant separation. Reducing separation can be obtained using a splitter in the internal contraction of the hypersonic inlet (Xie *et al*., 2018). Recently, Devaraj *et al*. (2020, 2021) used a flap at the intake's end to evaluate the influence of back pressure on the initial features of a hypersonic inlet at Mach 6. They observed shock train oscillations that resulted in unstarts at higher flap angles.

An effect of the shock train on the variable angle of attacks (AoAs) in a scramjet inlet-isolator with constant upstream and downstream conditions was investigated by Li (2022). This study reported that the oscillations of the first separation shock do not necessarily improve with increased excitation magnitude. Instead, the SWBLI's oscillation zone enlarged and internal pressure increased as the amplitude increased. The influence of AoA change on the quasi-steady motion characteristics of the shock train leading edge was examined numerically by Xu *et al*. (2017) and Guo *et al*. (2017). The simulation findings showed that the shock train's motion had a leaping characteristic, mostly driven by the strength of the local flow separation shifting. The background wave's reflection spots migrated downstream as AoA decreased, and one of them approached the shock train's separation zone. Adaptive control of a hypersonic flight vehicle based on AoA and partial loss of efficiency has also been studied by Liu *et al*. (2018). Their findings showed that whatever the AoA, the necessary performance can be ensured. In addition, the influence of AoA on the performance of a supersonic inlet was examined by Askari *et al*. (2019) and Ambe Verma *et al*. (2021). They discovered that when AoA increases, the pressure increases owing to variations in shock wave strength. The shock wave strength over the lower section of the body intensifies as AoA increases, whereas the shock wave strength over the upper portion of the body decreases.

Following the previous studies, this study aims to investigate characteristics of the SWBLI at the supersonic and hypersonic inlets. This research is focused on performance at various AoAs ranging from -10 to 10° with varying Mach numbers in the supersonic and hypersonic categories. The simulation was carried out on a two-dimensional double ramp inlet at Mach values of 4, 5, and 6 in free-stream mode. The numerical procedure was validated with published experimental data and the SWBLI was examined. The findings of this study are important to improve the knowledge of supersonic and hypersonic inlet flow characteristics.

2. PERFORMANCE OF INLET-ISOLATOR

The overall pressure ratio and kinetic energy efficiency of an inlet-isolator were used to classify its performance. Both performance indicators were based on a quasi-one-dimensional flow tube across the intake, as shown in Figure 1. Station 0 is the collected free-stream before compression. Station 3 links the inlet-isolator to the combustor and is located downstream of the internal compression area. Therefore, all performance estimations were based on the flow conditions at Station 3.

2.1 Total Pressure Efficiency

Total pressure efficiency, π_c is defined as the ratio of backpressure at Station 3, P_{t3} to Station 0, P_{t0}. It represents the total pressure loss due to the compression process (Idris *et al*., 2014). The shock waves

boundary layer interaction and, to a lesser extent, the high losses in the stationary flow caused by slip conditions on the wall's surface both have significant impact on the total pressure efficiency. The following equation is used to calculate π_c:

$$\pi_C = \frac{P_{t3}}{P_{t0}} \tag{1}$$

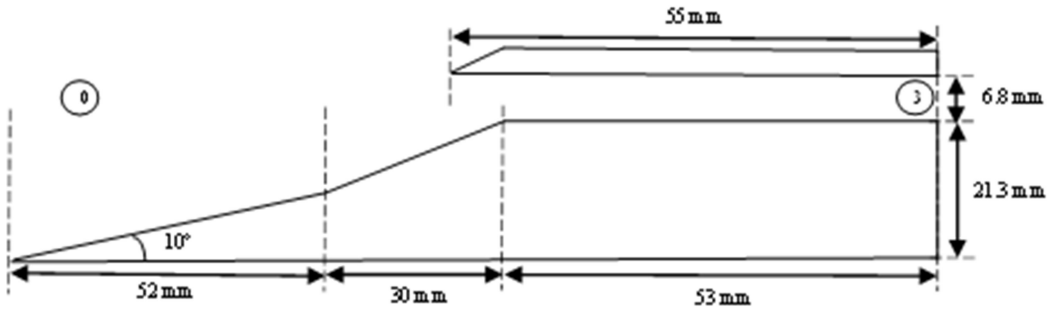


Figure 1: Generic scramjet inlet-isolator dimension.

2.2 Kinetic Energy Efficiency

Kinetic energy efficiency, $\eta_{KE\,(ad)}$ is defined as the ratio of the kinetic energy contained in the flow at Station 3 to the kinetic energy initially accessible in the free-flowing flow if it were extended isentropically to free-stream pressure (Idris *et al*., 2014). The function of a scramjet engine is to create thrust by boosting the nozzle flow at hypersonic speeds when free flow already includes a considerable amount of kinetic energy. The thrust created would be less than optimum if a considerable quantity of kinetic energy is consumed during the compression process. The following equation is used to determine $\eta_{KE\,(ad)}$:

$$\eta_{KE(ad)} = 1 - \left(\frac{2}{\gamma-1}\right)\left(\frac{1}{M_0^2}\right)\left[\left(\frac{T_x}{T_0}\right) - 1\right] \tag{2}$$

where:

$$T_x = T_3\left(\frac{P_o}{P_{t3}}\right)^{\frac{\gamma-1}{\gamma}} \tag{3}$$

2.3 Compression Process Efficiency

The total efficiency of the adiabatic compression process, $\eta_{C(ad)}$ is determined by the amount of energy consumed by the compression process. This value is derived by comparing the total energy contained at Station 3 to the starting energy in the volume of air collected in free flow (Idris *et al*., 2014):

$$\eta_{C(ad)} = 1 - \frac{(\gamma-1)M_0^2}{2}\left(\frac{1-\eta_{KE(ad)}}{\frac{T_3}{T_0}-1}\right) \tag{4}$$

3. NUMERICAL ANALYSIS

In this study, the free-stream Mach numbers of 4, 5, and 6 at inlet-isolator configurations were investigated. The design baseline was at Mach number 5 with the angle of attack (AoA) = 0. At Mach = 5, the stagnation pressure and inlet temperatures were P_∞ = 0.65 MPa and T_∞ = 375 K respectively.

These values correspond to Reynolds number of 13.2×10^{6} m^{-1}. At Mach numbers 4 and 6, the stagnation pressure and inlet temperature were adjusted accordingly.

ANSYS Fluent was utilised for this study. The viscous model for turbulence used in this study was the k-omega (k–ω) shear stress transport (SST) model. The viscosity ratio of turbulence was set as unity. The Courant–Friedrichs–Levy (CFL) number was initially set to 0.5 and subsequently increased by the same amount at every 1,000 iterations to maintain stability. The initialisation conditions for turbulence computation by the solver were inviscid solutions for each parametric case. A pressure inlet, pressure far-field, two pressure outlets, constant temperature walls and symmetry defined the computing domain, as shown in Figure 2. The symmetry area between the pressure inlet border and the first compression ramp wall assisted in keeping the iterations steady. The properties at the two pressure outputs were estimated with free flow in consideration, expecting that the flow will escape the isolator and expand to free-stream conditions.

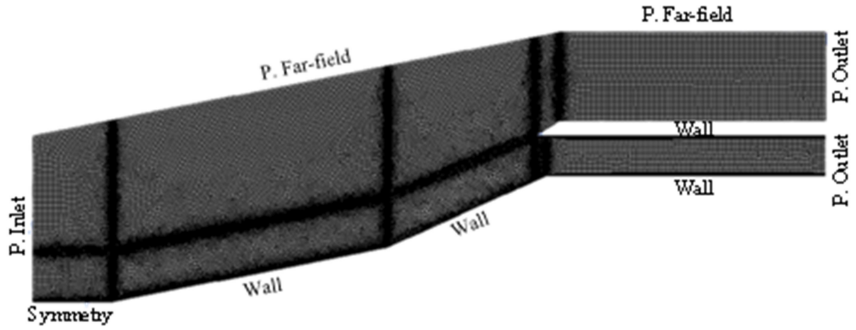


Figure 2: Mesh for the baseline case.

The baseline (AoA = 0°), as well as AoA = -10°, -8°, -6°, -4°, 4°, 6°, 8° and 10° used quadrilateral mesh cells with a higher density grid clustered around a significant flow turning zone.

In order to lower the computational cost, mesh size optimisation was implemented. Figure 3 shows that the element size of 0.0005 m fits with the smallest element size of 0.0002 m. Thus, element size of 0.0005 m was used for all the simulation cases in order to lower the computational cost.

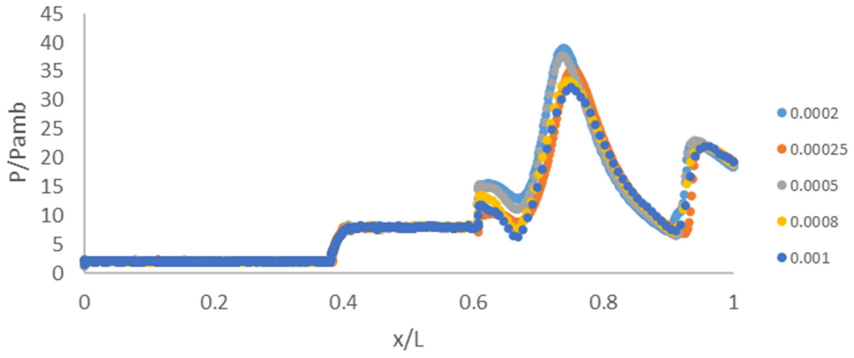


Figure 3: Grid independent test.

4. RESULTS AND DISCUSSION

4.1 Validation

Figure 4 shows distributions of the mean wall pressure along the inlet-isolator surface. The present numerical results are consistent with the experimental data reported by Idris *et al*. (2014).

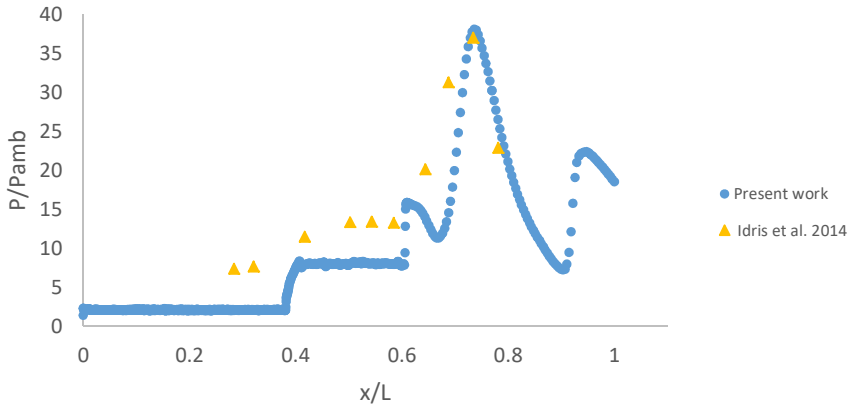


Figure 4: Validation of the simulation results with experimental data.

4.2 Centreline Pressure Profile

Figure 5 shows the pressure profile along the model's centreline, which perfectly demonstrates the shock pattern inside the ramp section. The pressure profiles offered critical signals on the distortion level since the background wave within the isolator formed as a result of the cowl tip shock impingement downstream of the shoulder (Van Wie & Aultt, 1996). There is indeed a lot of consistency between all of the AoA cases. On the ramp, a precise difference in the pressure measurements may be noticed. When compared to AoA = 4°, 6° and 8°, the pressure profile on the ramp of AoA = 10° has a significantly larger magnitude. This is as the pressure profile increases as AoA increases. Meanwhile, when the AoA declines from 0° to -4°, so does the pressure profile. For the AoA = -6°, -8°, and -10° cases, a similar pattern of pressure profile declines is seen. The amplitude of the pressure profile on the ramp was dramatically reduced when AoA = -10° was used.

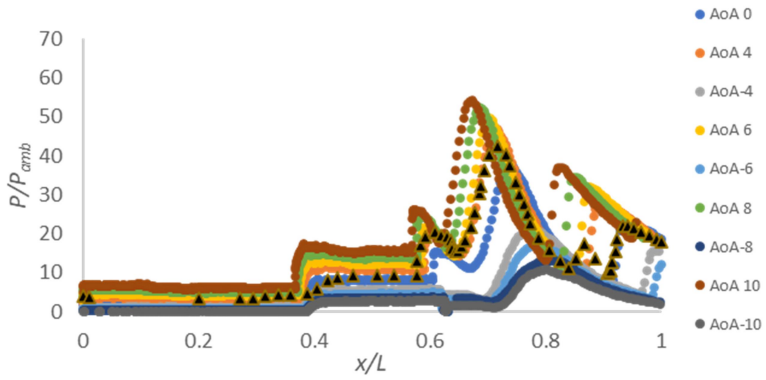


Figure 5: Static pressure normalised along the centreline of the scramjet inlet model for various cases along the ramp surfaces.

Figure 6 shows the pressure profile along the model's centreline, which perfectly illustrates the shock pattern inside the isolator section. The numerical study also shows that the separation onset positions at the isolator entry are somewhat further upstream. There are three pressure peaks inside the isolator section: (1) the shoulder separation bubble's re-attachment point; (2) the separation impingement locations; and (3) the third separation bubble's re-attachment shock.

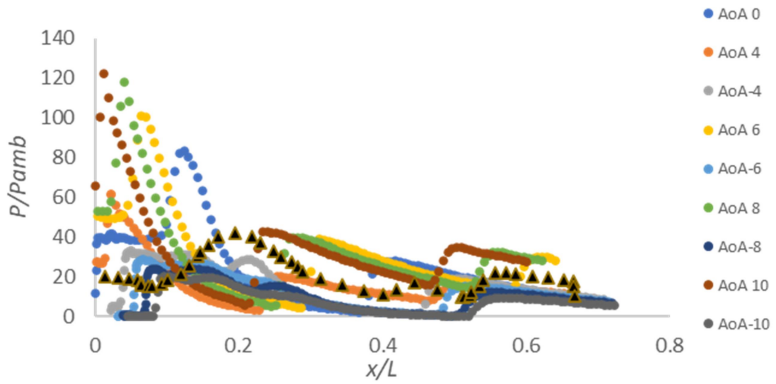


Figure 6: Static pressure normalised along the centreline of the scramjet inlet model for various cases along the isolator surfaces.

4.3 Internal Shock Structures

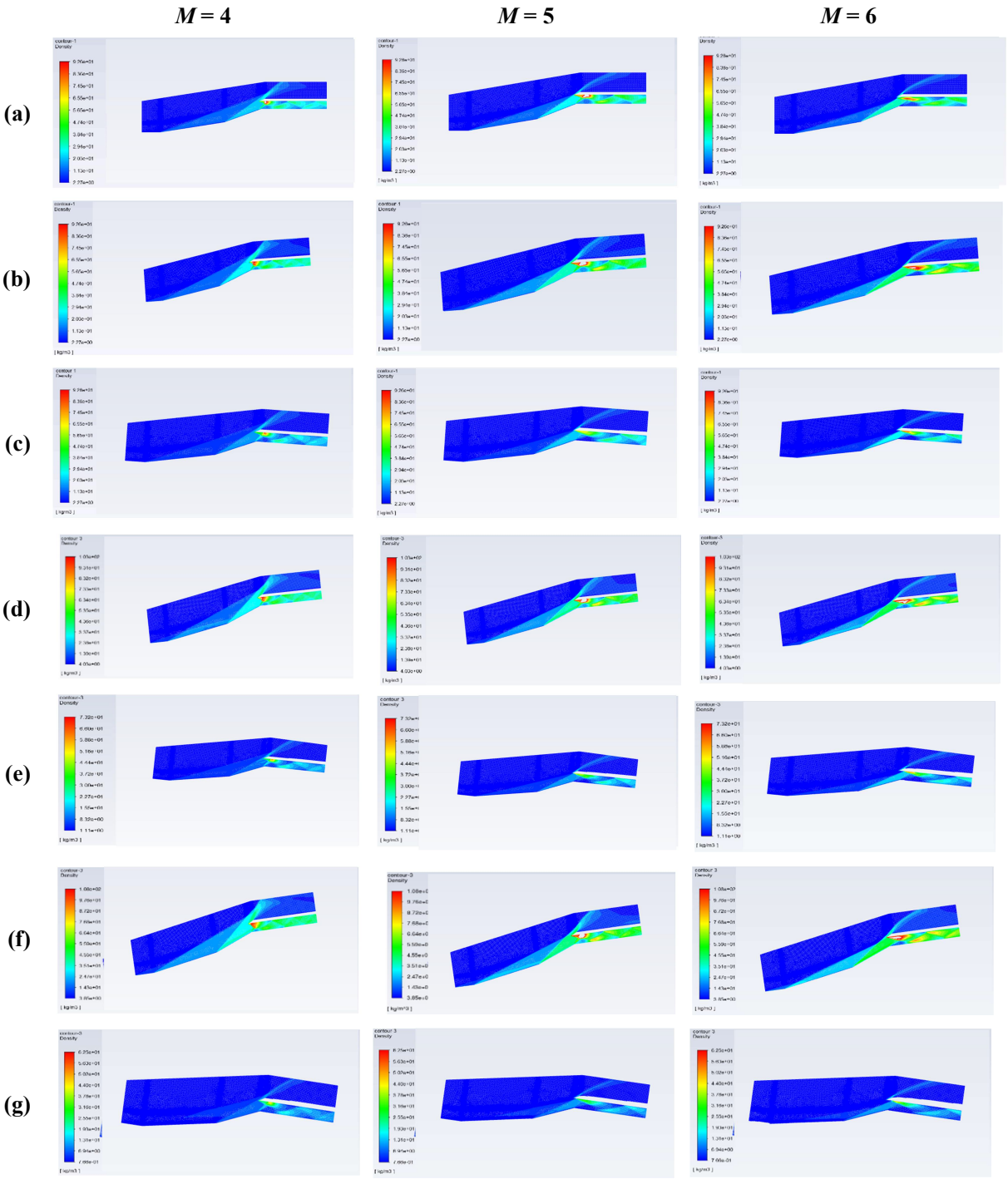
Figure 7 depicts the density contours at all AoAs, concentrating on just the throat and isolator exit. The design based on Mach 5 demonstrates that the shocks from the external compression ramp impact hit the cowl tip in the AoA = 0 case, as shown in Figure 7(a), allowing the intake to catch a nearby mass flow rate.

The minimal boundary layer motion of the compression ramp can be seen in this analysis. The major elements of the internal flow field, as depicted in Figure 7(a), are a succession of SWBLI between both the floor and ceiling of the isolator (Tan *et al*., 2012). The cowl shock impinges on the inlet's shoulder, causing separation (Tan *et al*., 2009). Both the created and reflected separation shocks impacted the isolator's ceiling. A small separation bubble is created by the reflected shock, which caused a modest separation shock and a second reflected shock (Bachchan & Hillier, 2004). The shocks from the external compression ramp were extremely near to reaching the cowl when the Mach number dropped to 4. At Mach 6, the shocks from the external compression ramp reflected the shock inside the isolator, resulting in a smaller separation bubble. The shocks got stronger because of the acceleration of flow aftershocks and the higher Mach number. A larger separation bubble was created by the reflected shock.

When an aircraft travels at different Mach values, different AoAs are employed. The cowl tip separation was reduced when the AoA on the windward side was increased, as seen in Figures 7(b), 7(d), 7(f) and 7(h) for AoA = 4, 6, 8, and 10°. The compression ramp generated shocks that travelled upstream, preventing the cowl tip from separating. In the AoA = 4° case, Figure 7(b) displays an intriguing shock production. The external compression ramp shocks projected the shocks very close to the cowl at Mach 5. A Mach stem occurred at the intersection of the cowl tip and shoulder separation shocks, causing a significant loss of stagnation pressure. For the AoA = 4° case, this flow turning angle was unsustainable, causing the flow into the isolator to decelerate to subsonic speeds through a Mach stem. As shown by the oblique shock reflections downstream, this subsonic pocket was limited and did not expand towards the isolator outlet. Shockwaves from the external compression ramp did not strike the cowl at Mach 4, but they were reflected within the isolator at Mach 6. For the AoA = 6° case depicted in Figure 7(d), the external compression ramp reflected the shock expelled from the impacted cowl at Mach 4 and 5, whereas the shocks hit the cowl at Mach 6. For Mach 4, 5, and 6, all of the external compression ramps reflected the shock without hitting the cowl in the AoA = 8 and 10° cases depicted in Figures 7(f) and 7(h).

The cowl tip separation increased in size when the AoA is changed from 0 to -4, -6, -8, and -10°, as illustrated in Figures 7(c), 7(e), 7(g) and 7(i) for AoA = -4, -6, -8, and -10° respectively. Figure 7(c) displays the case with AoA = -4°. The shocks from the external compression ramp reflected the shock

hit the cowl at Mach 5. At Mach 4, the shockwaves left the external compression ramp without hitting the cowl, but at Mach 6, the shocks were reflected within the isolator. For the AoA = -6, -8, and -10° cases, shown in Figures 7(e), 7(g) and 7(i), the external compression ramp reflected shocks inside the isolator without hitting the cowl at Mach 5 and 6, while at Mach 4, the shock waves reflected hit the cowl.



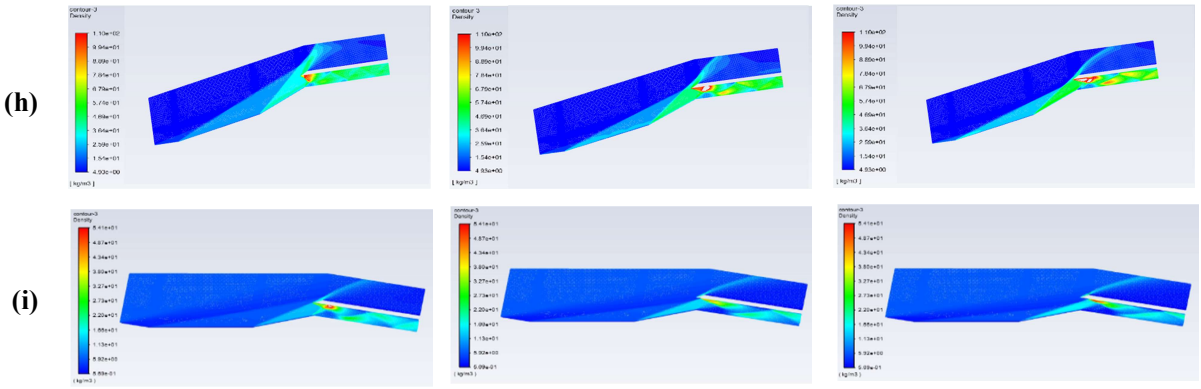


Figure 7: Density contours for various cases of AoA (a) 0° (b) 4° (c) -4° (d) 6° (e) -6° (f) 8° (g) -8° (h) 10° (i) -10°.

4.4 Flow Properties and Inlet-Isolator Performance

Figure 8 shows the total pressure efficiency at various AoAs between Mach 4 and 6. It shows that 22 to 53% of total pressure was produced in the baseline case. When the AoA was increased to 4°, the total pressure increased from 62 to 100%, and at AoA = 6°, the total pressure increased from 63 to 165%. Moreover, at AoA=8 and 10°, the total pressure decreased between 44 and 81%. In the meantime, the total pressure produced on the leeward side at AoA of -4° ranged from 12 to 59%. The highest peak was seen at AoA = -6°, with values ranging from 143 to 171%. As the AoA approached AoA= -8 and -10°, the overall pressure fell.

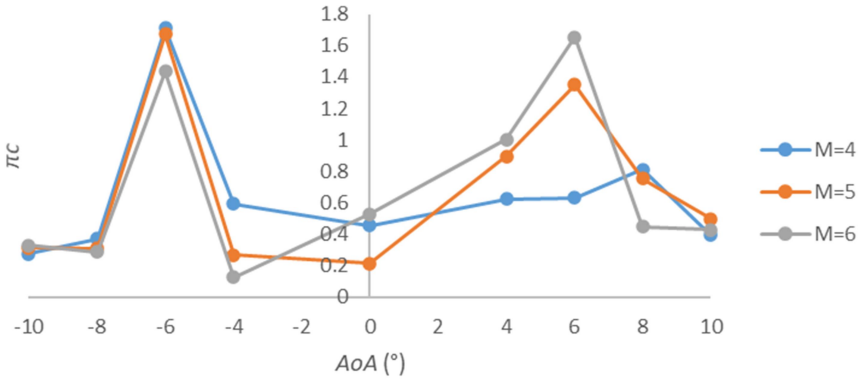


Figure 8: Total pressure efficiency at different AoA cases.

By analysing the graph of kinetic energy efficiency in Figure 9, all cases reveal a similar trend in which kinetic energy increases as the AoA increases. The kinetic energy generated ranged from 0.75 to 0.79 at AoA = 0°. The kinetic energy increased when the windward side was applied, from AoA = 4 to 10°. However, when the leeward side was utilised, the kinetic energy decreased. Furthermore, Figure 10 shows a similar pattern in terms of compression process efficiency. As AoA increased, the the compression process efficiency improved. In the baseline case, compression process efficiency ranged from 0.6 to 0.69. The compression process efficiency increased from 0.69 to 0.77 when the AoA was between 4 and 10°. On the leeward side, the compression process efficiency ranged from 0.45 to 0.64.

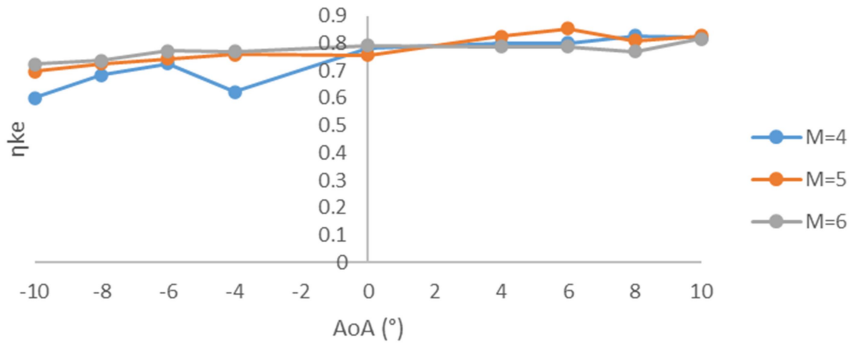


Figure 9: Kinetic energy efficiency at different AoA cases.

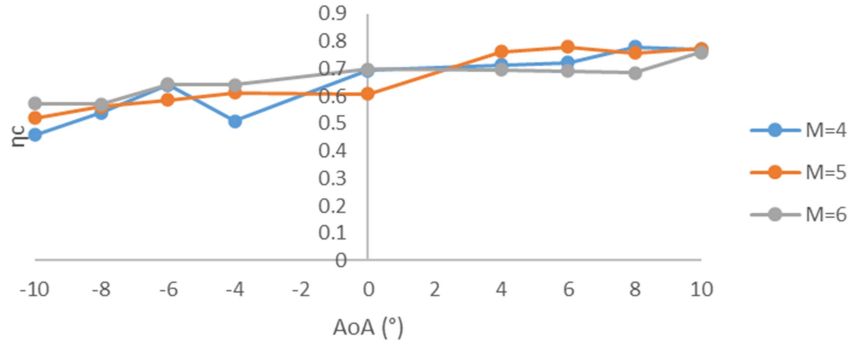


Figure 10: Compression process efficiency at different AoA cases.

The performance of the baseline scenario is moderate, if not unacceptable, with just 22% of the total pressure. It has weak kinetic energy efficiency of 0.75, which translates to compression process efficiency of 0.61. The greatest compression process efficiency, according to Heiser & Pratt (1994), was around 0.9 for a conventional three compression shock intake. The existence of the three separation bubbles inside the isolator section was assumed to be the explanation for such poor performance. They caused the flow to stagnate and lose kinetic energy.

When a modest AoA was used, the performance of the free-stream flow improved substantially. For example, in the AoA = 4° case, the isolator exit static pressure developed to roughly two times the free-stream. If the compression were increased, the propulsion unit would produce more thrust. In this situation, the compression of the inlet-greater isolator did not come at the expense of critical total pressure and flow kinetic energy, which both improved as compared to the AoA = 0° case. Due to the higher total pressure and kinetic energy efficiency, the compression system efficiency may have increased significantly, approaching the limitations for such inlets.

The performance was somewhat enhanced by increasing the AoA by 4° further. In comparison to the AoA = 4° case, this results in a somewhat improved compression system efficiency, but at the consequence of a significant decrease in stagnation pressure and kinetic energy. The overall compression enhances the static pressure leaving the isolator at around 50 kPa, which is optimal for a wind tunnel-scale scramjet engine's high cycle efficiency (Smart, 2012).

5. CONCLUSION

Numerical studies were conducted in a 2D inlet isolator to investigate the SWBLI at various AoAs. It was observed that various AoAs would affect the SWBLI at the cowl at Mach number lower or higher than design Mach number of 5. The location of the shock wave hitting the cowl inlet was significantly affected. Moreover, at design Mach number, Mach 5, increment in the AoA would move the shockwave inside the isolator, while decrease in AoA would expel the shockwave from the isolator. It was found that the total pressure efficiency, kinetic energy efficiency and compression process efficiency were significantly affected by Mach number and AoA whereby an increment in AoA increases the kinetic energy efficiency and compression process efficiency.

ACKNOWLEDGEMENT

The authors would like to thank the Universiti Pertahanan Nasional Malaysia (UPNM) for giving full support to this research work. This material is based upon work supported by the Air Force Office of Scientific Research under award number FA2386-21-1-4016.

NOMENCLATURE

M	Mach number
P_∞	Stagnation pressure (Pa)
P_O	Pressure at Station 0 (Pa)
P_{t3}	Pressure at Station 3 (Pa)
T_3	Temperature at Station 3 (K)
T_0	Temperature at Station 0 (K)
T_∞	Stagnation temperature (K)
γ	Specific heat ratio
π_C	Total pressure efficiency
$\eta_{KE\ (ad)}$	Kinetic energy efficiency
$\eta_{C(ad)}$	Compression process efficiency

REFERENCES

Ambe Verma, K., Murari Pandey, K., Ray, M. & Kumar Sharma, K. (2021). Effect of transverse fuel injection system on combustion efficiency in scramjet combustor. *Energy*, **218**:119511.

Askari, R., Soltani, M.R., Mostoufi, K., Fard, A.K. & Abedi, M. (2019). Angle of attack investigations on the performance of a diverterless supersonic inlet. *J. Appl. Fluid Mech.*, **12**: 2017–2030.

Bachchan, N. & Hillier, R. (2004). Hypersonic inlet flow analysis at off-design conditions. *AIAA Appl. Aerodyn. Conf.*, **2**: 1180–1192.

Devaraj, M.K.K., Jutur, P., Rao, S.M.V., Jagadeesh, G. & Anavardham, G.T.K. (2020). Experimental investigation of unstart dynamics driven by subsonic spillage in a hypersonic scramjet intake at Mach 6. *Phys. Fluids*, **32**:026103.

Devaraj, M.K.K., Jutur, P., Rao, S.M.V., Jagadeesh, G. & Anavardham, G.T.K. (2021). Investigation of local unstart in a hypersonic scramjet intake at a Mach number of 6. *Aerosp. Sci. Technol.*, **115**:106789.

Guo, S. T., Li, Z. F., Gao, W. Z. & Yang, J. M. (2017). Analogy between effects of attack

angle and mach number on inlet starting. *Tuijin Jishu/J. Propuls. Technol.*, **38**: 983–991.
Guo, Y., Xu, B., Han, W., Li, S., Wang, Y. & Zhang, Y. (2020). Robust adaptive control of hypersonic flight vehicles with asymmetric AOA constraint. *Sci. China Inf. Sci.*, **63**: 212203.
Heiser, W.H. & Pratt, D.T. (1994). *Chapter 5: Compression Systems or Components. In Hypersonic Airbreathing Propulsion*. AIAA Education Series.
Hohn, O. M. & Gulhan, A. (2017). Experimental investigation of sidewall compression and internal contraction in a scramjet inlet. *J. Propul. Power*, **33**:501–513.
Huang, R., Li, Z., Zhan, D., Yang, J., Yu, A. & Wu, Y. (2017). Measurements of the streamwise vortices in a hypersonic inward turning inlet. *21st AIAA Int. Space Planes Hypersonics Technol. Conf. Hypersonics 2017*, 1-9 March 2017..
Idris, A. C., Saad, M. R., Zare-Behtash, H. & Kontis, K. (2014). Luminescent measurement systems for the investigation of a scramjet inlet-isolator. *Sensors-Basel*, **14**: 6606–6632.
I'm, S. Kyun, & Do, H. (2018). Unstart phenomena induced by flow choking in scramjet inlet-isolators. *Prog. Aerosp. Sci*., **97**:1–21.
Li, N. (2022). Response of shock train to the fluctuating angle of attack in a scramjet inlet-isolator. *Acta Astronaut*, **190**:430–443.
Liu, J., An, H., Gao, Y., Wang, C. & Wu, L. (2018). Adaptive control of hypersonic flight vehicles with limited angle-of-attack. *IEEE-Asme T. Mech.*, **23**: 883–894.
Saravanan, R., Desikan, S.L.N., Francise, K.J. & Kalimuthu, R. (2021). Experimental investigation of start/unstart process during hypersonic intake at Mach 6 and its control. *Aerosp. Sci. Technol*, **113**: 106688.
Smart, M. K. (2012). How much compression should a scramjet inlet do? *AIAA J.*, **50**: 610–619.
Tan, H.J., Sun, S. & Huang, H.X. (2012). The behaviour of shock trains in a hypersonic inlet/isolator model with complex background waves. *Exp. Fluids*, **53**: 1647–1661.
Tan, H.J., Sun, S. & Yin, Z.L. (2009). Oscillatory flows of rectangular hypersonic inlet unstart caused by downstream mass-flow choking. *J. Propul. Power*, **25**:138–147.
Van Wie, D.M. & Aultt, D.A. (1996). Internal flowfield characteristics of a scramjet inlet at Mach 10. *J. Propul. Power*, **12**:158–164.
Xie, W.Z., Wu, Z.M., Yu, A.Y. & Guo, S. (2018). Control of severe shock wave/boundary layer interactions in hypersonic inlets. *J. Propul. Power*, **34**:614–623.
Xu, B., Shi, Z., Sun, F., & He, W. (2019). Barrier Lyapunov function based learning control of hypersonic flight vehicle with AOA constraint and actuator faults. *IEEE Trans. Cybern.*, **49**:1047–1057.
Xu, K., Chang, J., Zhou, W. & Yu, D. (2017). Mechanism of shock train rapid motion induced by variation of attack angle. *Acta Astronaut* **140**:18–26.
Zhang, Y., Tan, H.J., Sun, S., Chen, H. & Li, C.H. (2017). Experimental and numerical investigation of a fluidically variable hypersonic inlet. *AIAA J.*, **55**:2597–2606.
Zhang, Y., Tan, H.J., Sun, S. & Rao, C.Y. (2015). Control of cowl shock/boundary layer interaction in hypersonic inlets by bump. *AIAA J.*, **52**:3492–3495.

FLIGHT TESTING OF BASELINE MODEL OF VERTICAL TAKE-OFF AND LANDING (VTOL) UNMANNED AERIAL VEHICLE (UAV)

Zulhilmy Sahwee*, Mohd Hariz, Shahrul Ahmad Shah, Nadhiya Liyana Mohd Kamal & Nurhakimah Norhashim

Unmanned Aerial System Research Laboratory, Avionics Section, Malaysian Institute of Aviation Technology, Universiti Kuala Lumpur, Malaysia

*Email: zulhilmy@unikl.edu.my

ABSTRACT

Advanced unmanned aerial vehicles (UAVs) have become increasingly popular as low-risk platforms for new technology demonstration and research purposes. This paper presents the conceptual design and flight testing of an electric-powered experimental flying wing UAV, known as PLANK-V. The UAV is a modular type that shares the same wing as the conventional flying wing. Flight tests were conducted to evaluate the UAV's performance and initial control parameter tuning. Due to the experimental nature and avionics complexity of such systems, the flight tests require the support of an experienced pilot and ground operation crew, who monitor and assess the behaviour and performance of the aircraft and its subsystems. Safety aspects are critical in any flight, especially in the initial flight test phase. Before performing any flight tests of either manned or unmanned aircrafts, a comprehensive and well-trained test pilot should use a pre-flight checklist as a required safety document in the flight test plan. This paper aims to present the preliminary flight test result on the PLANK-V, which uses a H-configuration vertical take-off and landing (VTOL) arm for the vertical lifting propulsion system. The preliminary design was addressed from several points of view: a conceptual design was carried out, which emphasises the ease of manufacturing; aerodynamic performances improvement for subsequent iterations; while propulsion system design for vertical and horizontal flight, and mechanical design was addressed in order to produce the prototype. From the maiden-flight test campaign of the PLANK-V, the flight test procedures and experience gained during the flight tests were gathered, and recommendations were put forward. The flight data analysis and feedback from the pilot were also valuable tools for future improvements.

Keywords: Unmanned aerial vehicle (UAV); vertical take-off and landing (VTOL); flight test; experimental design; quadplane.

1. INTRODUCTION

In any aerial vehicle design, the proof of the design needs to be validated through flight tests, which is a very demanding task (Palaia *et al.*, 2019; Dündar *et al.*, 2020; Ewoud, 2022). In this research, flight tests were conducted to observe the flight performance of the initial PLANK-V unmanned aerial vehicle (UAV) prototype platform. This prototype was equipped with a vertical propulsion unit capable of vertical take-off and landing (VTOL). The flight tests were performed to evaluate the performance of the VTOL propulsion system, which was used for the hover flight phase. There were two phases of the flight test, which were essential in the initial flight tests: the first phase was the hover flight test, while the second phase was the transition test. In order to analyse the flight performance, the flight data of the UAV was recorded throughout the test flight. The flight data was then retrieved from the flight controller after the flight was completed.

The flight test operation was divided into three stages: pre-flight preparation, in-flight experiments and post-flight data analysis. These stages were required to ensure safety while collecting the flight test operation data. The pre-flight stage's purpose was to prepare the UAV before the flight operation. This preparation was essential to ensure that all the components and systems ran smoothly before the

flight to prevent any accidents during flying. Then, the in-flight stage was performed by various flight patterns to collect sufficient data on UAV performance. The flight data was collected during flying, and it was saved in an external storage. Lastly, the post-flight stage was performed to analyse the data collected during the flight test. The flight data was retrieved from the external storage and analysed using Mission Planner 1.3.65 from Ardupilot. The analysed flight data was used to assist in the improvement of the configuration and parameters of the UAV for stable and safe flight.

2. MATERIALS AND METHODS

2.1 Conceptual Design

PLANK-V was used in this research as the baseline platform. It is a multi-configuration flying wing developed by the unmanned aerial system (UAS) research cluster of UniKL MIAT (Asri *et al*., 2019; Sahwee *et al*., 2019, 2021). The initial prototype design and prototyping process comprised of the design process, prototype fabrication, and flight tests to ensure that the design was suitable for flying, which took about two months. The prototyping process is essential for the platform to be used for other optimisation in future studies. Using a separate lift thrust (SLT) quadplane VTOL as a basic concept, the general structure of this model was taken from a multi-platform UAV developed by Mansor (2019), a flying wing design that consisted of a wing, fuselage and vertical stabiliser. Then, a H-configuration quadrotor structure was installed on the flying wing that acted as a vertical lifting system. The computer aided design (CAD) design of this model was made using CATIA V5 and is shown in Figure 1. The design and fabrication of this model took into account a few design constraints, such as material cost, ease of fabrication process and fabrication equipment limitations.

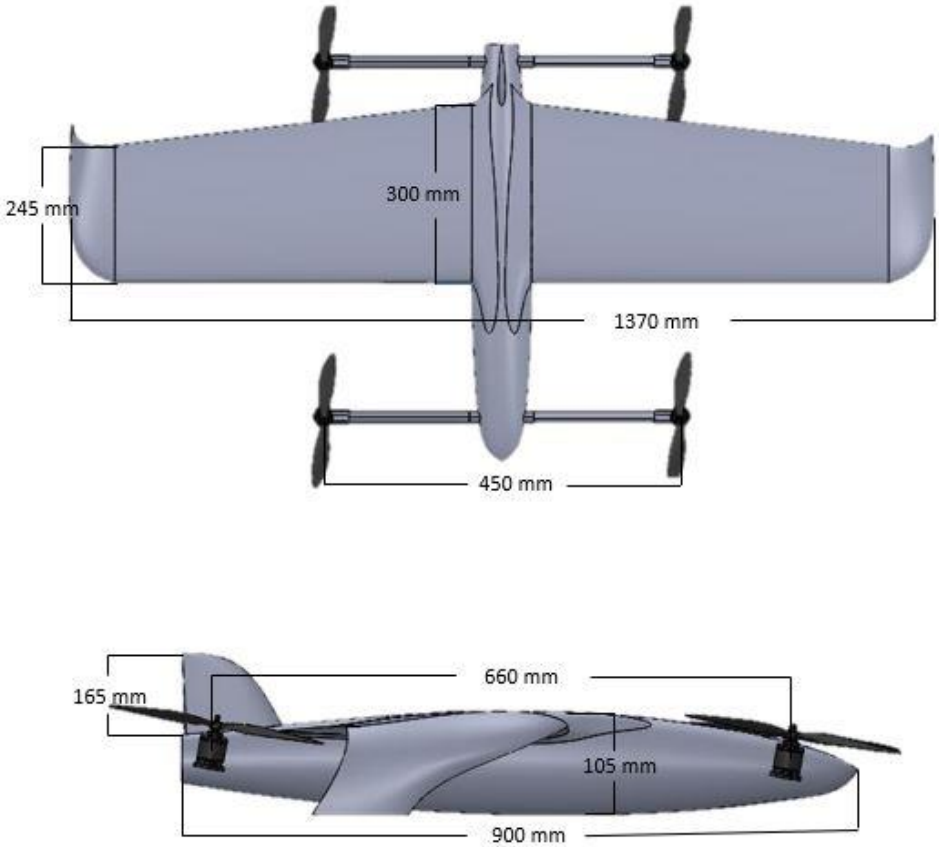


Figure 1: Basic dimensions of the PLANK-V UAV platform.

2.2 Prototype and Flight Test

In order to design an aerodynamically efficient VTOL aircraft, it is important to characterise the drag introduced from the installation of the VTOL components (Hadi *et al*., 2016). The additional VTOL components used compared to a traditional fixed-wing aircraft are VTOL arms, electric propulsion system and propellers. Prior to investigating drag consideration on the VTOL components, the VTOL UAV base design was developed, as shown in Figure 2, as a baseline platform. It is an initial prototype design of a medium cost and low weight VTOL UAV. It is called PLANK-V because "Plank" represents the plank form design of the wing and "V" represents the number 5 in Roman numerals, which indicates the five propulsion units used for this UAV. There are four propulsion units for vertical flight and one propulsion unit for forward flight. In addition to aerodynamic performance, the flight performance of the installed components should not be neglected. The flight performance initially focused on selecting a suitable combination of propellers, motor, battery and electronic speed controller (ESC) to provide sufficient vertical lift for hovering and forward flight. Since the PLANK-V is designed based on a quadplane SLT VTOL UAV, further optimisation work is planned for future improvement to reduce parasitic drag on the VTOL components. This is because the VTOL structure and components will be inactive during forward flight and add to the total parasite drag of the UAV.

Figure 2: SLT VTOL UAV fabricated in-house.

2.2.1 Pre-Flight Preparation

The flight test stages of this UAV were focused on evaluating the performance of the propulsion unit and overall control tuning parameters. Every flight test experiment was a resource-demanding task due to its complexity and unpredictability in nature. As suggested by some researchers, the test flight regime would have to include a few flight phases (Hendarko *et al*., 2018; Bliamis *et al*., 2021; Ewoud, 2022). The pre-flight process contains the pre-flight check, recognising and preparing for the emergency procedure, and finally, the flying site selection. This is done because the UAV stability setting needs to be properly established before proceeding with a more complex transition flight phase and to reduce the risk of uncontrollable flight that could lead to an incident during the flight test. This process ensures a safe flight operation even before the flight test is initiated (Hendarko *et al*., 2018).

Before the flight test, a pre-flight check is performed on the aircraft and its surrounding area. This check involved an overall visual inspection of aircraft conditions. The inspection was conducted to help the pilot determine that the aircraft is in good condition to fly. The mechanical integrity of the structure, fastener condition, electrical wiring condition, movement of control surfaces, updated software and environmental conditions are some of the necessary pre-flight checks that need to be conducted. Some of the main items in the inspection are listed in Table 1 based on the recommendation by McGovern (2017).

Table 1: Pre-flight check items.

No.	Pre-Flight Check Procedure
1.	Verify that the battery is fully charged and adequate for the mission.
2.	Inspect the airframe structure, flight control surface and linkages for any signs of damage.
3.	Check the propeller for chips, cracks, looseness and any deformation.
4.	Inspect the communication link of the transceiver.
5.	Check for correct movement of control surfaces.
6.	Calibrate the compass of the UAV.
7.	Repair or replace any part found unfit to fly prior to take-off.

2.2.2 Emergency Procedure

In order to ensure a safe flight operation, an emergency procedure was planned before the flight test as suggested by few researchers (Mcgovern, 2017; Kim *et al*., 2019; Jayavarman *et al*., 2020; Bliamis *et al*., 2021). The emergency procedure was necessary for the pilot and flight crew to handle any emergencies in the best possible way. It also helps to prevent any accident, and protect the UAV and its surrounding area from heavy damage caused if any accidents occur. The UAV pilot should always be aware and prepared to execute the emergency procedure in any instance during emergencies. The pilot should brief the flight crew about the emergency procedures before the start of the flight operations and have a mission abort site for landing in the case of an emergency. Possible emergencies that could happen during the flight are as follows:

a) Loss of datalink communication
b) Autopilot software bug / error / failure
c) Loss of propulsion
d) Control surface failure
e) The intrusion of another manned / unmanned aircraft into the mission space
f) Mechanical impairment control of the UAV

For the maiden flight test operation, the emergency procedure is very important since the aircraft was not fully tuned, which could lead to flight instability during the testing. Hence, a detailed emergency procedure was planned. Several failsafe options for a well-tuned UAV, such as return to land (RTL) and loiter modes, are available for the autopilot software. However, this option was not applicable for the maiden test flight UAV, as it required a GPS aid with a well-tuned parameter to execute it. Thus, during emergencies, the pilot would have to take complete control of the UAV, and in the worst-case scenario, the pilot would need to execute a controlled-crash landing at a safe place. Therefore, selecting the flight test area is essential to reduce the damage.

2.2.3 Flight Test Area

Another important criterion in any flight test preparation is the selection of the flight testing area. The most important requirement for the flight test area is that it needs to be isolated from the manned aircraft airways, land traffic and residential areas. This is crucial to prevent any damage or even casualties caused by an accident during the flight test. A large empty area is needed for fault recovery if a system failure occurred during the flight (Kugler *et al*., 2018).

There were two test areas identified to be used for different flight phases. The hover test flight requires a small area to test the vertical propulsion system. Thus, this test was conducted at the UniKL MIAT compound (2°51′04.71″ N 101°44′10.54″ E) that had adequate space for vertical flight, as shown in Figure 3 (a).

Conversely, the transition test requires a large area to allow for forward flight and fault recovery. The transition test was carried out in a large, unpopulated area at the Sepang district (2°53′01.39″ N 101°40′40.61″ E). This area provides a large flat grassland where the flight tests could be safely conducted. It is also far from any traffic, residential areas and aviation air traffic activity, thus ideally suited for the research flight testing activities. The flight test area was about 1,000 m long and 300 m wide, as shown in Figure 3 (b).

Figure 3: (a) Hover flight test location. (b) Forward flight test location.

2.3 In-Flight Experiment

The in-flight experiment consisted of two flight phases: hovering and transition flights. The hover phase focused on the ability of the vertical propulsion unit to maintain its altitude in hover mode, while small movements of roll, pitch and yaw are induced. The transition phase was planned to study the transitional stability from hovering to forward flight. Apart from autopilot parameters setting for the transition phase, other main focuses were on the vertical motor thrust angle for smooth flight phase transition and adequate thrust generation by the quad motor for the transition process to happen successfully.

2.3.1 Hover Phase

The hover test phase for this VTOL UAV was to verify the chosen propulsion unit's capability in actual flight conditions. In order to safely test the ability of the vertical propulsion unit, the test was performed in three stages with weight increment for each stage. The weight increment was done in stages, beginning with the bare quad motor and H-frame fuselage, followed by adding an additional forward flight battery. The weights of the stages are listed in Table 2. Finally, the main wing was added for the complete VTOL configuration. These tests were performed in stages to prevent overloading of the propulsion unit.

Table 2: Hover test configuration weight.

	Basic Configuration (g)	External Battery (g)	Wing (g)	Total Weight (g)
Test 1	1,713	-	-	1,713
Test 2	1,713	207	-	1,920
Test 3	1,713	207	350	2,270

Based on Table 2, the weight for the first flight test included the weight of the avionics system, UAV frame and vertical propulsion unit. For the second test, the maximum capacity of the battery planned for the VTOL UAV was used for the forward flight power source, which was a Li-Ion 3,000 mAh with weight of 207 g. Then, to complete the hovering VTOL UAV test, the right and left wings were attached to the fuselage for the third test.

2.3.2 Transition Phase

The second phase of this experiment was the transition phase. It was the most crucial phase for the VTOL UAV as it required complex settings for control stability parameters and mechanical integrity for the UAV to successfully transition from hover to forward flight. The recovery time is a limiting factor and very critical, as during this phase, the thrust of the vertical motor is reduced while the lift generated by the wing is still low. Therefore, the vertical motor was deflected slightly forward initially at a smaller angle to assist the forward thrust in countering this phenomenon. The forward deflection could help the forward movement during the transition phase (ArduPilot Dev Team, 2022). Figure 4 shows the slightly deflected vertical thrust angle that was tested in this experiment.

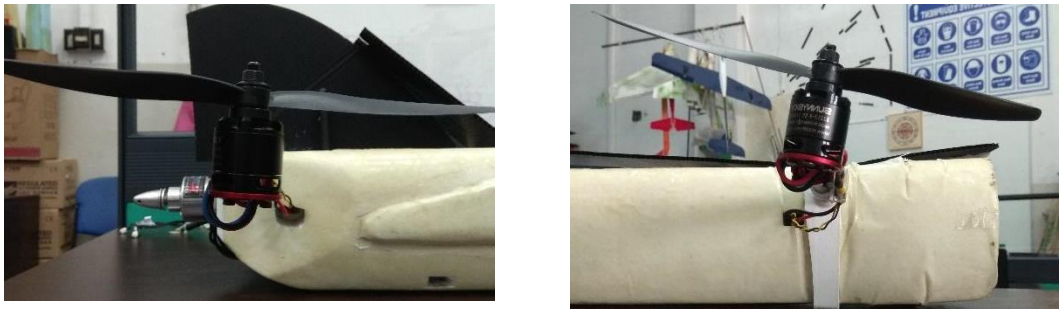


Figure 4: (a) 7° and (b) 10° vertical motor deflection angle.

The lift produced by a bigger deflection angle requires a demand for more power from the vertical motor in hovering. This is due to the resultant thrust that is shifted away from the vertical position, as shown in Figure 5. As a result, the resultant thrust angle influences the power required by the vertical propulsion system as mentioned by Serrano (2018).

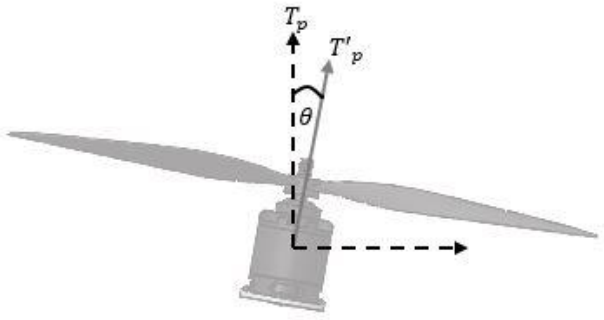


Figure 5: Vertical propulsion system thrust angle.

The deflection angle of the vertical motor is a complex interaction between total weight, the lift generated by the wing shape, external wind speed, propeller sizing and other external factors. Thus, it is best to determine its angle by performing a transition flight test to determine the initial configuration. Two angles were used for this test, which were 7° for the maiden flight and 10° for the

subsequent flight. Based on visual observation, pilot feedback and flight data analysis carried out after the maiden flight, it was found that the 10° angle produced a better transition from hovering to forward flight. The initial 7° angle was used based on the recommendation from Ardupilot's online forum for quadPlane VTOL. Two power supply configurations were used to accommodate different power requirements due to the different deflection angles. The 7° vertical motor angle configuration used a single power source to power up both vertical and forward flight systems, which was a 4,400 mAh Li-Po battery with C-rating of 40C and maximum current of 176 A. In contrast, the 10° motor angle used two separate batteries to provide a higher current. It used the same 4,400 mAh Li-Po battery for the vertical flight system and 1,550 mAh Li-Po battery with C-rating of 75C and maximum current of 1,16.25 A for the forward flight propulsion system. These two configurations are shown in Table 3.

Table 3: Transition test configuration.

	Deflection Angle	**Battery Configuration**	**Weight (kg)**
First flight test	7°	Single battery – 4,400 mAh Li-Po (40C)	2.135
Second flight test	10°	Dual battery – Quad motor (4,400 mAh Li-Po (40C)) Pusher motor (1,550 mAh Li-Po (75C))	2.252

3. RESULTS AND DISCUSSION

3.1 Hovering Test

During the hover test, the PLANK-V was flown initially in QHOVER mode during take-off and altitude hold, then changed to QSTABILISE mode during the landing process. QHOVER mode was used for take-off and hovering as it was a semi-auto mode with an altitude hold function using the autopilot barometric pressure sensor to ease the test process. However, the QSTABILISE mode was used during landing for the pilot to familiarise himself and gain complete control of the UAV. Approximately 50% throttle input was used for the UAV to take off during the test, which indicates that the propulsion system selection was adequate. The UAV was flown to an altitude of 45 m and hovered at that altitude for 2 min. The flight data was stored in the microSD card inside the flight controller and is shown in Figure 6.

The first stage of the hover test flight used approximately 20 A to hover, which increased to 23 A in the second stage and finally 30 A in the third stage. The flight controller managed to respond accordingly to the weight increase of the UAV by increasing the throttle signal output to the motor in order to maintain the altitude during the hovering phase. The current consumed by the UAV increased with the increase of UAV weight. A detailed graph showing the relation between thrust generation and UAV weight is shown in Figure 7.

The flight data in Figure 7 was collected from a propulsion system using a Sunnysky X2212 motor paired with an 8 in diameter with pitch of 4.5 (8 x 4.5) slow-fly propeller. Figure 7 indicates the ratio of thrust produced by the propulsion unit with respect to UAV weight at an average of 1.03 thrust to weight ratio (T/W) for all three stages. The T/W ratio from the actual flight test was slightly lower than presented by Serrano (2018), with a value of 1.2 to 1.5. This was mainly due to the hover condition without any input demand from the pilot, thus the thrust generated was lower. This showed that the propulsion unit was acceptable to be used for the PLANK-V as only 50% throttle was needed for the hovering test.

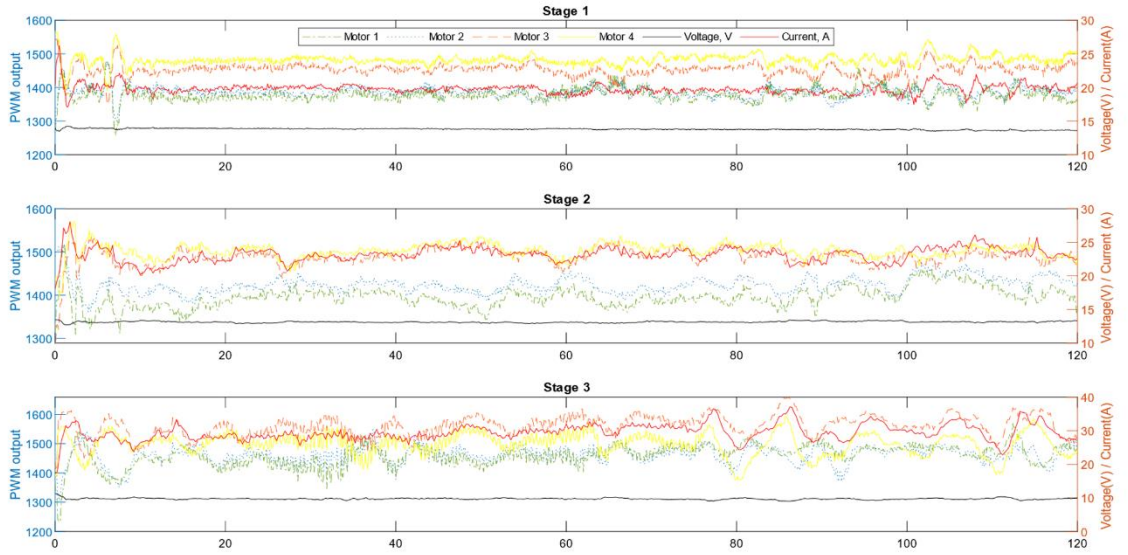


Figure 6: Hover test flight data.

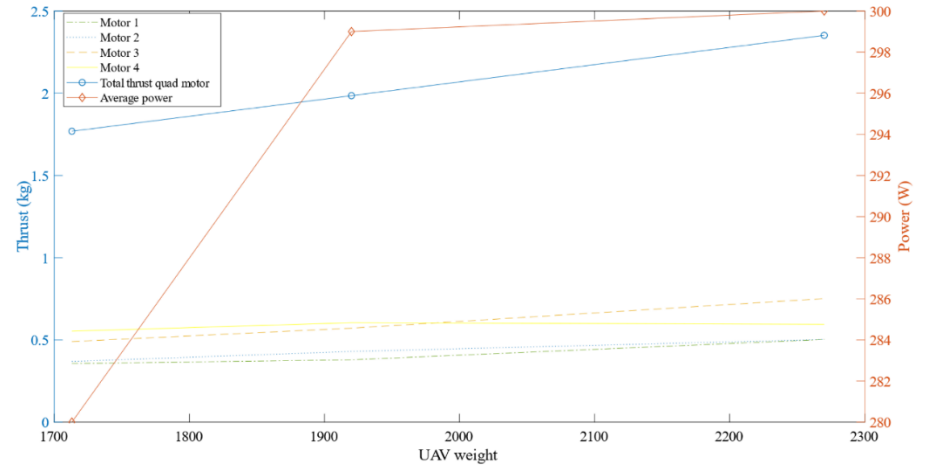


Figure 7: Graph of thrust and power vs weight.

3.2 Transition Test

Once the hovering phase had been completed, a flight transition phase was initiated. During this phase, the lift conversion generated by two propulsion systems occurred. The vertical flight used the vertical motor to provide lift during the take-off and landing, whereas the forward flight used the wing to generate lift with the assistance of a horizontal propulsion system. The interchange of lifting forces between these systems was a complex process that needed to be balanced through the flight experiment.

Two flight modes used for both tests were the QSTABILISE mode for hover and fly-by-wire mode A (FBWA) mode for forward flight. The UAV climbed to about 25 m in the QSTABILISE mode before the transition switch was initiated to the FBWA mode to start the transition phase. Two tests were conducted, with the first test flight configuration using vertical motor deflection angle of 7°, while the second test was at 10°.

3.2.1 Transition Test 1

The first transition test used a 7° vertical motor angle with a single power source configuration. This configuration weighed 2.135 kg, as only one battery was used to power up the UAV. The flight log data of the transition phase of this test is shown in Figure 8. Once the transition switch was triggered to the FBWA mode, the pusher motor pulse width modulation (PWM) signal value increased gradually, while the vertical motor PWM output decreased. The quad motor shut off once the airspeed reached 14 m/s, which indicated that the transition to fixed-wing mode was completed.

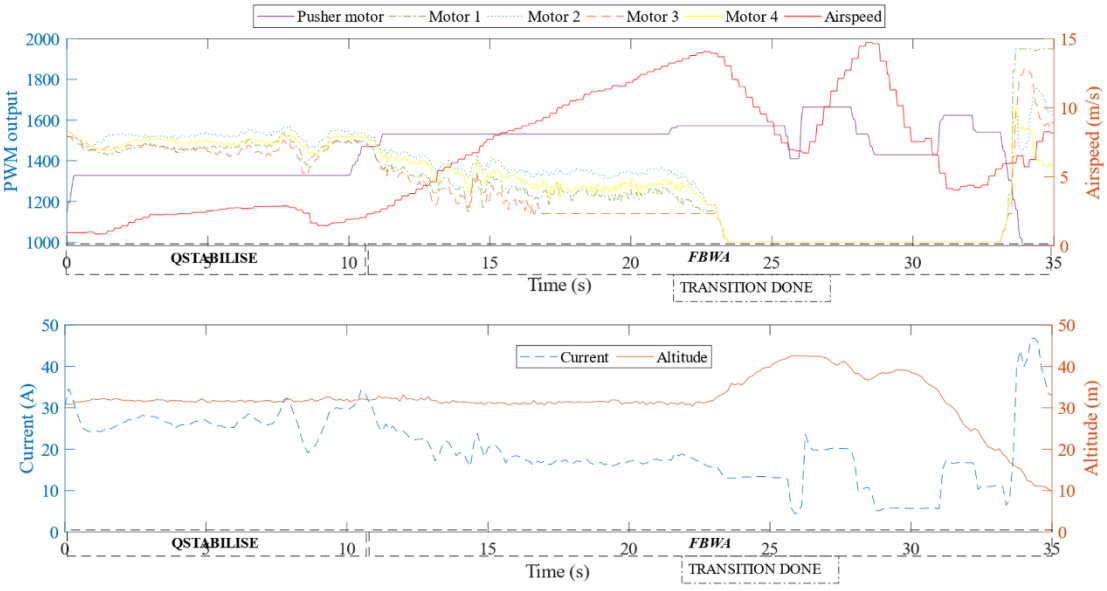


Figure 8: Vertical to forward flight data for Transition Test 1.

However, once the transition was completed, the forward speed dropped to 4 m/s, affecting the lift generation. The lift reduction caused the UAV to lose altitude rapidly, from 40 to 8 m within 5 s. This condition occurred because the single 1,550 mAh battery was insufficient to supply the current needed for the pusher motor during the transition. The current dropped to 4 A after the transition was completed. In order to prevent altitude loss, the pilot initiated the QSTABILISE mode. Since the altitude of the UAV was too low, only limited recovery time was available, which led to a crash landing. The crash caused only minor damage to the UAV as the thick grass cushioned the impact. The damage includes one of the motor shafts being loosened and one of the motor propellers being unfastened. The UAV post-crash data is shown in Figure 8.

A post-crash analysis was performed to identify the improvement plan for the next flight test. Based on the log data, the current supplied to the pusher motor was insufficient for the initial transition to forward flight. The other observation from the flight indicated that the UAV pitched up slightly and one of the motor propellers loosened mid-air during the transition. The pitch up condition was because the centre of gravity (CG) placement had been moved backwards by 24 mm from the initial position. The tail-heavy aircraft was unfavourable, especially for the flying wing, leading to a tip stall. The loosened motor propeller added complexity to the recovery process. In addition, the transition was done while the UAV was making a slight turn, which reduced the possibility of it being safely recovered.

Based on the observation and analysis above, a few improvements were planned for the second transition test:

a) Dual independent power sources were employed for adequate power for the vertical and forward propulsion units.
b) The CG placement was also confirmed, together with properly tightened propellers and motors.
c) The transition phase was performed in straight-line flight conditions to give the pilot complete control to recover if any fault situation occurs.

3.2.1 Transition Test 2

For the second transition flight test, a 10° angle of the vertical motor was used along with a dual independent power source. The changes in motor angle helped provide more forward thrust during the transition, which aided the lift generated by the wing. The second test performed a complete transition phase from vertical to forward flight and finally returned to the vertical flight for landing. The flight log data of the vertical to forward transition phase is shown in Figure 9.

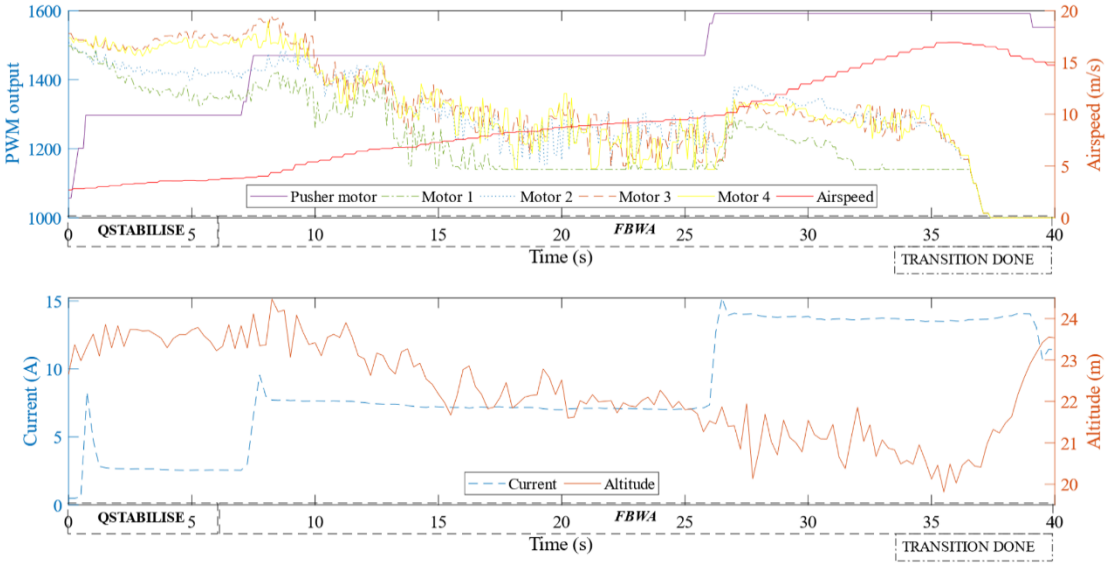


Figure 9: Vertical to forward flight data for Transition Test 2.

Despite the additional weight in this flight test due to the additional battery used, the pusher motor PWM input signal during the transition remained the same as the first transition test. This demonstrated that the quad motor deflection angle helped the forward transition process. The pusher motor consumed a current of 15 A during the transition until it reached the transition airspeed before it reduced to 12 A for cruising. The altitude data shows a cruising altitude of 23 m after the transition.

While the reverse transition from fixed-wing to hover was a success, the hover to land manoeuvre was not very successful. Based on the flight log data shown in Figure 10, the UAV altitude reduced drastically to zero in less than 1 s. From visual observation, the UAV nosedived after the transition, leading to heavy damage. The data shows that after the transition switch was initiated to the QHOVER mode, the PWM output of the pusher motor dropped to zero and stopped the pusher motor, which meant that the lift was dependent entirely on the four vertical motors.

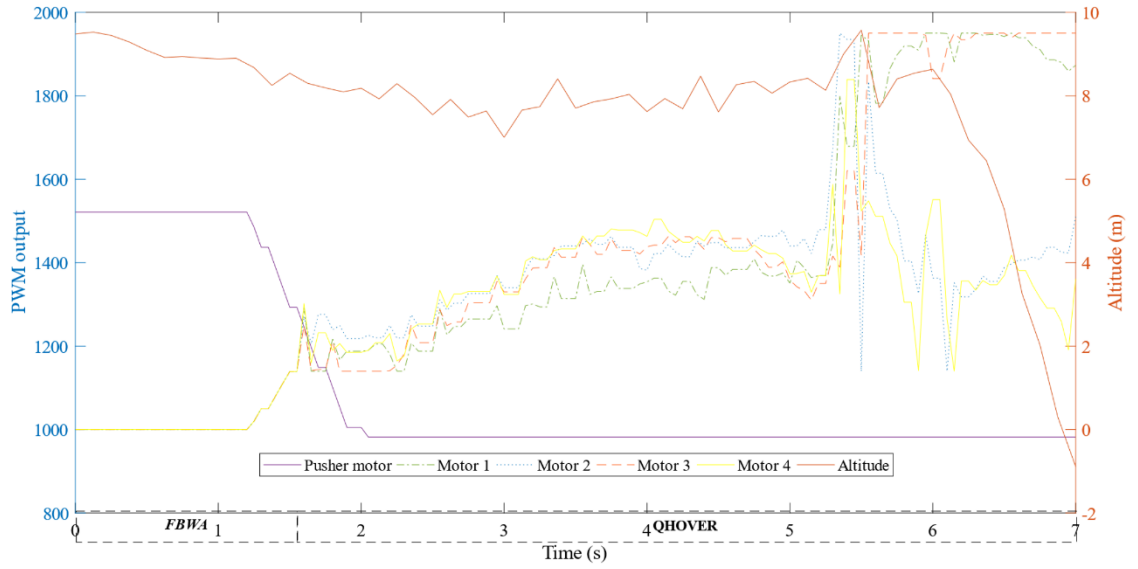


Figure 10: Forward flight to vertical data for Transition Test 2.

The crash broke the nose section of the UAV, and damaged one of the motor holders and propellers. The post-crash analysis found that the QHOVER mode caused the battery to be shifted forward and led to a nose-down crash. From visual observation, the conversion from straight flight to a standstill condition occurred drastically after the transition switch was initiated. This caused the battery to move forward and get disconnected. The movement of the battery also changed the CG of the UAV, causing it to be nose heavy. This resulted in the vertical motor to losing control. It is also shown in Figure 10 that the PWM output of Motors 1 and 3, which are the two front motors of the quadrotor, increased drastically at 5.5 s. This was because the front motors were trying to correct the CG shift that occurred. However, the thrust of both motors was not enough to counter the abrupt weight transfer, thus causing the crash.

4. CONCLUSION

The flight tests aimed to evaluate the baseline model of the PLANK-V VTOL platform in an actual flight condition. The results of the flight tests showed that the selected propulsion system and associated initial design are suitable for future optimisation. The flight test also narrowed down the vertical thrust angle for smoother transition from hover to forward flight. Even though the 10° motor deflection with dual battery setup was completed during the transition from vertical to forward flight, further setup improvement is still needed for the forward to vertical flight transition. The main change needed is to fix all components securely in the right place to prevent any shifting that will cause the CG to change with any drastic flight movement. Several recommendations for the UAV improvement should be focused on the reverse transition from the forward flight to hover through further autopilot parameter tuning.

REFERENCES

ArduPilot Developer Team (2021). *QuadPlane Tips*. Available online at: https://ardupilot.org/plane/docs/quadplane-tips.html (Last access date: 1 May 2022).

Asri, M.H., Sahwee, Z. & Mohd Kamal, N.L. (2019). Propulsion system drag reduction for vertical takeoff and land (VTOL) unmanned aerial vehicle (UAV). *Proc. 2019 Int. Conf. Comp. Drone App. (IConDA 2019)*, 19–22 December 2019, Kuching, Malaysia.

Bliamis, C., Zacharakis, I., Kaparos, P. & Yakinthos, K. (2021). Aerodynamic and stability analysis of a VTOL flying wing UAV. *IOP Conf. Ser.: Mater. Sci. Eng.*, **1024:** 012039.

Dündar, O., Mesut, B. & Tarık, Ü., (2020). Design and performance analyses of a fixed wing battery VTOL UAV. *Eng. Sci. Tech. Int. J.*, **23:** 1182-1193.

Ewoud S. (2017). *Cyclone Hybrid Vehicle Tailsitter*. Available online at: https://diydrones.com/profiles/blogs/cyclone-hybrid-vehicle-tailsitter (Last access date: 1 May 2022).

Hadi, G.S., Puspita, T.D., Muhammad, R.K., Aris B. & Agus B. (2016). Design of separate lift and thrust hybrid UAV. *J. Instrum. Automat. Syst.*, **2:** 45-51.

Hendarko, T., Indriyanto, S. & Maulana, F.A. (2018). Determination of UAV pre-flight checklist for flight test purpose using qualitative failure analysis. *IOP Conf. Ser.: Mater. Sci. Eng.*, **352**: 012007.

Jayavarman, Abdulkareem, S.M.A., Swee, K.P. & Yong G.H. (2020). Improving aerodynamic efficiency of a Skywalker drone. *AIP Conf. Proc.*, **2233:** 020011.

Kim, K., Seungkeun K., Jinyoung S., Jongmin A., Nakwan K. & Byoung S.K. (2019). Flight test of flying-wing type unmanned aerial vehicle with partial wing-loss. *Proc IMechE Part G: J Aerospace Eng.*, **233:** 1611–1628.

Kugler, M.E., David, S., Matthias, H. & Florian, H. (2018). Real-time monitoring of flight tests with a novel fixed-wing UAV by automatic flight guidance and control system engineers. *4th Int. Conf. Control, Automat. Robotics (ICCAR 2018)*, 20-23 April 2018, Auckland, New Zealand.

Mansor, Y., Sahwee, Z. & Mohd Asri, M. H. (2019). Development of multiple configuration flying wing UAV. *Proc. 2019 Int. Conf. Computer Drone App. (IConDA 2019)*, 19–22 December 2019, Kuching, Malaysia.

McGovern, S.M. (2017), UAS flight test for safety and for efficiency. *2017 Integr. Comm. Nav. Surveillance Conf. (ICNS 2017)*, 18-20 April 2017, Herndon, USA.

Palaia, G., Vittorio C., Vincenzo B. & Emanuele R. (2019). Preliminary design and testing of a VTOL light UAV based on a box-wing configuration. *Aircraft Eng. Aerospace Tech.*, **92:** 737-742.

Sahwee, Z., Mohd Kamal, N.L., Abdul Hamid, S., Norhashim N., Lott N. & Mohd Asri M.H. (2019). Drag assessment of vertical lift propeller in forward flight for electric fixed-wing VTOL unmanned aerial vehicle. *IOP Conf. Ser.: Mater. Sci. Eng*, **705:** 1.

Sahwee, Z., Mohd Asri M. H., Mohd Kamal, N.L., Norhashim N., Ahmad Shah S. & Wan Jusoh W. N. (2021). Drag reduction of separate lift thrust (SLT) vertical take-off and land (VTL) components. *Defence S&T Tech. Bull.*, **14:** 55–69.

Serrano, A.R. (2018). Design methodology for hybrid (VTOL + fixed wing) unmanned aerial vehicles. *Aeron. Aero. Open Access J.*, **2:** 165–76.

HANDOVER FEASIBILITY FOR CELLULAR-CONNECTED UNMANNED AERIAL VEHICLE (UAV)

Nadhiya Liyana Mohd Kamal*, Omran Alshalabi, Hatem Aqil Mior Ahmad Termizi, Zulhilmy Sahwee, Nurhakimah Norhashim, Shahrul Ahmad Shah & Sabarina Abdul Hamid

Unmanned Aerial System Research Laboratory, Avionics Section, Malaysian Institute of Aviation Technology, Universiti Kuala Lumpur, Malaysia

*Email: nadhiyaliyanamk@unikl.edu.my

.

ABSTRACT

The ability of unmanned aerial vehicles (UAVs) can be unlocked to fly beyond the line-of-sight (LOS) by utilising current cellular networks. The configuration of cellular network antennas is primarily optimised for ground users, with the antennas tilted downward. This causes frequent handover when cellular-connected UAVs fly at high altitudes. To this end, this paper presents the investigation of handover feasibility for cellular-connected UAVs at high altitudes using a simulation software known as Radio Mobile. This investigation establishes the reliable operating limits of UAVs by quantifying the handover performance of UAVs through observation of the received signal strength (RSS) value when the handover is initiated. Considering a suburban setting that is modelled by a macro-cellular network, the results obtained show that a UAV that takes off at horizontal distance of 100 m away from the serving cellular base station (BS) gives the best RSS performance as compared to horizontal distances of 50, 200, 500 and 750 m from the serving BS. Such performance is due to better reception from the down tilted antenna at the BS. Therefore, it is recommended that UAV operators fly their UAVs at a distance of at least 100 m away from the serving BS in a macro-cellular network to ensure handover feasibility of UAV.

Keywords: *Cellular-connected unmanned aerial vehicle (UAV); received signal strength (RSS); handover; macro- and micro-cellular network; 4G.*

1. INTRODUCTION

The development of unmanned aerial vehicles (UAVs), commonly referred to as drones, have produced a wide variety of applications. From the public's old understanding of drones as pure "games of hobbyists," "flying cameras of the wealthy," or "industrial operating devices," UAVs have now infiltrated large areas of our economy and are now an essential part of our daily lives. These include surveillance and monitoring, disaster relief, defence, agriculture, as well as delivery (Chen *et al*., 2017; Raffelsberger *et al*., 2019; Mehta *et al*., 2020; Abdalla *et al*., 2021).

In terms of wireless connectivity, UAVs in the market are usually equipped with Wi-Fi and Global Navigation Satellite System (GNSS). While this is mostly adequate for applications with limited coverage requirements, cellular networks, such as Fourth Generation (4G) Long Term Evolution (LTE), may offer a broad coverage area, which is necessary for autonomous flights expanding beyond the line-of-sight (LoS) (Hayat *et al*., 2019). However, cellular networks are designed to serve terrestrial users and thus encounter many challenges to support cellular-connected UAVs (Homayouni *et al*., 2021).

When a UAV is in a hovering flight, it undergoes handovers at different altitudes depending on the type of cellular network. The more natural way to understand handover is the concept of a cell phone system that is inside a car moving from one point to another point. When the cell phone moves out of one cell to the next cell, it must be possible to hand the call over from the base station (BS) of the first cell to that of the following cell with no disruption to the call (Östling, 1996). A smooth handover process is crucial to avoid interruption to any calls as the users move from one location to another. Failure for it to perform reliably can result in dropped calls, and this is one of the critical factors that can lead to user dissatisfaction. The same goes for UAVs that utilise cellular networks, where the connectivity between a moving UAV and the BS must be maintained via a reliable handover process to avoid any loss of communication and control link (Kuruvatti *et al*., 2020).

In this study, handover performance will be quantified in terms of the received signal strength (RSS) at which the handover is initiated by performing radio link simulations in the Radio Mobile software. The outcome of this investigation will be useful for UAV operators by providing them with a guideline of acceptable operational distance in macro-cellular networks. As the handover process is a crucial function of cellular networks (Ghanem *et al*., 2012), UAV operators need to be aware of the requirements. Identifying suitable operational distance with respect to a particular BS is important to establish an optimal flight path with a reliable communication link.

2. BACKGROUND INFORMATION

In terms of wireless connectivity, modern UAVs nowadays frequently employ the IEEE 802.11 Wireless Local Area Network (WLAN) technology for sensor data and proprietary radio technologies for command and control. However, given UAVs' three-dimensional mobility, high relative speeds, and changing altitudes, IEEE 802.11 does not always meet the service requirements of UAV applications envisioned (Hayat *et al*., 2019). Therefore, cellular network standards such as Third Generation (3G), LTE, and Fifth Generation (5G) that are available today are an alternative for UAV communications (Chen *et al*., 2018).

However, this work focuses mainly on the 4G network since the existing infrastructure is well developed and has wide coverage area with satisfying high-speed data transmission capability. The 4G network offers maximum real-world download speeds of up to around 100 Mbps, making it over 20 times faster than 3G. These features have led us to choose 4G as the network for UAVs in this study.

One of the 4G infrastructure features is frequency reuse. Lam *et al*. (2015) described frequency reuse as an effective way to optimise the spectrum. The advantage of this is many transmitters of small output power operating at the same frequency can be used. Hence, it reduces the minimum height of the transmitting antenna and limits escaping power to adjacent cells. The concept of cellular handover also applies to cellular-connected UAVs as they fly in free space. However, Hayat *et al*. (2019) stated that as UAV equipment is substantially different from terrestrial mobile devices, the assumptions made for terrestrial user equipment do not extend to aerial user equipment.

In this study, the investigation considers a UAV that ascends vertically from the ground with the handover process taking place as it ascends into higher altitude. The investigation will be extended to consider different horizontal distances between the BS and UAV. The outcome of this investigation is to quantify the horizontal and vertical distances of the UAV from the BS at which a handover is initiated and its corresponding RSS value. With this information, suggestions on the suitable operating limits can be provided to UAV operators.

3. METHODOLOGY

3.1 Simulation Using Radio Mobile

Radio Mobile is an open-source software (Brown, 2011) developed by Roger Coudè VE2DBE for radio amateurs based on the well-known Longley-Rice irregular terrain model for predicting radio propagation from 20 MHz to 20 GHz using several sets of freely available digital elevation models (DEMs). It also offers a comprehensive radio propagation model that allows for simulation of radio connections in various scenarios by changing the parameters of the radio link. It can also display coverage of the area from a given location on a real-world map (Brown, 2011) .

3.2 Scenarios Considered

This work focuses on macro-cellular networks in a suburban setting that is surrounded with buildings, hilly terrains, and different locations and heights of BS. The RSS is measured by considering both cases where the UAV is connected to the closest BS as well as the neighbouring BS. Comparison of these RSS values allows for identification if a handover is initiated. The cellular-connected UAV will fly vertically up to 120 m to comply with the UAV flight regulations set by the Civil Aviation Authority of Malaysia (CAAM) (CAAM, 2017). In order to simulate the UAV flying forward, horizontal distances of 50, 100, 200, 500 and 700 m away from the closest BS are set up as well. Other parameters that require configuration are the BS antenna, UAV receiver antenna and BS location referring to the standard specification of the macro-cellular networks since the simulation is set up in a sub-urban area.

3.3 Parameter Setup

Table 1 shows the parameters for macro-cellular networks based on the 3GPP standard and is based on the International Telecommunication Union (ITU) guidelines as well as some previous research papers related to this study (3GPP, 2011; ITU-R, 2012; Hayat *et al.*, 2019). The type of antenna used at the BS in this study is a corner antenna due to its high directivity and narrow band. However, the UAV uses an omnidirectional antenna because its radiation pattern is more equally distributed, which allows radio signal to be captured from the UAV regardless of its mobility.

Table 1: Parameters for macro-cellular networks based on the 3GPP standard.

Parameter	Value
Antenna Height (m)	40
Antenna Horizontal Coverage Range (°)	60
Antenna Vertical Coverage Range (°)	100
Antenna Transmitted Power (W)	20
Antenna Down Tilt (°)	-10
Cell Radius (m)	750
Type of Environment	• Typified by wide streets • Building heights are generally less than three stories making diffraction over roof-top likely • Reflections and shadowing from moving vehicles can sometimes occur

3.4 BS Location

In this study, two locations are selected as the BS locations with distance of 1.5 km apart to match the specification of the macro-cellular network (ITU-R, 2012). Xiamen University is chosen as BS University, which is the closest BS with the UAV when it takes off from the ground. Kota Warisan is chosen as BS Town, which is the neighbouring BS to the UAV when it flies. Figure 3 shows the radio frequency link and radio propagation between BS University, the 4G UAV and BS Town.

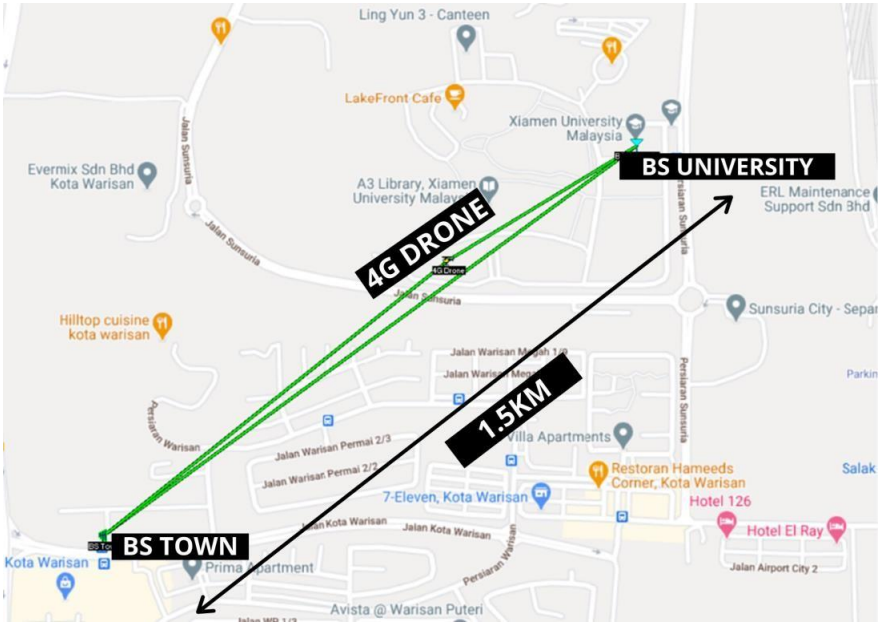


Figure 1: Network link between all the units in Radio Mobile.

4. RESULTS AND DISCUSSION

4.1 RSS for BS University and BS Town at Different UAV Horizontal Distance

Figures 2 to 6 show the RSS values when the UAV takes off at a specific horizontal distance away from the closest BS (BS University) and flies vertically from the ground (0 m) up to altitude of 120 m.

At 50 m away from the closest BS, Figure 2 shows that BS University gives higher RSS during take-off. However, as the UAV flies higher, the RSS received by the UAV from BS Town at altitudes of 70 and 120 m is better as compared to BS University. Hence, handover is initiated twice. The first is at altitude of 70 m, where there is some obstruction that affects the path loss as shown in Figure 7. Thus, the UAV tends to connect with the neighbouring BS (BS Town). The second is at altitude of 120 m, where the UAV seems to fall outside BS University's antenna.

At 100 m away from BS University, Figure 3 shows that BS Town gives better RSS when the UAV flies at high altitude of 120 m. Hence, there is only one point where handover is initiated due to flying at high altitude.

At 200 m away from the BS University, Figure 4 shows that handover was initiated twice as BS Town, which gives better RSS at altitudes of 70 and 90 m. However, in between these altitudes, the RSS signal appears to be larger for BS University, causing the UAV to revert back its connectivity to BS University.

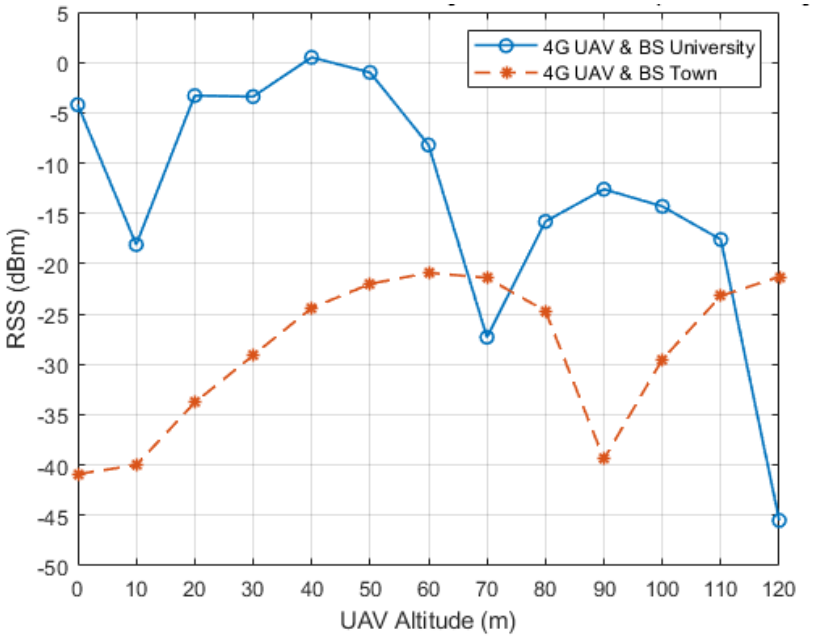


Figure 1: RSS values when the UAV is 50 m away from BS University.

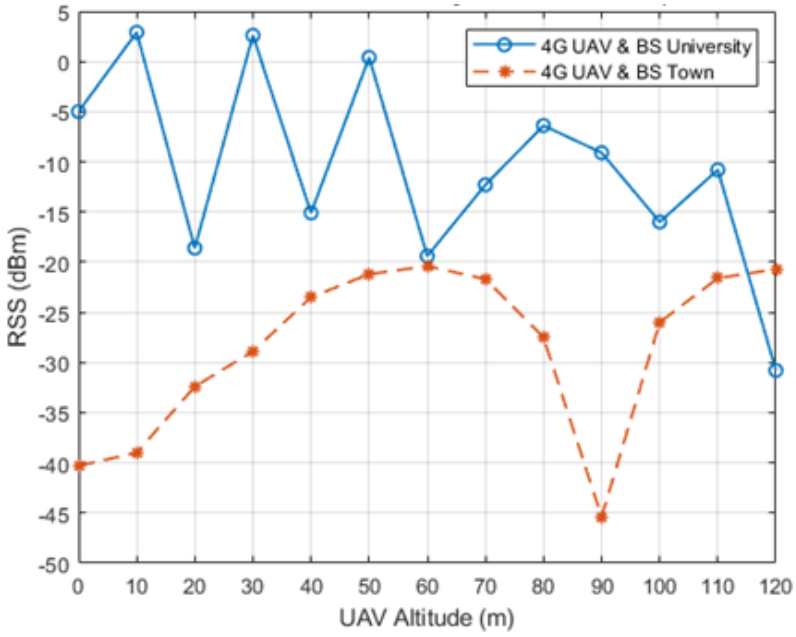


Figure 2: RSS values when the UAV is 100 m away from BS University.

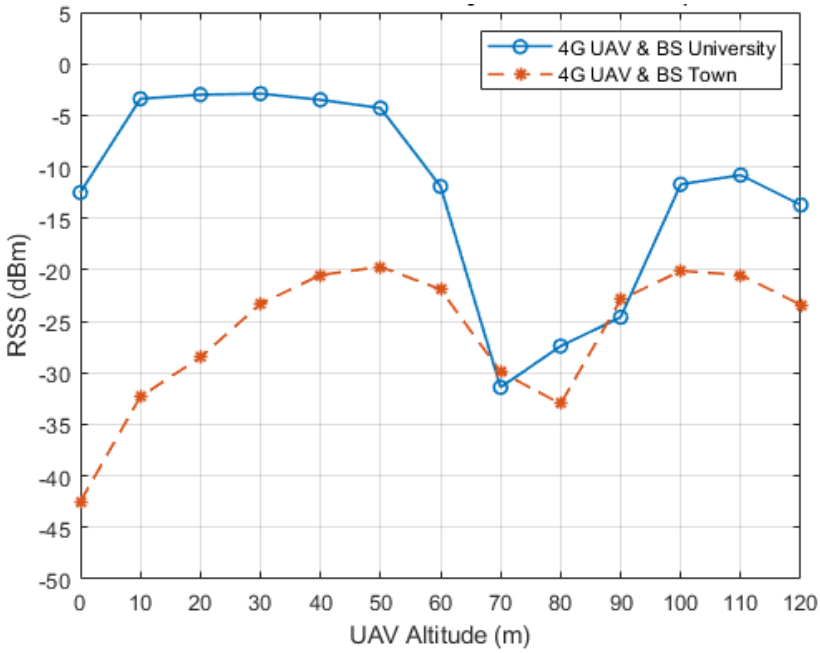


Figure 4: RSS values when the UAV is 200 m away from BS University.

At 500 m away from BS University, Figure 5 shows that there are more handovers initiated as compared to the previous scenario. This is as at horizontal distance of 500 m, the UAV is close to the midway point between BS University and BS Town. This means that based on the free space path loss theory, there is greater path loss and signal attenuation. However, when the UAV flies at an altitude of 100 m, it can be seen that both RSS values drop tremendously. This is because path loss and obstruction have increased significantly, as illustrated in Figure 8.

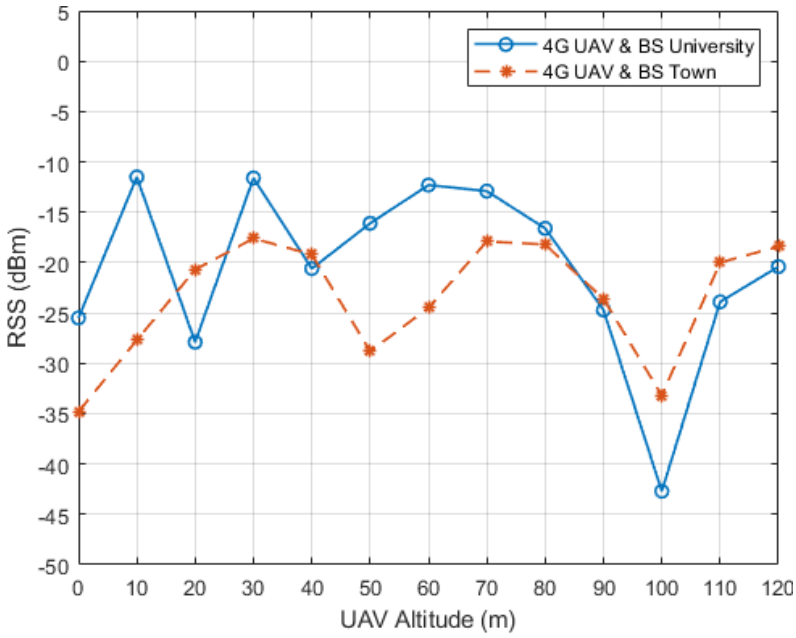


Figure 5: RSS values when the UAV is 500 m away from BS University.

At 750 m away from BS University, Figure 6 shows that most of the UAV signals are served by the neighbouring BS (BS Town). This is because at horizontal distance of 750 m, the UAV is located at the actual midway point between BS University and BS Town. The largest drop in RSS value for BS University can be seen when the UAV flies at altitude of 100 m, which is again due to the high value of path loss and obstruction, as shown in Figure 9.

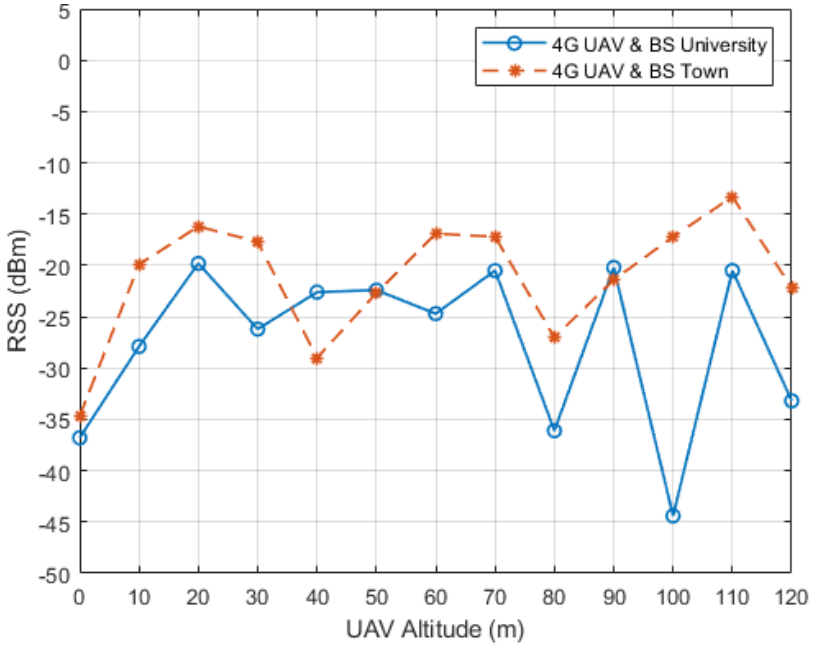


Figure 6: RSS value when the UAV is 750 m away from BS University.

Free Space=52.6 dB	Obstruction=14.4 dB TR	Urban=0.0 dB
PathLoss=71.2dB [4]	E field=91.7dBµV/m	Rx level=-27.3dBm

Figure 7: Radio line profile of the UAV with BS University at altitude of 70 m (50 m from BS University).

Azimuth=238.70°	Elev. angle=8.305°	Clearance at 0.20km
Free Space=70.0 dB	Obstruction=23.8 dB TR	Urban=0.0 dB
PathLoss=98.8dB [4]	E field=76.3dBµV/m	Rx level=-42.7dBm

Figure 8: Radio line profile of the UAV with BS University at altitude of 100 m (500 m from BS University).

Azimuth=220.97°	Elev. angle=4.654°	Clearance at 0.26km
Free Space=73.4 dB	Obstruction=18.4 dB TR	Urban=0.0 dB
PathLoss=97.1dB [4]	E field=74.6dBµV/m	Rx level=-44.4dBm

Figure 9: Radio line profile of the UAV with BS University at altitude of 100 m (750 m from BS University).

4.2 Average RSS of the UAV with the Closest BS (BS University) and Neighbouring BS (BS Town)

The handover is initiated when the UAV receives a stronger signal from the neighbouring BS (BS Town) than from the serving BS (BS university) based on the UAV's flight profile that would happen for a few reasons. First, when there is high obstruction preventing the UAV from getting a strong signal from the serving BS. Second, when the UAV falls out of the serving BS' antenna beam. Third, when the UAV flies at high altitude causing high free path loss.

Figures 10 and 11 show the average RSS values for BS University and BS Town respectively. It can be concluded that the UAV has the best signal strength of -10.6 dBm when is flies at horizontal distance of 100 m away from the serving BS (BS University) and has the worst signal strength of -27.33 dBm when the UAV flies at horizontal distance of 750 m away from the serving BS.

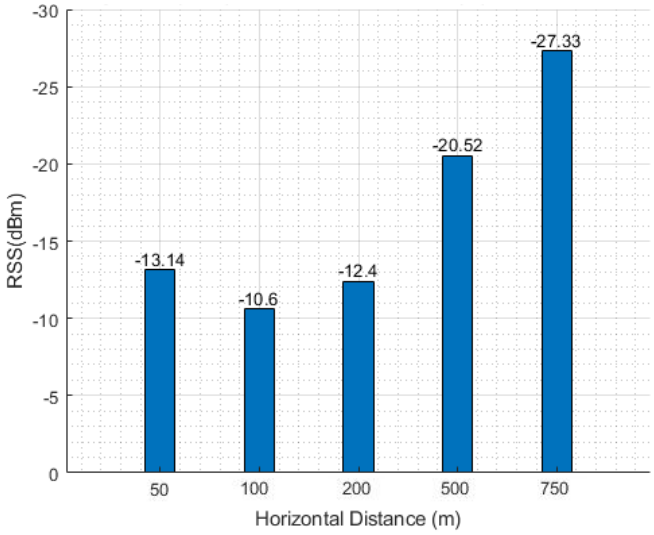


Figure 10: Average RSS received by the UAV from the serving BS (BS University) at five different horizontal distances.

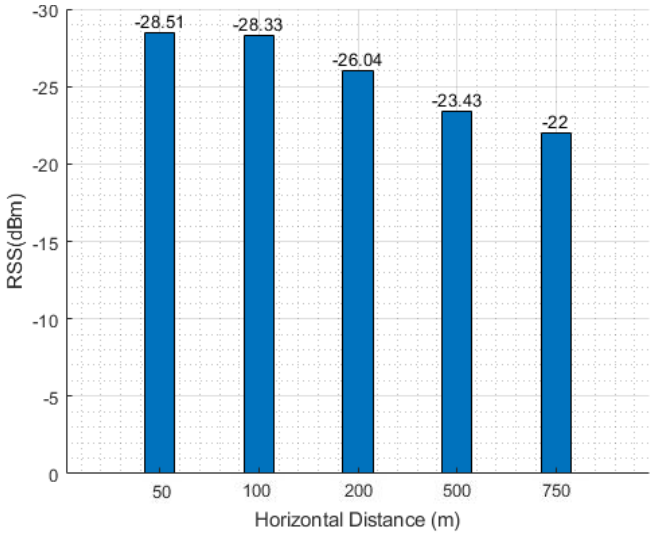


Figure 11: Average RSS received by the UAV from the neighbouring BS (BS Town) at five different horizontal distances.

4.3 Recommendations for UAV Operators

UAV operators utilising cellular-connected UAVs in macro-cellular networks are recommended to fly close to any BS, but it is best to fly at horizontal distance of 100 m away from the closest BS as it will give a strong and better RSS in average. Take-off and flying at this position also reduce frequent handover as compared to other positions. Therefore, flying at this distance has smaller attenuation and free space path loss. At this distance, the UAV also flies comfortably inside a large vertical antenna beamwidth that is radiated by the serving BS and the UAV has longer time before it falls outside the antenna radiation beam as the UAV operator increase's the UAV's altitude. Flying so close to the serving BS, such as horizontal distance of 50 m may not be ideal as the UAV easily falls out of the antenna radiation beam even though it has less free space path loss. Flying a UAV at horizontal distance of 50 m away from closest BS (BS University) gives poorer RSS (-13.4 dBm) on average as compared to flying the UAV away from BS University at horizontal distance of 100 m (-10.6 dBm) and 200 m (12.4 dBm).

UAV operators are not recommended to fly at horizontal distances of 500 and 750 m away from the closest BS, as it would be at the midway point between the two BS, which will give the worse RSS on average. Both these positions are found to have the most frequent handover caused by greater attenuation and larger free space path loss. Hayat *et al*. (2019) observed that the higher the altitude, the more handovers are initiated. This can also be seen in this study's results, where most of the time when the UAV is flying at an altitude of 120 m, handover will be initiated. UAV operators are not suggested to fly at 120 m and above because of poor signal coverage caused by massive attenuation, larger free space path loss, and the tendency of the UAV to fall outside the antenna radiation beam.

5. CONCLUSION

This study suggests a suitable operational horizontal distance for UAV operators to fly a cellular-connected UAV primarily in macro-cellular networks or in suburban areas by analysing the RSS when the UAV is close to the serving BS and comparing it when the UAV is far away from the serving BS. In addition, this work also compiles data that can give UAV operators an understanding of cellular-connected UAV handover feasibility at certain altitudes and distances. The network model is configured according to the specification of real world macro-cellular network properties. According to the results obtained, UAVs that fly far from the serving BS and at higher altitudes are subject to frequent handovers, where the neighbouring base station takes over the connectivity as a backup to resume the flight transmission as normal. Flying the UAV too close to the serving base station is not recommended because the UAV can easily fall out of the antenna's radiation beam due to the narrower vertical beam width. Taking off and flying the UAV at horizontal distance of 100 m from the serving BS is the ideal location as it gives more stable RSS and less signal attenuation.

REFERENCES

3GPP (2011). *3GPP TR 25.996 Version 10.0.0 Release 10: Spatial Channel Model for Multiple Input Multiple Output (MIMO) Simulations*. 3GPP, Sophia Antipolis, France.

Abdalla, A.S., Marojevic, V. & State, M. (2021). Communications standards for unmanned aircraft systems : the 3GPP perspective and research. *IEEE Comm. Std. Mag*., **5**: 70-77.

Brown, I. D. (2011). *Radio Mobile Handbook.* G3TVU, UK.

Civil Aviation Authority of Malaysia (CAAM) (2017). *FAQ on Unmanned Aircraft System (UAS) | Civil Aviation Authority of Malaysia*. Available online at: https://www.caam.gov.my/aviation-professionals/faq/faq-on-unmanned-aircraft-system-uas (Last access date: 24 April 2022).

Chen, J., Xie, J., Gu, Y., Li, S., Fu, S., Wan, Y. & Lu, K. (2017). Long-range and broadband aerial

communication using directional antennas (ACDA): Design and implementation. *IEEE T. Veh. Tech.*, **66:** 10793-10805.

Chen, L., Huang, Z., Liu, Z., Liu, D. & Huang, X. (2018). 4G network for air-ground data transmission: A drone based experiment. *2018 IEEE Int. Conf. Ind. Internet (ICII 2018)*, Seattle, Washington, USA.

Ghanem, K., Alradwan, H., Motermawy, A. & Ahmad, A. (2012). Reducing ping-pong handover effects in intra EUTRA networks. *8th Int. Symp. Commun. Syst., Netw. Digit. Sig. Proc. (CSNDSP 2012)*, 18-20 July 2012, Poznan, Poland.

Hayat, S., Bettstetter, C., Fakhreddine, A., Muzaffar, R. & Emini, D. (2019). Handover challenges for cellular-connected drones. *5th Workshop Micro Aerial Vehicle Netw. Syst. Appl. (DroNet 2019)*, 21 June 2019, Seoul, South Korea.

Homayouni, S., Paier, M., Benischek, C., Pernjak, G., Leinwather, M., Reichelt, M. & Fuchsjager, C. (2021). On the effect of cellular-connected drones on terrestrial users: field trials. *Int.Cong. Ultra Modern Telecomm. Contr. Sys. Workshops*, 27-31 October 2021, Brno, Czech Republic.

ITU-R. (2012). *P Series Radiowave Propagation: Propagation Data and Prediction Methods for the Planning of Short-range Outdoor Radiocommunication Systems and Radio Local Area Networks in the Frequency Range 300MHz to 100GHz.* ITU, Geneva, Switzerland.

Kuruvatti, N. P., Mallikarjun, S. B., Kusumapani, S. C. & Schotten, H. D. (2020). Mobility awareness in cellular networks to support service continuity in vehicular users. *3rd Int. Conf. Info. Comm.Tech. (ICOIACT 2020)*, 24-25 November 2020, Yogyakarta, Indonesia.

Lam, S.C., Subramanian, R., Sandrasegaran, K., Ghosal, P. & Barua, S. (2015). Performance of well-known frequency reuse algorithms in LTE downlink 3GPP LTE systems. *9th Int. Conf. Sig. Proc. Comm. Sys.(ICSPCS)*, 14-16 December 2015, Cairns, Queensland, Australia.

Mehta, P., Gupta, R. & Tanwar, S. (2020). Blockchain envisioned UAV networks: challenges, solutions, and comparisons. *Comp. Comm.*, **151:** 518–538.

Östling, P.E. (1996). Performance of RSS-, SIR-based handoff and soft handoff in microcellular environments. In Rappaport, T.S., Woerner, B.D. & Jeffrey, H.R. (Eds.), *Wireless Personal Communication*, Springer, Boston, Massachusetts, pp. 147-158.

Raffelsberger, C., Muzaffar, R. & Bettstetter, C. (2019). A performance evaluation tool for drone communications in 4G cellular networks. *Int. Symp. on Wireless Comm. Sys.*, 27-30 August 2019, Oulu, Finland.

ADJACENT SATELLITE INTERFERENCE IN GLOBAL MOBILE SATELLITE COMMUNICATIONS

Dimov Stojce Ilcev

University of Johannesburg (UJ), Johannesburg, South Africa

Email: ilcev@uj.ac.za

ABSTRACT

This paper introduces the specific influence of adjacent satellite interference (ASI) as a unique negative propagation consideration affecting all types of satellite networks including global mobile satellite communication (GMSC) systems, which particularly negatively affects geostationary Earth orbit (GEO) satellite constellations. Any satellite network including GMSC could potentially cause ASI effects to their fixed and mobile users. There are two types of ASI effects, namely uplink and downlink ASI, which are discussed and examined in this paper. Downlink ASI occurs when the mobile receiving antenna beamwidth is large enough to receive additional signal levels from adjacent GEO satellites. On the other hand, uplink ASI occurs when adjacent satellites receive and rebroadcast strong uplink signals from mobile Earth station (MES) antennas, often because the antennas are either too small or improperly pointed. Modern mobile satellite terminals use very small antennas mounted onboard ships, land vehicles (road and rail), and aircrafts, including mobile military platforms. In practice, these mobile satellite communication terminals and antennas need to be small to increase transportability and facilitate easy installation. Antennas of this dimension greatly limit the achievable link budgets of satellite networks. In addition, the pointing error and focus of such antennas often requires using efficient modulation and power spectral density reduction technology successfully to mitigate all ASI influences.

Keywords: Adjacent satellite interference (ASI); global mobile satellite communication (GSMC); geostationary Earth orbit (GEO) satellites; mobile Earth station (MES); global fixed satellite communication (GFSC).

1. INTRODUCTION

In global fixed satellite communication (GFSC) systems and particularly in all models of global mobile satellite communication (GMSC) systems, there are three main types of radio frequency interference (RFI) affecting geostationary Earth orbits (GEO) satellite constellations, which are adjacent satellite interference (ASI), co-channel interference (CCI) and cross-polarization interference (XPI). Any satellite user may significantly contribute to ASI influence and other propagation interferences by applying too much uplink power, using too small antenna systems for certain applications, and / or failing to properly set the polarizations of their terminals (Ayala *et al.* 2008; Cochetti, 2014; Acharya, 2017; Barbuddhe e*t al* 2020; AsiaSat, 2022).

Many modernized global mobile and fixed satellite communication networks provide integrated segments consisting of multiple ground Earth stations (GES) and space segment stations, such as GEO, medium Earth orbit (MEO) and low Earth orbit (LEO) satellite systems, as well as other satellite networks (Collin, 1985; Flock, 1987; Elbert, 1997, 2014; AsiaSat, 2022; CNS Systems, 2022). In such kind of satellite systems, the performances of their digital links are determined by the interference power level in both uplink and downlink. In fact, the contribution of each station is normally different to each other, so it is necessary to evaluate the interference sources individually. The hypothetical system architecture of the GEO multiple satellite network can consist multi-user

GES terminals connected to a GEO satellite. Figure 1 shows the configuration of four GES terminals, which are not interfering with the system, with each GES communicating with the respective GEO satellites. On the other hand, Figure 2 illustrates the satellite links with undesired interference, which in particular can come even from different GEO satellite networks. Many strategies can be used to mitigate and eliminate these negative interferences, such as angular separation, antenna gain patterns, type of modulation and multiple access system (Freeman, 1987; Ilcev, 2013, 2017; Ghasemi e*t al* 2016; Graham, 2017; CNS Systems, 2022).

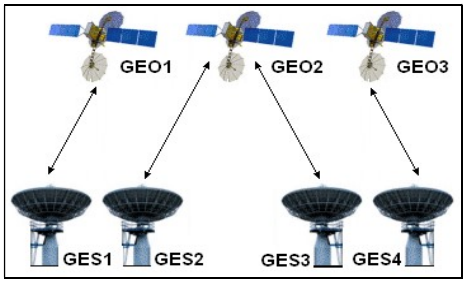


Figure 1: Satellite communication links without interference.
(Source: Ilcev, 2017)

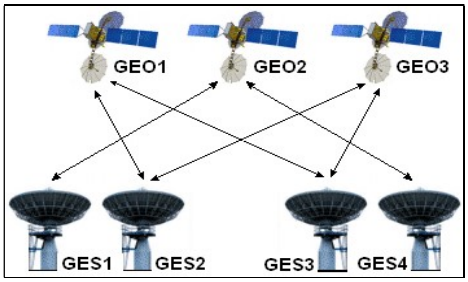


Figure 2: Satellite communication links with interference.
(Source: Wei-Chi *et al*., 2012)

2. INFLUENCES OF ASI IN GMSC

In mobile satellite networks, there are two types of channels to be considered, which are the variable channel between the MES terminals, and satellites or service links; and the constant channel between the GES terminals, and GEO satellites or feeder links. These two channels have many different characteristics that need to be taken into account during the system design examination. A more critical factor is the variable channels used by MES, since transmitter power, receiver gain, and satellite visibility are sometimes quite restricted in comparison to the constant link. Thus, the propagation path profile between MES terminals and satellites varies continuously whenever the MES terminals are in motion. These variations are most significant and frequent in the case of maritime and aeronautical environments. MES terminals, such as ships, land vehicles (road and rail) and aircrafts use small antennas as an essential tool for operational and economic reasons. As a result, a number of low gain-to-noise-temperature (G/T) value MES terminals with smaller mobile antennas have been developed (Flock, 1987; Ilcev, 2005, 2017; Acharya, 2017; Ghasemi *et al* 2016; Graham, 2017; CNS Systems, 2022).

However, such antenna systems are subject to the restriction of frequency utilization efficiency, coexistence between two or more satellite systems in the same frequency band, and / or overlap area where both satellites are visible. For coordination between two different satellite systems in the same frequency band, a highly reliable interference evaluation model covering both interfering and interfered with conditions is required. Investigation into this area has been undertaken in particular by

the International Telecommunication Union Radiocommunication Sector (ITU-R) Study Group 8. The advancement of such a model is an urgent matter for the ITU-R considering the number of GMSC systems that are being developed (Intelsat, 2015; Ilcev, 2016, 2018; ITU, 2016; Barbuddhe *et al* 2020; AsiaSat, 2022; CNS Systems, 2022).

In GMSC networks, the desired signal from a GEO satellite and interfering signal from a different adjacent satellite independently experience amplitude fluctuations due to multipath fading, necessitating a different treatment from that for fixed satellite systems. The main technical requirement is the formulation of statistics for differential fading, which is the difference between the amplitude of two GEO satellite signals. In fact, the method given by ITU-R (1999) presents a practical prediction method for signal-to-interference (S/I) ratio, where the effect of thermal noise and noise-like interference is taken into account, assuming that the amplitudes of the desired and interference signals affected by the sea reflected multipath fading follow the Nakagami–Rice distributions. Therefore, this situation is quite probable in maritime GMSC systems (ITU, 1996, 2017; Ilcev, 2005, 2017; Ma *et al* 2015; Kolawole, 2017; CNS Systems, 2022).

3. BASIC MODEL OF ASI IN GMSC

Modern GMSC systems, especially for maritime and aeronautical applications, usually deploy digital and adaptive-impairment mitigation techniques that tend to benefit system performance in a frequency re-use environment. In such a way, both downlink and uplinks with large increase in transmit power may generate substantial interference to adjacent satellite links (Flock, 2014; Ilcev, 2017, 2018; Barbuddhe *at al* 2020; CNS Systems, 2022).

This type of intersatellite interference is caused by the presence of side lobes in addition to the desired main lobe in the radiation pattern of the Earth ground station antenna. At this point, if the angular separation between two adjacent satellite systems is not too large, it is quite possible that the power radiated through the side lobes of the antenna's radiation pattern, whose main lobe is directed towards the intended satellite, interferes with the received signal of the adjacent satellite system. In a similar manner, transmission from an adjacent satellite can interfere with the reception of an Earth ground station through the side lobes of its receiving ground antenna's radiation pattern. The basic assumptions of the ASI intersatellite model in GMSC networks is defined by the mutual interference phenomenon. An example of downlink interference between adjacent satellite systems on the MES terminal side is shown in Figure 3, while an example of uplink interference on the satellite side is illustrated in Figure 4 (ITU, 2016; Ilcev, 2017; Acharya, 2017; Graham, 2017; Barbuddhe *et al* 2020; CNS Systems, 2022).

This example applies to multiple systems that share the same frequency band. More exactly, it is anticipated that the interference causes an especially severe problem when the interfering GEO satellite is at a low elevation angle viewed from the ship's MES depicted in Figure 3, because the maximum level of interference satellite signal suffered from multipath fading increases with decreasing elevation angle. On the other hand, another situation is GEO satellite interference between beams in multi-spot-beam operation, where the same frequency is allocated repeatedly. Figure 4 illustrates the interfered GEO satellite, which receives the desired signal from the shipborne MES and interference signals from the airborne MES (Elbert, 2014; ITU, 2016; Acharya, 2017; Ilcev, 2017; Barbuddhe *et al* 2020; CNS Systems, 2022).

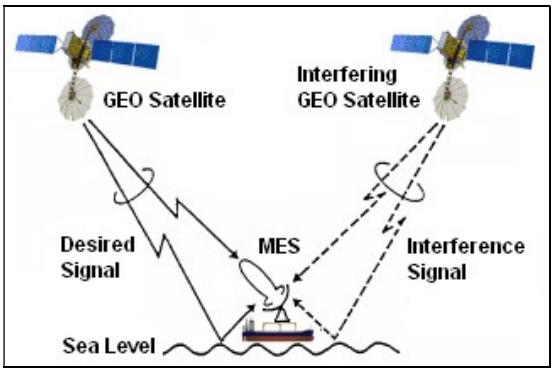


Figure 3: Satellite downlink interference.
(Source: ITU, 1996)

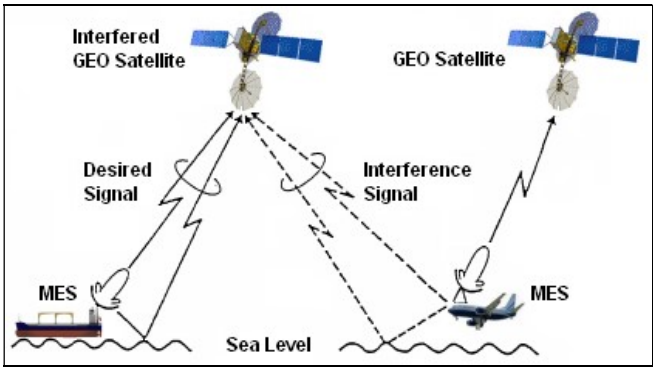


Figure 4: Satellite uplink interference.
(Source: ITU, 1996)

4. PERFORMANCE OF ASI VALUES

The interference due to adjacent satellites is a type of the critical phenomenon in both fixed and mobile satellite communication and networking. As stated earlier, ASI influence can be broken down into two parts, namely uplink and downlink ASI. This type of satellite transmission impairments is depicted in Figure 5, where satellites GEO 1 and GEO 2 are two adjacent satellites. The transmission from terminal GES 1 to satellite GEO 1 on its uplink, in addition to directing its radiated power towards the intended satellite through the main lobe of its transmitting antenna's radiation pattern, also sends some power, though unintentionally, towards satellite GEO 2 through the side lobe (Ilcev, 2017; ITU, 2017; Graham, 2017; Kolawole, 2017; Barbuddhe *et al* 2020).

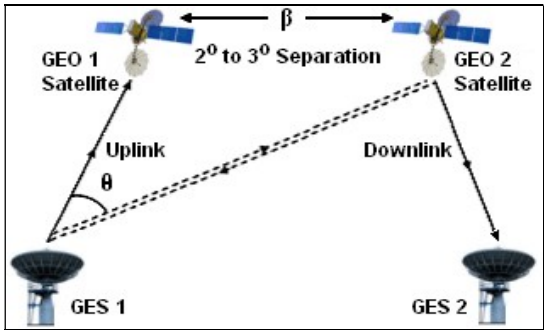


Figure 5: Impairments due to adjacent satellites.
(Source: CNS Systems, 2022)

In such a way, the desired and undesired paths are shown by solid and dotted lines respectively, while θ is the angular separation between the two GEO satellites as viewed by the GES terminal, while β is the angular separation between the satellites as viewed from the center of the Earth; in other words, β is simply the difference in longitudinal positions of the two satellites. Revisiting the problem of interference depicted in Figure 5, the transmission from satellite GEO 2 in its downlink, in addition to being received by the intended terminal GES 2 as shown by the solid line, also finds its way to the receiving antenna of the undesired terminal GES 1 through the side lobe as shown by the dotted line (Intelsat, 2015; Maini *et al* 2015; Graham, 2017; ITU, 2017; Kolawole, 2017; Ilcev, 2018).

Quite obviously, this would happen if the off-axis angle of the GES antenna radiation pattern is equal to or more than the angular separation θ between the adjacent satellites. At this point, the θ and β values are interrelated by the following equation:

$$\theta = cos{-}1\ [d\mathrm{A}2 + d\mathrm{B}2 - 2r\,2\ (1 - cos\,\beta)]\ /\ 2d\mathrm{A}d\mathrm{B} \tag{1}$$

where dA = slant range of satellite GEO 1; d_B = slant range of satellite GEO 2; and r = geostationary orbit radius (Saunders, 1999; Miani *et al*. 2007; Ilcev, 2013).

For a known value of θ, the worst-case acceptable value of the off-axis angle of the antenna's radiation pattern can be computed. Similarly, for a given radiation pattern and known off-axis angle, it is possible to find the minimum required angular separation between the two adjacent GEO satellites for them to coexist without causing interference to each other. Thus, taking the case of downlink satellite interference and determining the expression for carrier-to-interference (C/I) ratio, the desired carrier power C_D for the downlink channel in dBW can be expressed with the following equation:

$$C_D = EIRP - L_D + G \tag{2}$$

where the equivalent isotropically radiated power (*EIRP*) = measured radiated power of an antenna in a specific direction or desired *EIRP* (in dBW); L_D = downlink path loss for the beam from the desired satellite (in dB); and G = Earth station (GES) antenna gain in the direction of the desired satellite (in dB). The interfering carrier power for the downlink channel (I_D) in dBW is given by the following equation:

$$I_D = EIRP' - L_{D'} + G' \tag{3}$$

where *EIRP'* = interfering *EIRP* (in dBW); $L_{D'}$ = downlink path loss for the beam from interfering GEO satellite (in dB); and *G'* = Earth station antenna gain in the direction of the interfering satellite (in dB). The expression for *C*/*I* in the case of downlink can then be written as follows:

$$\begin{aligned}(C/I)_D &= (EIRP - L_D + G) - (EIRP' - L_{D'} + G') \\ &= (EIRP - EIRP') - (L_D - L_{D'} + (G - G')\end{aligned} \tag{4}$$

where value $(C/I)_D$ is the *C*/*I* for the downlink channel in dB. In fact, if the path losses are considered as identical, then the above relation will be as follows:

$$(C/I)_D = (EIRP - L_D + G) - (EIRP' - L_{D'} + G') \tag{5}$$

In addition, the term $(G - G')$ is the receive GES terminal antenna discrimination, which is defined as the antenna gain in the direction of the desired GEO satellite minus the antenna gain in the direction of the interfering GEO satellite. According to International Radio Consultative Committee (CCIR), in cases where the ratio of the antenna diameter to the operating wavelength is greater than 100, *G'* as a function of the off-axis angle θ should at the most be equal to $(32 - 25 \log \theta)$ dB. In fact, this is the

forward gain of an antenna as compared to an idealized isotropic antenna, where θ is in degrees (Elbert, 2014; Intelsat, 2015; Maini *et al* 2015; Ilcev, 2017; Kolawole, 2017; AsiaSat, 2022; Maral *et al* 2022).

The requirement of the US Federal Communications Commission (FCC) standards for the same is (29 − 25 log θ) dB. Figure 6 shows a typical GES antenna pattern, which is a plot of the gain versus the off-axis angle along with the CCIR requirements, which gives:

$$(C/I)_D = (EIRP - EIRP') + (G - 32 + 25 \log \theta) \quad (6)$$

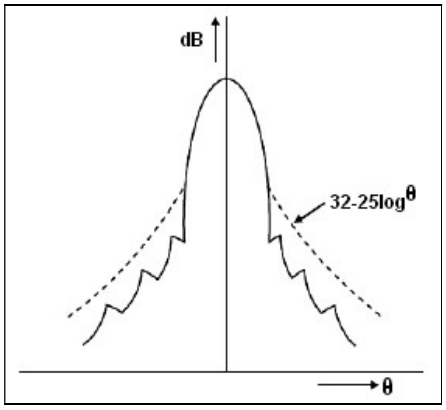


Figure 6: Typical GES antenna pattern.
(Source: CNS Systems, 2022)

A similar calculation can be made for the uplink interference, where a satellite may receive an unwanted signal from an interfering GES terminal. In the case of uplink, the expression for *C/I* can be written as:

$$(C/I)_U = (EIRP - EIRP') + (G - G') \quad (7)$$

where $(C/I)_U$ = *C/I* for the uplink channel in dB; *EIRP* =*EIRP* of the desired GES terminal in dBW; *EIRP'* =*EIRP* of the interfering GES terminal in the direction of the satellite in dBW; *G* = gain of the satellite receiving antenna in the direction of the desired GES terminal in dB; and G' = gain of the satellite receiving antenna in the direction of the interfering GES terminal in dB. The value of *EIRP'* is further equal to the following relation:

$$EIRP' = EIRP^{+} - G_I + (32 - 25 \log \theta) \quad (8)$$

Where the interference is noise like, it is possible to combine the effects of noise and interference. The combined carrier-to-noise ratio (*C/NI*) is given by the following quotation:

$$(C/NI) = [(C/N)^{-1} + (C/I)^{-1}]^{-1} \quad (9)$$

It may be mentioned here that the various terms in Equations 8 and 9 are not in decibels. Here, the power density (ASI_{DL0}) of the downlink ASI can be given by the following equation:

$$ASI_{DL0} = (EIRP/4\pi R^2) \text{ x } G_R(\theta) \; x \; \lambda^2/4\pi \text{ x } 1/BW \quad (10)$$

where *EIRP* is the saturation *EIRP* of the GEO satellite transponder; λ is the wavelength in free space; *R* is the distance from satellite GEO 1 to receiving Earth station Rx2, which is location dependent; $G_R(\theta)$ is the antenna off-axis gain of receiving GES terminal Rx2 at θ; and *BW* is the interference noise bandwidth. The value of *EIRP* depends on the designed and manufactured

characteristics of the satellite. However, the G_R (θ) value is determined by the side lobe performance of the receiving antenna. The industrial practice is to envelope the antenna side lobe by 2, where θ is the off axis angle (Elbert, 2014; Maini *et al*., 2015; Acharya, 2017; Graham, 2017; Ilcev, 2017; Prajapati, 2019; CNS Systems, 2022; Maral *et al.* 2022).

The *EIRP* rule, concerning transmitting *EIRP* density levels requires only very few parameters from the mobile satellite interfering terminal, while the $\Delta T/T$ method needs more necessary parameters for calculation, both from both interfering and victim satellite systems. Various national and international agencies have imposed regulations on ASI issues. In such a way, the $\Delta T/T$ method is used for accessing the necessity of ASI coordination, which is sophisticated and inconvenient to carry out. The $\Delta T/T$ method measures ASI by the increase of noise temperature at the victim receiver caused by the interfering satellite system, to the initial total noise temperature at the victim receiver. It imposes a threshold of 6% as the maximum allowable ΔT/T value. The total noise temperature increase at the victim receiver (R_v) is obtained by the following equation:

$$\Delta T/T = \gamma \Delta T_S + \Delta T / \gamma T_S + T_E \tag{11}$$

where value $\Delta T = \gamma \Delta T_s + \Delta T_E$ is the total absolute noise temperature increase;$T = \gamma T_s + T_E$ is the initial noise temperature of the victim system without consideration of ASI; and γ is the link transmission gain from the output of receiving antenna of satellite to the output of receiving Earth station antenna, which can be calculated with the following equation:

$$\gamma = (EIRP_S/OBO) / [SFD_S(\lambda^2/4\pi)/IBO \times G_V] \times (g_V/l_d) \tag{12}$$

where $EIRP_S$ and SFD_S are satellite saturation *EIRP* and saturation flux density respectively; λ is the uplink wavelength; and *IBO* and *OBO* are input and output back off respectively (Acharya, 2017; Graham, 2017; Ilcev, 2018; Pratt *et al.* 2019; CNS Systems, 2022; Maral *et al.* 2022).

5. PREVENTING ASI IMPAIRMENTS

Satellite interference is a problem that affects all satellite operators and providers, with impairments that all industry players must address as well. In the recent past, much discussion has focused on how to identify the sources of RF interference and their quick elimination. To this end, satellite operators and bandwidth managers will need to mitigate accidental interference by adding a unique carrier identifier (CID) to the signal transmission. CID is a signal embedded into a transmission path that allows satellite operators and end users to work together to identify existing sources of interference (Ayala *et al*., 2008; ITU, 2010, 2016; Richharia, 2014; Ma *et al*, 2015; Ilcev, 2017; CNS Systems, 2022).

This was proven to be technically feasible in 2006, but if manufacturers are expected to include CID in their monitoring and measurement systems, it would be up to the satellite operators and broadcasters to decipher the complexity of the real world when using CID on the receiving side. The regulatory bodies governing fixed and mobile satellite communications include the ITU and their member states. With the high density of communication and other types of satellites in orbit, as well as many more Ka band satellites planned for launch, ASI will be a key concern in the future. New deployed satellite terminals that wish to transmit must meet the emission regulations. The ASI impairments will provide more challenges for small terminals where the antenna side lobe powers are large with respect to their main lobes, thereby limiting the maximum power they are allowed to transmit (Saunders, 1999; Tham, 2014; Acharya, 2017; ITU, 2017; Ilcev, 2018; AsiaSat, 2022; CNS Systems, 2022).

When satellite terminals are on the move, including MES onboard applications, allowable emissions are constrained further as the mechanical satellite antenna pointing accuracy experiences shocks and vibrations that need to be accounted for during movement through land, various sea states or air turbulence. In such a way, to ensure that Earth stations on moving platforms (ESOMP) or our hypothetical MES do not cause harmful interference on adjacent satellite networks, they must operate according to regulatory guidelines, such as the off-axis EIRP Spectral Density (ESD) limits defined by the local regulators, or with other recognized operation limits coordinated with neighboring satellite systems (Saunders, 1999; Ayala *et al.* 2008; Tham, 2014; Ghasemi *at al* 2016; ITU, 2017; Ilcev, 2018; Prajapati, 2019; CNS Systems, 2022).

Ensuring that the interference criteria are met with different co-frequency services is critically important for system designers. Designers must address the conflicting demands to ensure that the resulting interference is within acceptable limits, while at the same time providing an adequate ESD levels that offer reasonable data rates that are acceptable to end users. In order to limit interference to adjacent GEO satellites, ITU has established limits on the ESD limits of a satellite transmit terminal in its off-axis directions. Due to satellite antenna beam characteristics, terminals with large-aperture antennas are not constrained by the main beam but by the side lobes, hence they can transmit higher ESD levels. However, as the main lobe of small antennas is wide, these terminals can be severely limited by the ESD in the boresight direction, *i.e.*, the direction of the maximum gain of the antenna (Wei-Chi *et al.,* 2012; Tham, 2014; Ghasemi *at al* 2016; Ilcev, 2017; Prajapati, 2019; Barbuddhe *at al* 2020; AsiaSat, 2022).

6. CONCLUSION

Recent demand for on-the-move or mobile satellite communication applications has generated interest in a new type of satellite terminal, which can be mounted onboard ships, ground vehicles and aircraft. They can also be transportable, portable in suitcases, semi-fixed and personal applications. In general, they can deploy small, high-performance and compact antennas with tracking systems, servo controllers and petitioners, and include the respective intermediate frequency (IF) and RF equipment.

Antenna size and other transmission parameters are selected to provide two-way mobile communications via GEO satellites and under various seas, terrains and air operational conditions. In fact, terminals mounted on mobiles may cause additional interference to adjacent satellites due to motion-induced antenna pointing errors. From a satellite operator’s perspective, this interference should be maintained at a minimum level.

On the other hand, service providers will seek to design their systems such that the terminals provide enough transmit power to support end-user applications at reasonable data rates and free of interferences. The current and future mobile and fixed satellite communication users should not assume that bandwidth will be completely free of interference since low-level sources are ever-present and virtually unavoidable. However, in the vast majority of instances, interference can be mitigated by careful terminal selection, link design, grooming and coordination of capacity, and by setting and following appropriate operating constraints. Thus, satellite operators, providers and mobile customers can and have successfully worked together to reduce and overcome even high levels of ASI in closely-spaced orbital locations.

REFERENCES

Acharya, R. (2017). *Satellite Signal Propagation, Impairments and Mitigation*. Academic Press, Cambridge, Massachusetts.

AsiaSat (2022). *Adjacent Satellite Interference (ASI) – Effects and Causes on Ku-band DTH Network.* Asia Satellite Telecommunications Holdings Limited, Hong Kong.

Ayala, M. R., Gonzales, E.A. & Mendez, J.A. (2008). Calculation of C/I for mobile satellite system. *WSEAS Int. Conf. Commun.*, Athens, Greece.

Barbuddhe, V., Zanjat, S.N. & Karmore, B.S. (2020). *Satellite Communication – Basic of Satellite Communication.* LAP Lambert Academic Publishing, Chisinau, Moldova.

Cochetti, R. (2014). *Mobile Satellite Communications.* Wiley, Chichester, UK.

Collin, R.E. (1985). *Antenna and Radiowave Propagation.* McGraw-Hill, New York, US.

CNS Systems (2022). *Satellite Antennas and Propagation*. CNS Systems, Durban, South Africa.

Elbert, B.R. (1997). *The Satellite Communication Applications Handbook.* Artech House, London.

Elbert, B.R. (2014). *The Satellite Communications Ground Segment and Earth Station Handbook.* Artech House, Boston.

Flock, W.L. (1987). *Propagation Effects on Satellite Systems at Frequencies Below 10 GHz.* NASA, Washington.

Freeman, R.L. (1987). *Radio Systems Design for Telecommunications (1-100 GHz).* John Wiley, Chichester.

Ghasemi, A. & Abedi, A. (2016). *Propagation Engineering in Wireless Communications.* Springer, Boston.

Graham, L. (2017). *Antennas and Propagation: Technology and Applications.* Larsen and Keller Education, New York.

Ilcev, D. S. (2013). *Global Aeronautical Communications, Navigation and Surveillance (CNS).* Volume 1 & 2. AIAA, Reston.

Ilcev, D.S. (2017). *Global Mobile Communications, Navigation and Surveillance (CNS).* Durban, South Africa.

Ilcev, D.S. (2005). *Global Mobile Satellite Communications for Maritime, Land and Aeronautical Applications*. Springer, Boston.

Ilcev, D.S. (2016). *Global Mobile Satellite Communications for Maritime, Land and Aeronautical Applications, Vol. 1 & 2.* Springer, Boston.

Ilcev, D. S. (2018). *Mobile Satellite Antenna and Propagation.* CNS System, Manual, Durban, South Africa.

Intelsat. (2015). Adjacent Satellite Interference in Mobile/VSAT Environments. McLean, Virginia.

ITU (International Telecommunication Union) (2016). *Interference Calculation Methods.* International Telecommunication Union (ITU), Geneva.

ITU (International Telecommunication Union) (2017). *Radiowave Propagation for the Design of Earth-Space Telecommunication Systems.* International Telecommunication Union (ITU), Geneva.

ITU (International Telecommunication Union). (1996). *Radiowave Propagation Information for Predictions for Earth-to-Space Path Communications.* International Telecommunication Union (ITU), Geneva.

ITU (International Telecommunication Union Radiocommunication Sector - ITU-R) (1999). *Recommendation P.680: Propagation Data Required for the Design of Earth-Space Maritime Mobile Telecommunication Systems*. International Telecommunication Union (ITU), Geneva, Switzerland.

Kolawole M. O. (2017). *Satellite Communications Engineering.* CRC Press, Boca Raton, Florida.

Ma, Y., Luo, W. & Zhu, J. (2015). *ASI Regulations Comparison for GSO Satellite Communication System at Ku-band. Int. Conf. Control, Electr., Renew. Ener. Communications (ICCEREC)*, Bandung, Indonesia, pp. 171-176.

Maini, A.K. & Agrawal, V. (2015). *Satellite Technology - Principles and Applications.* Wiley, Chichester, UK.

Maral, G., Bousque, M. & Sun, Z. (2022). *Satellite Communications Systems: Systems, Techniques and Technology.* Wiley, Chichester, UK.

Prajapati, G.K. (2019). *Satellite Communication System and its Applications.* LAP Lambert Academic Publishing, Chisinau, Moldova.

Pratt, T. & Allnutt, J. E. (2019). *Satellite Communications*. Wiley, Chichester. UK.

Richharia, M. (2014). *Mobile Satellite Communications.* Wiley, Chichester, UK.

Saunders S.R. (1999). Antennas and Propagation for Wireless Communication Systems. Wiley, New York.

Tham, D.W.H. (2014). *Carrier to Interference (C /I Ratio) Calculations*. International Telecommunication Union (ITU). Geneva.

Wei-Chi, I.P., Vong. C.Y. & Zhang R. (2012). *Measured downlink adjacent satellite interference of C-band satellites with 2° orbital separation. World Acad. Sci., Eng. Tech. Conf.*, Paris.

RAIN ATTENUATION AT C, KU AND KA BANDS DETERMINED USING EARTH-SATELLITE LINK BEACON SIGNALS IN TROPICAL REGION

Nur Hanis Sabrina Suhaimi[1*], Khairayu Badron[2], Ahmad Fadzil Ismail[2] & Yasser Asrul Ahmad[2]

[1]Science and Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia
[2]International Islamic University Malaysia (IIUM), Malaysia

[*]Email: sabrina.suhaimi@stride.gov.my

ABSTRACT

Spectrum congestion in frequencies below 10 GHz is pushing satellite network operators to migrate to higher bands. Nonetheless, availability and system performance can be seriously degraded at higher frequency bands when the operating link experiences severe rainfall intensity. A study has been carried out to investigate the different consequences of rain on C, Ku, and Ka band satellite links. Cumulative distribution function (CDF) analysis was performed to obtain the time exceedance percentage, which represents the percentage of link availability in determining the required fade margin. It was observed that rain attenuation experienced by the C band link was very minimal and can be assumed to be negligible, implying that this link was not severely affected by rain. On the other hand, the attained Ku and Ka bands rain fade at 99.9% availability were 10 and 29 dB respectively. The Ku band link attenuation was 32 dB at 99.99% availability. The attenuation at 99.99% availability for the Ka band link was not able to be compiled as rain attenuation exceeded 33 dB, the receiver system started to provide a saturated value. A few extrapolation techniques were used to determine the Ka band attenuation values above 33 dB. Second-degree polynomial was found to fit well with the measurement data since it has lower root mean square error (RMSE) as compared to other techniques applied.

Keywords: *Rain attenuation; tropical region; C, Ku and Ka bands; Earth space satellite link; fade margin.*

1. INTRODUCTION

Satellite communication is a very important technology in defence applications to meet the military's requirements, particularly in areas that cannot be reached by terrestrial communication links. A military communication satellite system requires high data transfer rates for usage by military aircraft, helicopters, ships and personnel (Boyd, 2017). For example, an unmanned autonomous vehicle (UAV) requires instant and secure transfer of large amounts of data during mission phases. Thus, high-capacity satellites are very essential for government use, especially in the defence field, which can significantly improve mission performance (DSSI, 2018).

Most satellite engineers and manufacturers are now shifting to high throughput satellites (HTS), which can cater to larger bandwidths as well as high-speed data transmission. This recent advancement in satellite communication systems has led engineers and designers to assemble links that operate at higher frequencies above 10 GHz, such as Ku, Ka and V bands, since lower frequencies such as L, S, C and X bands are already congested (Ojo *et al*., 2008; Sujimol *et al*., 2015; Ahmad *et al*., 2019; Samat & Jit Singh, 2020). Higher frequencies that provide larger bandwidth can enable frequency reuse and broad-spectrum availability that support high data-rate broadcast, multimedia as well as telecommunication applications (Yussuff *et al*., 2019; Samat & Jit Singh, 2020). However, the challenge in using frequencies above 10 GHz is that the links are susceptible to atmospheric interference including rain (Austin, 2017). As Malaysia is located in the equatorial region where it experiences heavy rainfall intensity per year, rain attenuation is a major challenge for satellite

links that operate at higher frequencies and such circumstances lead to the reduction of the quality of service (QoS) as well as link availability due to serious signal fading and interference (Abubakar *et al*., 2019; Kalaivaanan *et al*., 2020; Usha & Karunakar, 2020).

Studies of rain attenuation have been conducted in Malaysia. They comprise of C band (4 GHz), Ku band (12 GHz) and Ka band (20 GHz) link performance analysis. These studies also compiled information on link vulnerability due to rain (Afahakan *et* al., 2016; Badron *et* al., 2011; Mohamed Yunus *et* al., 2016; Shrestha & Choi, 2017). Deploying new satellite systems involves very complicated procedures and can be very costly (Abubakar *et* al., 2019; Pinder *et* al., 1999). The findings from these studies can be useful in link budget design for higher frequency applications with minimal cost. Currently, attenuation data in Malaysia is scarce and limited. Some of the previously conducted researches only involved restricted periodic data as well as selected statistics, and therefore assessments of the preliminary qualitative information might implicate statistical uncertainty (Cuervo *et al*., 2016). This study hopes to address the research gap where it highlights the impact of rainfall intensity on millimetre-wave transmissions in tropical regions, especially in the equatorial region. Many researchers have raised concerns implying that the International Telecommunication Union Radiocommunication Sector's (ITU-R) predictions for fade margin in tropical regions (ITU-R, 2017) are inaccurate (Nauval *et al*., 2017; Samat *et al*., 2019; Kalaivaanan *et* al., 2020).

This study is carried out to investigate the different consequences of rain on C, Ku, and Ka band satellite links. A compilation of the monthly and annual statistical fluctuations of rain attenuation was acquired during the study. Cumulative distribution function (CDF) analysis is employed in this study as it is one of the first order rain attenuation statistics that can be deployed to evaluate satellite link performance (Das & Maitra, 2016).

2. METHOD

The measurements were based on the beacon signal data of the 4.198 GHz C band and 12.201 GHz Ku band from the MEASAT 3 satellite, as well as the 20 GHz Ka band from the MEASAT-5 satellite. The C and K band Earth stations are located at the MEASAT Teleport and Broadcast Centre, while the Ku band earth station is located at the ASTRO Broadcast Centre. Both centres are located adjacent to each other in Cyberjaya, Malaysia. The Earth stations are constructed at an area of about 20 m above sea level with latitude and longitude of N 2.9350° and E 101.6580°. The locations of the MEASAT 5 and MEASAT 3 satellites are at E 119.5 and 91.5° orbital positions respectively (Ahmad *et al*., 2019). Rainfall intensity data was retrieved from a rain sensor installed at the ASTRO Broadcast Centre.

The received signal strength from the beacon signal during clear sky was individually compared for each day to determine the minimum value of the received signal for every month in 2016. The average minimum received signal acquired for the 12 months throughout 2016 was then denoted as the clear sky value. The clear sky value was deducted from the received signal strength to obtain the attenuation. The rain attenuation was accumulated to obtain the frequencies of occurrence and probability density function (PDF). The PDF was then transformed into a cumulative distribution function (CDF) and the time percentage was calculated to obtain a probabilistic value. This was then used to obtain the time exceedance percentage, which represents the percentage of link availability in determining the required fade margin. Time exceedance percentages of rain attenuation at 0.1, 0.3, 0.01 and 0.03% were established from the obtained CDF. The formula to determine time exceedance is given by:

$$\text{Time exceedence} = \frac{\text{CDF}}{24 \times 60 \times n} \times 100\% \quad (1)$$

where n refers to how many days are in a particular month / year. For example, for a one-year duration, at 0.01% of time exceedance for 1 min-based rain attenuation and rainfall intensity, the

required value of samples is 53. The monthly and annual experimental data were plotted for 2016. This time exceedance graph is used to determine whether the fade margin of the link is capable of achieving the required QoS.

3. RESULTS AND DISCUSSION

3.1 Rainfall Rate

The CDF of rain intensity for 2016 that was collected in Cyberjaya was assessed and investigated. The rainfall intensity (mm) is converted into rainfall rate (mm/hr). The monthly rainfall rate is portrayed in Figure 1, where is it found that at time exceedance of 0.1%, the month of November experienced the heaviest rainfall intensity, followed by March, May, and January with rainfall rates of 80, 67, 67 and 60 mm/hr respectively. The month of September experienced the least rainfall with rainfall rate of 10 mm/hr. At time exceedance of 0.01%, the highest rainfall rates were found for November and March at 122 mm/hr, while for May and January, it was 109 and 106 mm/hr respectively. The month of June was observed to experience the least rainfall with rainfall rate of 64 mm/hr. The rain amount for the year 2016 is higher than 2,000 mm, which is a typical representation of a tropical rain climate. It is observed that in 2016, the rainfall rate in Cyberjaya is high in January, March, May, and November. At time exceedances of 0.1 and 0.01%, the annual rainfall rates at Cyberjaya were 50 and 103 mm/hr respectively.

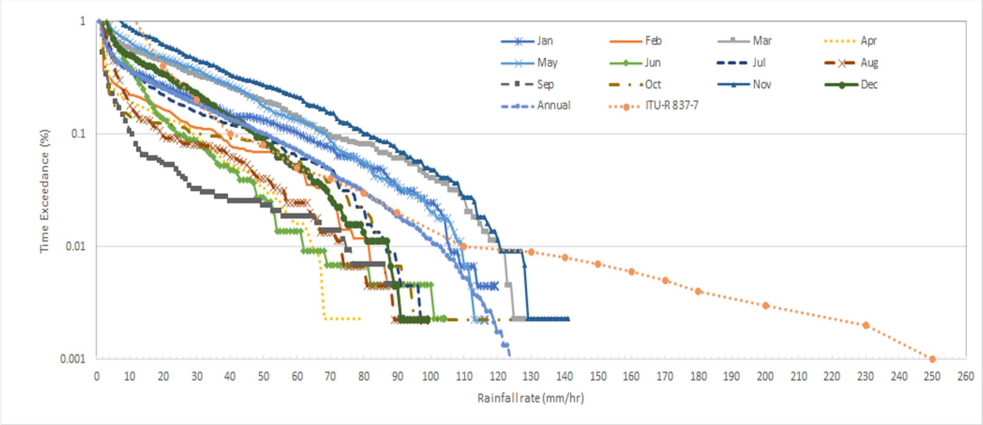


Figure 1: Monthly and annual CDF of rainfall rate for 2016 in Cyberjaya as compared with the ITU-R prediction model.

The rainfall rate for 2016 was regarded as lower as compared to the previous years or a normal year, where it was deemed to be strongly influenced by the natural climate variability of Super El Nino that occurred until the middle of the year (Samat & Singh, 2020). ITU-R's (2017) proposed rainfall rate for Malaysia is within close agreement at time exceedances of 0.1 and 0.01%. However, ITU-R overestimates the value for time exceedance of 0.001%. It has been suggested that the point of rainfall measurement varies according to the area. For example, in 2016, another study reported that for time exceedance of 0.01%, the rainfall rate obtained at a location identified as Puncak Niaga, which is located about 4.93 km away from the Cyberjaya measurement station, was 136 mm/hr (Samat & Singh, 2020). The wind will move clouds that contain water vapour from one place to another place. Therefore, the rain attenuation along the path is contributed by the rain data retrieved along the path near the location of the Earth station (Samat & Singh, 2020).

3.2 Rain Attenuation of C Band

The monthly rain attenuation statistics for the C band link are presented in Figure 2, where the CDF values of rain attenuation are below 2 dB for all percentages of time exceedance. This indicates that C

band is not affected by rain. As compared with the rain attenuation prediction values from ITU-R (2017), it can be seen that the measured values are higher. This might suggest that the prediction method of rain attenuation of C band as proposed by ITU-R somewhat underestimates the instantaneous rain attenuation value.

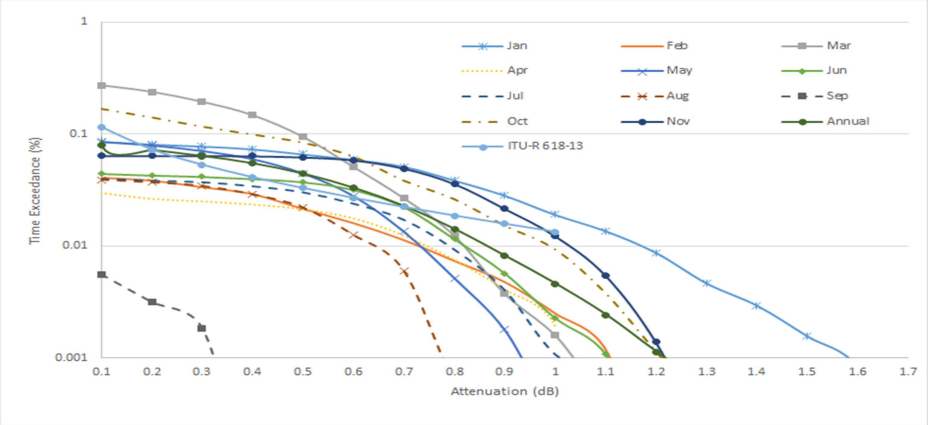


Figure 2: Monthly and annual CDF of C band for 2016 in Cyberjaya as compared with the ITU-R prediction model.

3.3 Rain Attenuation of Ku Band

From the monthly statistics for 2016 as presented in Figure 3, the month of May experienced the highest attenuation for time exceedance of 0.1%, followed by November, March and January, which are 18, 17, 15 and 10 dB respectively. Meanwhile, for time exceedance of 0.01%, the month of May experienced the highest attenuation, followed by January, November and March, which are 42, 41, 36 and 32 dB respectively. According to the rainfall rate statistics, January, March, May, and November are also the months with the highest rainfall rates. On the other hand, the month of September experienced the lowest attenuation as well as the lowest rainfall rate. From the annual statistics as shown in Figure 3, the Ku band attenuation for time exceedances of 0.1, 0.01 and 0.001% are 10, 32 and 43 dB respectively. The measured attenuation of Ku band is higher than the predicted values from ITU-R (2017). This shows that the prediction method of rain attenuation for the Ku band link proposed by ITU-R underestimates the exact rain attenuation value.

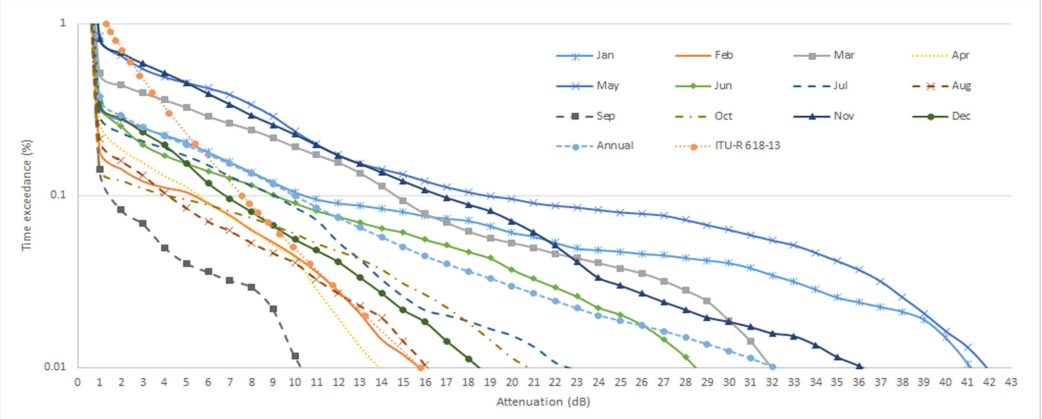


Figure 3: Monthly and annual CDF of Ku band for 2016 in Cyberjaya as compared with the ITU-R prediction model.

3.4 Rain Attenuation of Ka Band

From the monthly statistics for 2016 as presented in Figure 4, the months of January, May and March experienced the highest rain attenuations for time exceedance of 0.1%, which is 33 dB, followed by November with the value of 31 dB. All these months experienced higher rainfall rates as shown in Figure 1. However, the rain attenuation for time exceedance of 0.01% could not be obtained since the graph is saturated at 33 dB. The month of September experienced the lowest attenuation value for time exceedance of 0.01% with 24 dB, and the second-lowest for time exceedance of 0.1% with 13 dB after February with 5 dB. From the annual statistics, as shown in Figure 4, the attenuation for time exceedance of 0.1% is 29 dB, while the attenuation could not be captured for time exceedances of 0.01 and 0.001%. The measured attenuation of Ka band is higher than the predicted values from ITU-R (2017). This demonstrates that the prediction method to determine rain attenuation for the Ka band that has been proposed by ITU-R overestimates the percentage of exceedance for rain attenuation below 15 dB and underestimates rain attenuation above 15 dB. The graph also shows that the data is saturated at the value of 33 dB due to the receiver's level of sensitivity. Extrapolation of existing data needs to be conducted if another method of rain attenuation prediction such as the frequency scaling technique comes into consideration.

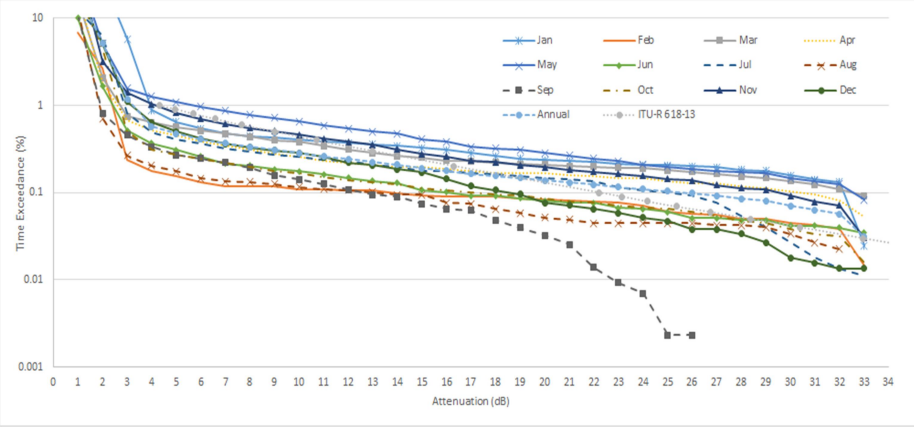


Figure 4: Monthly and annual CDF of Ka band for 2016 in Cyberjaya as compared with the ITU-R prediction model.

For the extrapolation method, a few techniques have been applied, such as linear, power and polynomial equations with the results summarised in Table 1. From all the techniques of extrapolation, second-degree polynomial fits well with the measurement data since R^2 is almost 1 and the root mean square error (RMSE) is smaller than the other techniques applied. The technique of determining the extrapolation graph using second-degree polynomial is as follows:

$$A_{extrapolate} = A_1 x P^2 + A_2 x P + A_3 \quad (2)$$

where P is rain on probability of time exceedance, while constants A_1, A_2 and A_3 are -0.0003586, 0.4717, and 2.531 respectively. According to the extrapolation technique performed for Ka band as shown in Figure 5, for annual CDF for Ka band in 2016, time exceedances of 0.01 and 0.001% are 47 and 55 dB respectively. The extrapolation data is compared to a previous model (Ahmad *et al*., 2019) and predicted values from ITU-R (2017), as shown in Figure 5, with the attenuation results summarised in Table 2.

Table 1: Summarised table for techniques used to determine the best extrapolation graph for extended Ka band.

Method	Linear Extrapolation	First Order Power Law Extrapolation	Second-Order Power Law Extrapolation	Second Degree Polynomial extrapolation
Correlation	0.998356401	0.996950263	0.998358506	0.998411732
R^2	0.996715504	0.993909826	0.996719707	0.996825987
RMSE	0.634472434	0.900234653	0.63407276	0.623710631

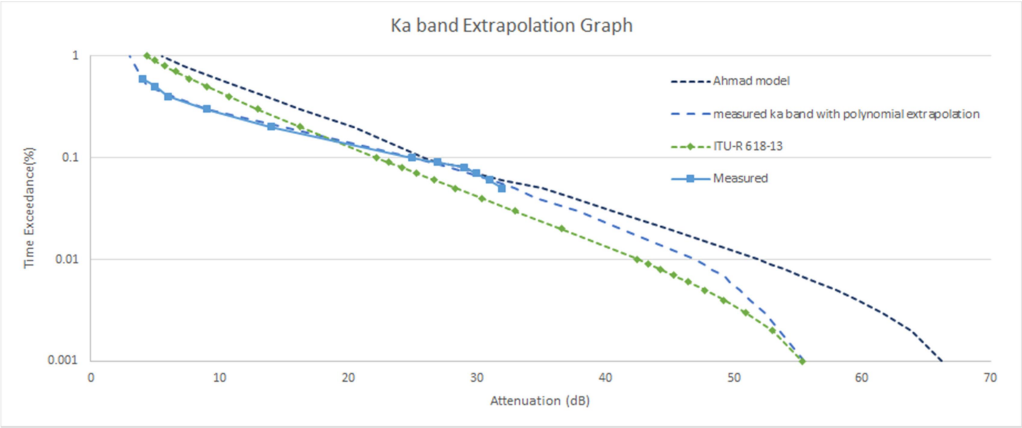


Figure 5: Comparison of extrapolation data technique for extended Ka band data with the models from Ahmad (2019) and ITU-R (2017).

Table 2: Comparison of measured and extrapolated rain attenuation for time exceedances of 0.1, 0.01 and 0.001% with the models from Ahmad (2019) and ITU (2017).

Time Exceedance (%)	Measured Ka Band with Extrapolation Data (dB)	Ahmad's (2019) Prediction Model with Revised Specific Attenuation (dB)	ITU-R's (2017) Predicted Attenuation (dB)
0.1	25	28	22
0.01	47	53	43
0.001	56	66	56

From Table 2, for time exceedances of 0.1 and 0.01%, it is shown that Ahmad's (2019) model overpredicts attenuation, while ITU-R's (2017) model underpredicts attenuation. However, for time exceedance of 0.001%, predicted attenuation by the ITU-R model fits well with the measured values from the extrapolation data, but Ahmad's model overpredicts the attenuation. The QoS at link availabilities of 99.7, 99.9, 99.97 and 99.99% for rain attenuation for the C, Ku and Ka band links are displayed in Table 3. It is shown that attenuation for time exceedances of 0.1, 0.3, 0.01 and 0.03% increase as frequency increases. Higher frequencies are crucially susceptible to rain events, which can reduce the signal link availability. The best QoS should be 99.99% of link availability or equivalent to attenuation not exceeding 0.01% for communication purposes (Suhaimi *et al*., 2018). Therefore, the fade margins needed to enable Ku and Ka bands to perform better are 32 and 47 dB respectively.

Table 3: Annual CDF of rain attenuation for C, Ku, and Ka bands.

QoS (%)	Exceeded Percentage (%)	Attenuation (dB)			
		C Band	Ku Band	Ka Band	Ka Band with Extrapolation Value
99.7	0.3	NA	2	10	9
99.9	0.1	0.08	10	29	25
99.97	0.03	0.6	19	35	38
99.99	0.01	0.9	32	NA	47

4. CONCLUSION

First-order statistics involving monthly and annual CDF of rainfall intensity and rain attenuation for 2016 at specific percentages of time exceedance were presented. It has been ascertained that the annual rainfall rate for time exceedance of 0.01% for Cyberjaya was 103 mm/hr, whereas the ITU-R predicted value is 120 mm/hr. The rain attenuation of C band can be considered trivial and insignificant since the overall value is below 2 dB. However, the same cannot be said for the case of the Ku and Ka band links. It was found that attenuations due to rain for Ku and Ka bands are most severe in May, which experienced considerably high rainfall rate. The Ka band receiver caps all attenuation values above 33 dB once the measurement exceeded the receiver's sensitivity level. In this paper, a few extrapolation techniques were tested on the Ka band measured data so that the estimation of rain attenuation value beyond 99.99% availability can be determined, with second-degree polynomial being found to fit well with the measurement data since it has lower RMSE as compared to other the techniques applied.

Based on the findings from this paper, the dynamic range and sensitivity of the receiver need to be assessed to utilise and employ a Ka band link satellite receiving system in the future. High sensitivity level of the receiver is required to capture attenuation above 33 dB for the Ka band satellite link so that communication link availability of 99.99% can be achieved for 20 GHz Ka band beacon signals. Sufficient QoS value for Ka band satellite links in Malaysia should achieve 99.9% or less link availability if a smaller terminal is being considered. The results from this study offer significant findings related to the implementation of Ku and Ka bands in Malaysia, as well as a reference for future research in the field of attenuation due to rain, particularly in tropical and equatorial regions.

ACKNOWLEDGEMENT

The authors would like to thank MEASAT Satellite System Sdn. Bhd. (MEASAT) and ASTRO Broadcast Centre (ASTRO) for providing the data for this study.

REFERENCES

Abubakar, I., Din, J. Bin, Yin, L.H. & Alhilali, M. (2019). Rain attenuation in broadband satellite service and worst month analysis. *Indones. J. Electr. Eng. Comput. Sci.* , **15**: 1443–1451.

Afahakan, I., Udofia, K. & Umoren, M. (2016). Analysis of Rain Rate and Rain Attenuation for Earth-Space Communication Links Over Uyo - Akwa Ibom State. *Niger. J. Technol.*, **35**: 137.

Ahmad, Y.A., Ismail, A.F. & Badron, K. (2019). Two-year rain fade empirical measurements and statistics of earth-space link at Ka-band in Malaysia. *ASM Sci. J.,* **12**: 35–46.

Austin, O. (2017). Rain attenuation analysis from system operating at ku and ka frequencies rain attenuation analysis from system operating at Ku and Ka frequencies bands. *Am. J. Adv. Res.,* **1**: 7-12.

Badron, K., Ismail, A.F., Din, J. & Tharek, A.R. (2011). Rain induced attenuation studies for V-band satellite communication in tropical region. *J. Atmos. Sol. Terr. Phys.*, **73**: 601–610.

Boyd, A.H. (2017). *Satellite and Ground Communication Systems: Space and Electronic Warfare Threats to the United States Army*. The Institute of Land Warfare, Arlington, Virginia, US.

Cuervo, F., Schonhuber, M., Capsoni, C., Yin, L.H., Jong, S.L., Din, J. & Martellucci, A. (2016). Ka-band propagation campaign in Malaysia - First months of operation and site diversity analysis. *10th Eur. Conf. Antennas Propag.*, 10-15 April 2016, Davos, Switzerland

Das, D. & Maitra, A. (2016). Fade-slope model for rain attenuation prediction in tropical region. *IEEE Geosci. Remote. Sens. Lett.*, **13**: 777–781.

DSSI (Defence & Security Systems International) (2018). The Future of Military Satellite Communications. Availabile online: https://www.defence-and-security.com/features/featurethe-future-of-military-satellite-communications-6097723 (Last access date: 13 April 2022).

ITU-R (International Telecommunication Union Radiocommunication Sector) (2017). *Recommendation ITU-R P.618-13: Propagation Data and Prediction Methods Required for the Design of Earth-Space Telecommunication Systems*. International Telecommunication Union (ITU), Geneva, Switzerland.

Kalaivaanan, P.M., Sali, A., Raja Abdullah, R.S.A., Yaakob, S., Jit Singh, M. & Al-Saegh, A.M. (2020). Evaluation of Ka-band rain attenuation for satellite communication in tropical regions through a measurement of multiple antenna sizes. *IEEE Access*, **8**: 18007–18018.

Mohamed Yunus, M., Din, J., Lam, H.Y. & Jong, S.L. (2016). Analysis of inter-fade intervals at Ku-band in heavy rain region. *2015 IEEE Int. RF Microwave Conf. (RFM),* pp. 276-279.

Nauval, F., Marzuki, M. & Hashiguchi, H. (2017). Regional and diurnal variations of rain attenuation obtained from measurement of raindrop size distribution over Indonesia at Ku , Ka and W bands. *Prog. Electromagn. Res. M*, **57**: 25–34.

Pinder, J., Ippolito, L.J., Horan, S. & Feil, J. (1999). Four years of experimental results from the New Mexico ACTS propagation terminal at 20.185 and 27.505 GHz. *IEEE J. Sel. Areas Commun.*, **17**: 153–162.

Samat, F. & Jit Singh, M.S. (2020). Impact of rain attenuation to Ka-band signal propagation in tropical region: A study of 5-year MEASAT-5's beacon measurement data. *Wirel. Pers. Commun.*, **112**: 2725–2740.

Samat, F. & Singh, M.J. (2020). Site diversity performance in Ka band using a 7.3-m antenna diameter at tropical climate: a comparison of prediction models. *Acta Geophys.,* **68**:1213–1221.

Samat, F., Singh, M.S.J. & Sountharapandian, T. (2019). Rain attenuation prediction model assessment on 3-Year Ka-band signal of MEASAT-5 at tropical region using 7.3-m antenna. J. Metrol. Soc. India, **35**: 201-212

Shrestha, S. & Choi, D. (2017). Characterization of rain specific attenuation and frequency scaling method for satellite communication in South Korea. *Int. J. Antennas Propag.,* **3***:*1-16.

Suhaimi, N.H.S., Badron, K., Ismail, A.F., Ahmad, Y.A., Yassin, M.R.M. & Rahmat, M.H. (2018). Determination of fade margin for Ka band operating in equatorial region. *Mal. J. Fund. Appl. Sci.*, **10**: 229-238.

Sujimol, M.R., Acharya, R., Singh, G. & Gupta, R.K. (2015). Rain attenuation using Ka and Ku band frequency beacons at Delhi Earth Station. *Indian J. Radio Space Phys.,* **44**: 45–50.

Usha, A. & Karunakar, G. (2020). Preliminary analysis of rain attenuation and frequency scaling method for satellite communication. *Indian J. Phys.,* **95**: 1033–1040.

EVALUATION OF PERFORMANCE OF GLOBAL POSITIONING SYSTEM (GPS) SPEED METERS

Dinesh Sathyamoorthy*, Hafizah Mohd Yusoff , Ahmad Firdaus Ahmad Kazmar, Mohd Zuryn Mohd Daud & Maizurina Kifli

Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia

*Email: dinesh.sathyamoorthy@stride.gov.my

ABSTRACT

In this study, Global Positioning System (GPS) simulation is used to evaluate the performance of three commercial GPS speed meters: 1) S1: SkyRC GSM020; 2) S2: XOSS G+; and 3) S3: Magene C406. It is found that with decreasing GPS signal power level, speed errors increase due to decreasing carrier-to-noise density (C/N_0) levels for GPS satellites tracked by the receiver, which is the ratio of received GPS signal power level to noise density. In addition, varying speed error patterns are observed for the each of the readings. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over location and time, resulting in GPS accuracy being location / time dependent It is found that the R3 receiver provided the lowest position errors for both the open area and obstruction area scenarios due to it having higher receiver sensitivity and lower receiver noise, which allows it track higher C/N_0 levels for the available GPS satellites. All three GPS speed meters are found to have relatively similar speed errors. However, Speed Meter S1 is observed to be able to operate at lower GPS signal power levels as compared to the other speed meters. This could be as it higher receiver sensitivity and lower receiver noise, which allows it track higher C/N_0 levels for the available GPS satellites.

Keywords: *Global Positioning System (GPS) simulation; speed measurement; Doppler shift; GPS signal power level; GPS satellite geometry.*

1. INTRODUCTION

Global Navigation Satellites Systems (GNSS) receivers are becoming smaller, cheaper and more reliable, and hence, are being increasingly used for measurement of speed. Speed is defined as the rate of change of position, with its determination requiring measurements of distance and time components (Witte & Wilson, 2004; Keskin *et al*., 2018; Akkamis *et al*., 2021). GNSS speed meters typically employ the Doppler shift method, whereby the GNSS receiver continuously tracks the carrier frequencies of the available GNSS satellites. The difference between the known satellite carrier frequency and the frequency determined at the receiver, known as the Doppler shift, is directly proportional to the speed of the receiver along the direction to the satellite. A minimum of four tracked satellites are required to determine the 3D speed vector of the receiver. A significant advantage of Doppler speed measurement is that it is insensitive to distances from satellites, phase delays and a number of factors that are major sources of error for GNSS positioning using satellite range measurement. However, its accuracy is not constant as it is dependent on the number and geometrical distribution of available satellites (Zhang *et al*., 2006; Huang *et al*., 2013; Gaglione, 2015; Keskin *et al*., 2018).

A number of studies have been conducted to evaluate the accuracy of GNSS speed measurements (Witte & Wilson, 2004; Huang *et al*., 2013; Steinmetz *et al*. 2014; Keskin *et al*., 2018; Alphin *et al*., 2020; Akkamis *et al*., 2021; Barbosa *et al*., 2022). These studies were conducted via field evaluations using live GNSS signals. However, such field evaluations are subject to various error parameters,

such as ionospheric and tropospheric delays, GNSS satellite clock and ephemeris errors, GNSS satellite positioning and geometry, radio frequency interference (RFI), and obstructions and multipath, which are uncontrollable by users (Aloi *et al.* 2007; Kou & Zhang 2011; Pozzobon *et al.*, 2013; Arul Elango & Sudha, 2016).

The ideal GNSS receiver evaluation methodology would be using a GNSS simulator, which can be used to generate multi-satellite GNSS configurations, transmit GNSS signals that simulate real world scenarios, and adjust the various error parameters. This would allow for the evaluations of GNSS receiver performance under various repeatable conditions, as defined by users. As the evaluations are conducted in controlled laboratory environments, they will not be inhibited by unwanted signal interferences and obstructions (Aloi *et al.* 2007; Kou & Zhang 2011; Dinesh *et al*., 2012; Pozzobon *et al.*, 2013; Bi & Yuan *et al*., 2021). To this end, Dinesh *et al*. (2015) employed Global Positioning System (GPS) simulation to demonstrate the effectiveness of GPS speed measurement using the Doppler shift method as compared to the track points method.

In this paper, the study in Dinesh *et al*. (2015) is extended to evaluate the performance of three commercial GPS speed meters: 1) S1: SkyRC GSM020 (SkrRC, 2021); 2) S2: XOSS G+ (XOSS, 2022); and 3) S3: Magene C406 (Magene, 2021). All three GPS receivers employ the GPS L1 coarse acquisition (C/A) signal, which is an unencrypted civilian GPS signal widely used by various GPS receivers. The signal has a fundamental frequency of 1,575.42 MHz and a code structure that modulates the signal over a 2 MHz bandwidth (DOD, 2001; USACE, 2011; Kaplan & Hegarty, 2017).

2. METHODOLOGY

The apparatus used in the study are an Aeroflex GPSG-1000 GPS simulator (Aeroflex, 2010) and a notebook running GPS Diagnostics v1.05 (CNET, 2008) The study is conducted in STRIDE's mini-anechoic chamber (Kamarulzaman, 2010) to avoid external interference signals and unintended multipath errors. The test setup employed is as shown in Figure 1. Simulated GPS signals are generated using the GPS simulator and transmitted via the coupler. The following assumptions are made for the tests conducted:

i) No ionospheric or troposheric delays
ii) Zero unintended GPS satellite clock or ephemeris error
iii) No obstructions or multipath
iv) No interference signals.

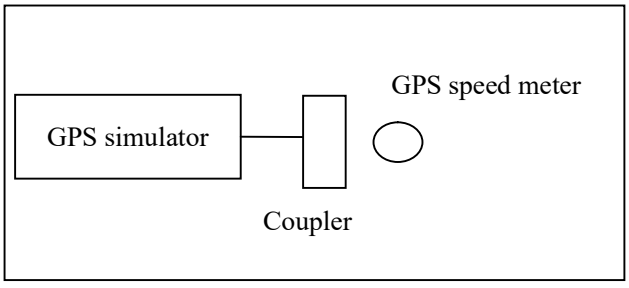


Figure 1: The test setup employed.

The tests are conducted for the route of Kajang (N 2° 58' E 101° 48') to Hanoi (N 20° 54' E 105° 49') for coordinated universal times (UTC) of 0000, 0300, 0600 and 0900. The almanac data for the period is downloaded from the US Coast Guard's web site (USCG, 2022) and imported into the GPS simulator. The GPS signal power levels is set at -130 dBm. The speeds used for the test are 5, 25, 50, 90, 130, 200 and 300 km/h, while the GPS signal power levels used are -130, -135, -140-, -145, -150, -155 and -160 dBm. Each reading is conducted three times for a period of 15 min with the average speed computed.

3. RESULTS & DISCUSSION

For the tests conducted, the recorded speed errors are shown in Figures 2 - 4. With decreasing GPS signal power level, speed errors increase due to decreasing carrier-to-noise density (C/N_0) levels for GPS satellites tracked by the receiver, which is the ratio of received GPS signal power level to noise density. Lower C/N_0 levels result in increased data bit error rate when extracting navigation data from GPS signals, and hence, increased carrier and code tracking loop jitter. This, in turn, results in more noisy range measurements and thus, less precise positioning (DOD, 2001; Petovello, 2009; USACE, 2011; Kaplan & Hegarty, 2017).

All three GPS speed meters are found to have relatively similar speed errors. However, Speed Meter S1 is observed to be able to operate at lower GPS signal power levels as compared to the other speed meters. This could be as it higher receiver sensitivity and lower receiver noise, which allows it track higher C/N_0 levels for the available GPS satellites. While Speed Meter S1 is able to display maximum speed of up to 300 km/h, Speed Meters S2 and S3 are only display maximum speeds of 99.9 and 140 km/h respectively.

Varying speed error patterns are observed for the each of the readings. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over location and time, resulting in GPS accuracy being location / time dependent (DOD, 2001; Huihui *et al.*, 2008; Dinesh *et al.*, 2010; USACE, 2011; Kaplan & Hegarty, 2017).

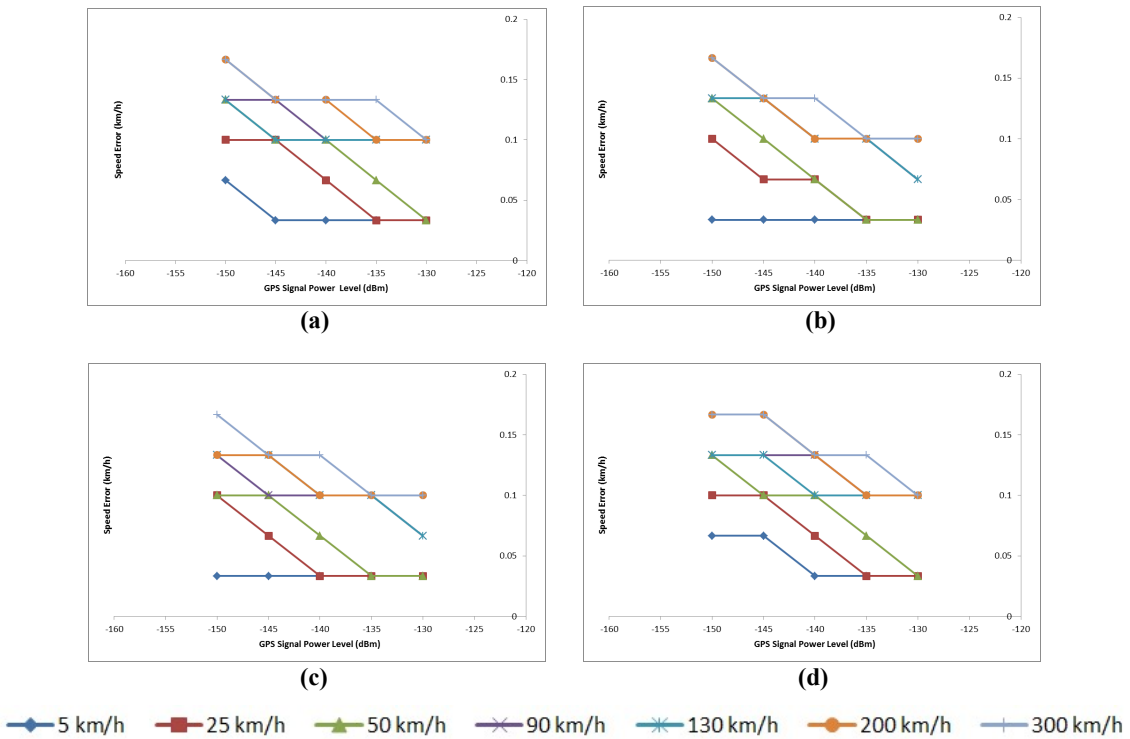


Figure 2: Recorded speed errors for Speed Meter S1 for the open area scenario for UTC times of: (a) 0000 (b) 0300 (c) 0600 (d) 0900.

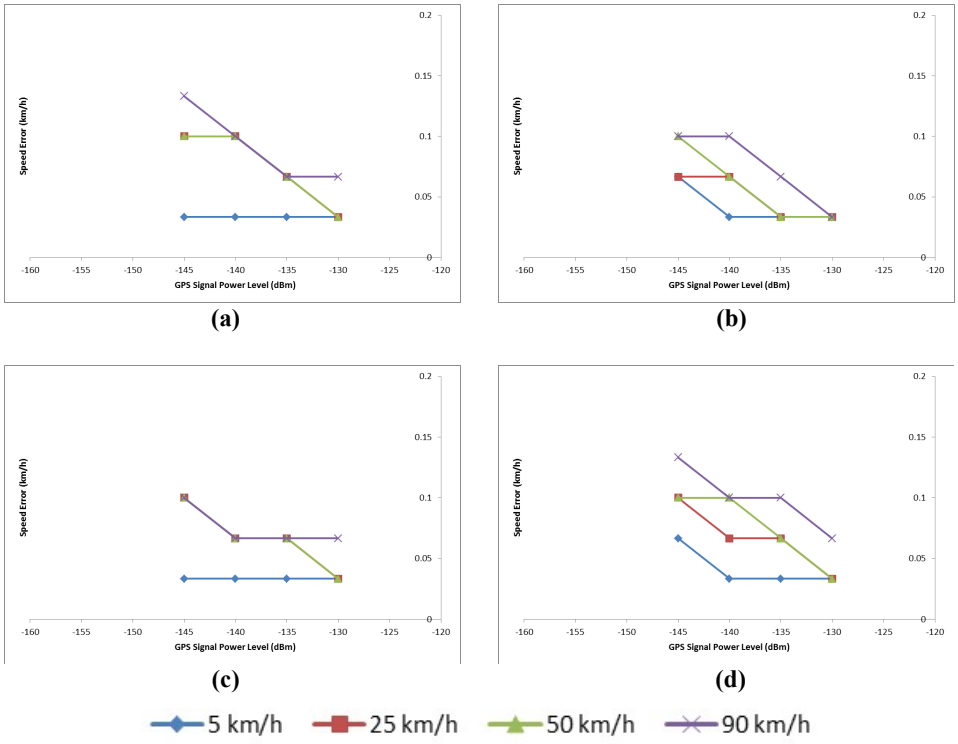


Figure 3: Recorded speed errors for Speed Meter S2 for the open area scenario for UTC times of: (a) 0000 (b) 0300 (c) 0600 (d) 0900.

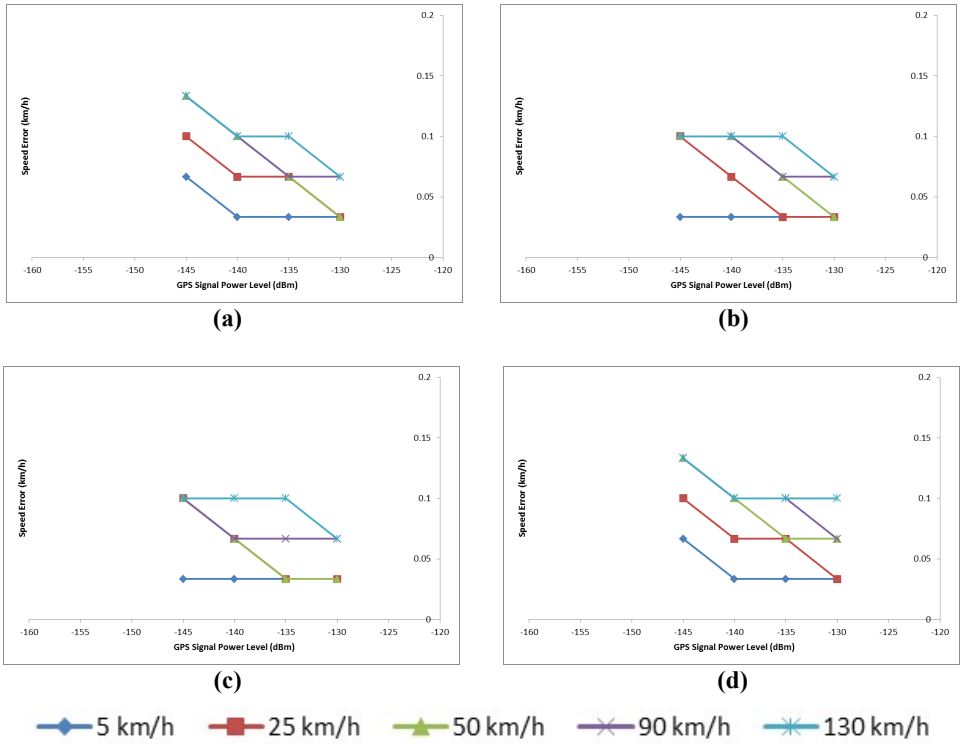


Figure 4: Recorded speed errors for Speed Meter S3 for the open area scenario for UTC times of: (a) 0000 (b) 0300 (c) 0600 (d) 0900.

The speed errors observed in this study are found to be smaller than corresponding errors reported in previous studies (Witte & Wilson, 2004; Huang *et al*., 2013; Steinmetz *et al*. 2014; Alphin *et al*., 2020; Barbosa *et al*., 2022). This is as in this study, the effects of GPS vulnerabilities, such as ionospheric and tropospheric delays, GPS satellite clock and ephemeris error, multipath, and radio frequency interference (RFI), were not considered. Furthermore, the study was conducted for smooth straight line paths with consistent speed. It has been reported that rapid changes of speed and circular paths can degrade the accuracy of GPS speed measurement (Witte & Wilson, 2004; Huang *et al*., 2013; Steinmetz *et al*. 2014; Keskin *et al*., 2018). To this end, further studies need to be conducted to evaluate the effects of various GPS vulnerabilities and movement types on the accuracy of GPS speed measurement in user-controlled scenarios.

It should be noted that the tests conducted in this study were for only three GPS speed meters. Additional tests using a wider range of GPS speed meters are needed to further validate the findings of this study. Furthermore, a limitation faced in this study was that the GPS simulator used only allows the transmission of the GPS L1 C/A signal. The proposed future work is for the procurement of a GNSS simulator that will allow transmission of other GPS signals, in particular L2C and L5, along with signals of other GNSS systems (GLONASS, BeiDou and Galileo).

4. CONCLUSION

In this study, it was found that with decreasing GPS signal power level, speed errors increase due to decreasing C/N_0 levels for GPS satellites tracked by the receiver, which is the ratio of received GPS signal power level to noise density. In addition, varying speed error patterns were observed for the each of the readings. This is due to the GPS satellite constellation being dynamic, causing varying GPS satellite geometry over location and time, resulting in GPS accuracy being location / time dependent. All three GPS speed meters were found to have relatively similar speed errors. However, Speed Meter S1 was observed to be able to operate at lower GPS signal power levels as compared to the other speed meters. This could be as it higher receiver sensitivity and lower receiver noise, which allows it track higher C/N_0 levels for the available GPS satellites.

REFERENCES

Aeroflex (2010). *Avionics GPSG-1000 GPS / Galileo Portable Positional Simulator*. Aeroflex Inc., Plainview, New York.

Alphin, K.L., Sisson, O.M., Hudgins, B.L., Noonan, C.D. & Bunn, J.A. (2020). Accuracy assessment of a GPS device for maximum sprint speed. *Int. J. Exerc. Sci.*, **13**: 273–280.

Aloi, D.N., Alsliety, M. & Akos, D.M. (2011). A methodology for the evaluation of a GPS receiver performance in telematics applications. *IEEE T. Instrum. Meas.*, **56**: 11-24.

Arul Elango, G. & Sudha, G.F. (2016). Design of complete software GPS signal simulator with low complexity and precise multipath channel model. *J. Electr. Syst., Inform. Tech.*, **3**: 161-180.

Akkamis, M., Keskin, M. & Sekerli, Y.E. (2021). Comparative appraisal of three low-cost GPS speed sensors with different data update frequencies. *AgriEng.*, **3**: 423-437.

Barbosa, L.A., Costa, D.C., & Oliveira, H.C. (2022). Evaluation of low-cost GNSS receivers for speed monitoring. *Case Stud. Transp. Policy*, **10**: 239-247.

Bi. Y. & Yuan, J. (2021). A portable GPS signal simulator design based on ZYNQ. 2nd Int. Symp. Comp. Eng. Intell. Comm. (ISCEIC 2021), 6-8 August 2021, Nanjing, China.

CNET (2008). *GPSDiag 1.0*. Available online at: https://download.cnet.com/GPSDiag/3000-2130_4-10055902.html (Last access date: 9 June 2022).

Dinesh, S., Mohd Faudzi., M. & Zainal Fitry, M.A. (2012). Evaluation of the effect of radio frequency interference (RFI) on Global Positioning System (GPS) signals: Comparison of field evaluations and GPS simulation. *J. Defence Secur.*, **5**: 71-86.

Dinesh, S., Shalini, S., Zainal Fitry, M.A., Asmariah, J. & Siti Zainun, A. (2015). Evaluation of the accuracy of Global Positioning System (GPS) speed measurement via GPS simulation. *Defence S&T Tech. Bull.*, **8**:, 121-128.

DOD (Department of Defence) (2001). *Global Positioning System Standard Positioning Service Performance Standard, Command, Control, Communications, and Intelligence*. Department of Defence (DOD), Washington D.C.

Gaglione, S. (20015). How does a GNSS receiver estimate velocity? *Inside GNSS*, **10**: 38-41.

Huang, D. (2013). *Evidential Problems with GPS Accuracy: Device Testing*. Master's dissertation, AUT University, Auckland.

Kamarulzaman, M. (2010). *Technical Specification for STRIDE's Mini-Anechoic Chamber*. Science & Technology Research Institute for Defence (STRIDE), Ministry of Defence, Malaysia.

Kaplan, E.D. & Hegarty, C.J. (2017). *Understanding GPS: Principles and Applications*. Artech House, Norwood, Massachusetts.

Keskin, M., Akkamis, M. & Sekerli, Y.E. (2018). An Overview of GNSS and GPS based velocity measurement in comparison to other techniques. *Int. Congr. Ener. Res.*, 31 October - 2 November 2018, Alanya, Turkey.

Kou, Y. & Zhang, H. (2011). Verification testing of a multi-GNSS RF signal Simulator. *Inside GNSS*, **6**: 52-61.

Magene (2021). *Magene C406 Smart GPS Bike Computer*. Qingdao Magene Intelligence Technology Co. Ltd. China.

Pozzobon, O., Sarto, C., Chiara, A.D., Pozzobon, A., Gamba, G., Crisci, M. & Ioannides, R. (2013). Developing a GNSS position and timing authentication testbed: GNSS vulnerability and mitigation techniques. *Inside GNSS*, **8**: 45-53.

SkyRC (2021). *GSM020 Instruction Manual. SkyRC* Technology Co. Ltd., Shenzhen, China.

Steinmetz, E., Jarlemark, P., Emardson, R., Skoogh, H. & Herbertsson, M. (2014). Assessment of GPS derived speed for verification of speed measuring devices. *Int. J. Instr. Tech.*, **1**: 212-227

USACE (US Army Corps of Engineers) (2011). *Engineer Manual EM 1110-1-1003: NAVSTAR Global Positioning System Surveying.* US Army Corps of Engineers (USACE), Washington D.C.

USCG (US Coast Guard) (2022). *GPS NANUs, Almanacs, & Ops Advisories*. Available online at: https://www.navcen.uscg.gov/gps-nanus-almanacs-opsadvisories-sof (Last access date: 8 June 2022).

Witte, T.H. & Wilson, A.M. (2004). Accuracy of non-differential GPS for the determination of speed over ground. *J. Biomech.*, **37**: 1891-1898.

XOSS (2022). *XOSS G+ User Manual*. XOSS Hong Kong Co. Ltd., Hong Kong.

Zhang, J., Zhang, K., Grenfell, R. & Deakin, R. (2006). On the relativistic Doppler effect for precise velocity determination using GPS. *J. Geodesy*, **80**: 104-110.

A WORKFLOW TO DEVELOP AND IMPLEMENT AN E-HEALTH INFORMATION SYSTEM IN WAR-TORN COUNTRIES: A CASE STUDY IN IRAQI KURDISTAN

Gorgees Akhshirsh[1,2], Bayar Azeez[1,2], Antonia Bezenchek[3,4], Iuri Fanti[3], Shahla O. Salih[5,6], Faiq B. Basa[1,7], Andrea Malizia[1], Stefania Moramarco[1*] & Leonardo Emberti Gialloreti[1]

[1]Department of Biomedicine and Prevention, University of Rome Tor Vergata, Italy
[2]Computer Systems Engineering, University of Kurdistan – Hawler, Iraq
[3]Informa-PRO, Italy
[4]EuResist Network, Italy
[5]Department of Civil Engineering and Computer Science Engineering, University of Rome Tor Vergata, Italy
[6]Department of Statistics and Informatics, University of Sulaimaniya, Iraq
[7]Rizgary Teaching Hospital, Iraq

[*]Email: stefania.moramarco@gmail.com

ABSTRACT

Conflicts and terrorism, especially when protracted, can deeply debilitate countries' security and safety, with multidimensional impact even on the public healthcare systems. The long-term effects can last for years after the cessation of emergencies, with health data not available and / or not fully reliable, causing targeted health interventions to be almost non-existent. Despite health information systems (HIS) being paramount in contributing to national security by guiding public health decision-makers, policy formulation, resource allocation and quality control, many Middle East countries, especially when they are faced with security instability, at present still do not collect electronic records. As a case study from the field, we describe the workflow - development, implementation, challenges and lessons learned - to create, maintain and advance a HIS in the Iraqi Kurdistan, a war-torn region in the Middle East. After a pilot phase, in 2018, a HIS based on the open-source software District Health Information System 2 (DHIS2) was set up in the region. It collects diseases registered in public health facilities and health data coded using the international WHO nomenclature ICD-10. The HIS was adapted to the local scenario, with user interfaces provided in Arabic and Kurdish-Sorani languages. The Pentaho Data Integration tool was used to effectively automate the process of data integration and bulk import from local systems already in use. The aim of this study is to provide lesson learned from the field to support evidence-based public health decisions even in other war-torn countries.

Keywords: *Public health; epidemiological surveillance; e-health; electronic records; District Health Information System 2 (DHIS2)*

1. INTRODUCTION

In the recent years, the traditional meaning of security has moved to wider scopes, including global health coverage. As a matter of fact, in 2014, the Global Health Security Agenda (GHSA) has been endorsed at both national and international levels to accelerate progress toward a world safer and securer through promotion of health. Therefore, global security has become a worldwide priority (GHSA, 2022). This implies that more acceptable, tangible and mainstream interpretations of national security exist. This is particularly the case for war-torn countries or countries emerging from wars, where years of conflict have deeply debilitated public health systems, with long-term effects on populations' health, which last for years after the cessation of emergencies (Gialloreti *et al.*, 2020).

Electronic health information systems (HIS) are paramount for guiding public health decision-makers in policy formulation, resource allocation and quality control (WHO, 2014a). Electronic health records (EHR) also entail demographic, social, political, and economic advantages; however, registration systems are not

implemented at scale (WHO, 2021a). As a result, it becomes challenging to generate accurate data on even the most basic health indicators (Mahapatra *et al*., 2007; AbouZahr *et al*., 2007). As an example, Iraq is suffering from what has been defined as the "single most critical failure of development over the past 30 years" (Setel *et al*., 2007), i.e., lack of demographic and epidemiological information, such as vital statistics on birth, disease, mortality, and cause of death. As reported by Asaad *et al.* (2020), civil registration systems in Iraq have never been formally evaluated until 2012, when the World Health Organization (WHO) requested an assessment. The evaluation showed a system malfunction, particularly concerning the completeness of birth and death registrations and the causes of death. The WHO offered some recommendations to improve the quality of the system. One of them was to envisage the computerization of the system as the only way to produce accurate and consistent data.

In 2005, the new Constitution of Iraq recognized the Autonomous Region of Iraqi Kurdistan (KRI) and its Kurdistan Regional Government (KRG) as part of the Republic of Iraq. In KRI, health service delivery and health financing mix public-private participation and investments. The Ministry of Health (MoH) of the KRI government follows the Iraqi MoH's basic organizational structure and system. The KRG decides health policies implemented by MoH (Shabila *et al*., 2010).

The current health situation of the KRI population is still not well known. Health data is collected sporadically and usually only in aggregated form; the available information is generally inferred by means of patchy surveys' estimations (RAND Corporation, 2014a). The consequence is a paucity of reliable health statistics and epidemiological surveillance being almost non-existent, limiting decision-makers considerations to guide the country. Health policies are often based on insufficient evidence (Lopez & Setel, 2015), while the impact of the provided care is not efficiently monitored or evaluated (EMRO, 2006; Webster, 2011). Therefore, much effort in data registration and evaluation is needed to guarantee a well-functioning public health system in Iraq and ensure universal health coverage (UHC) (WHO, 2014a). WHO is already supporting regional initiatives in the Eastern Mediterranean Region to develop national HIS and foster progress in expanding digital public health (WHO, 2014b; Murray *et al*., 2020).

After a two-year pilot phase, in 2018, the operational phase of a project for developing a health information system in Iraq was launched (Emberti Gialloreti *et al*., 2020). This case study describes the workflow - development, implementation, challenges, and lessons learned - to create, maintain and advance an e-health tool setup in the Primary Health Centers (PHC) and Public Hospitals (PH) of the Iraqi Kurdistan (KRI) to develop a digital health monitoring and epidemiological surveillance system.

2. MATERIALS AND METHODS

2.1 Conceptual Design of the Required Statistics

In the KRI, public health data - if collected - is mainly paper-based. Hence, format, accuracy, completeness, and accessibility of information are among the main challenges when processing health statistics. Data collected in PHC or PH - when present - are only in aggregated form. In order to overcome the challenge of aggregated data, an informatic system was set up to collect all the individual diseases registered during each diagnostic examination. The main advantage of this approach is that different levels of aggregation can be applied according to the requirements (Sahay *et al*., 2019). The system also collects data on births, deaths, and vaccinations.

2.2 Choice of the Suitable Tool

A free and open-source software platform for collecting, managing, analyzing and using data was employed, which is the District Health Information System 2 (DHIS2) (DHIS2, 2019). DHIS2 is among the world's largest health management information system platforms used by 72 low- and middle-income countries (Sahay *et al*., 2019). It adopts the WHO International Classification of Diseases as the primary standard for data reporting (Rashidian, 2019). The DHIS2 platform was chosen because it allows users to enter data directly from the periphery on the central servers, using only a web browser or a mobile app on even slow or

discontinuous internet connections. It also shows real-time statistics about entered data and supports external app development. Security and privacy are intrinsic to the DHIS2 open-source software. Furthermore, data security is guaranteed by the deployment of providers certified by the University of Oslo, which carries out a scheduled vulnerability assessment and penetration tests.

DHIS2 offers three different approaches for data acquisition (Sahay *et al*., 2019):

a) Aggregated data acquisition: It allows the quick collection of bulk data but without the possibility of generating more detailed statistics later.
b) Events acquisition with the person's registration: This approach stores data with the maximum granularity. However, in KRI, there is still no citizens' unique ID code.
c) Event data acquisition without registration: It allows the registration with the maximum granularity while allowing aggregations later and without privacy concerns for the patients.

In order to decide which approach was more suitable for this HIS, some of these approaches were tested using a small subset of health facilities. Eventually, the chosen implementation was based on the acquisition of event data without registration, where the recording of patients' data is not needed. Events are recorded using the DHIS2 Capture App.

A folder ID connected to the patient's folder in the specific hospital / health center (Health Unit) is introduced to avoid multiple registrations of the same events. This code is unique for the health unit and can be connected to the patient's personal data information only by the physicians of the health unit where the patient is visited and is treated. For all other users of the platform, the patient is anonymized.

2.3 Choice of the Appropriate Paradigm of DHIS2

Before choosing the most suitable paradigm, several prototypes were created based on different models. This was essential to test and compare the obtained output and to evaluate if it tallies with our main requirements as follows:

a) Statistical results used for decision making.
b) Easiness to enter data even with slow or discontinuous internet connections.
c) Flexibility to create different views starting from the same raw data.
d) Possibility to check and correct the entered data in case of errors.
e) Possibility to expand some metadata during later stages and not only during planning.

2.4 Setting up the HIS Skeleton

Five main statistical topics were chosen, and five DHIS2 programs were created: Births, Immunizations, Disease Surveillance (diseases diagnosed at health centers), Hospital Discharges (diseases diagnosed in hospitals), and Deaths.

2.5 Identification of the Nomenclature and Creation of the Local Language Version

The international WHO nomenclature was used to code the health events, i.e., the *International Statistical Classification of Diseases and Related Health Problems, 10th Rev. (ICD-10)* (WHO, 2016). It is recognized as the global standard for health data, clinical documentation, and statistical aggregation. In order to adapt the system to the local scenario—the language spoken in Iraqi Kurdistan is mainly Kurdish-Sorani, with Arabic often used when dealing with medical terms—the first step was to make the system usable for people who do not understand English. The required elements for the translation in Kurdish and Arabic languages are presented in Table 1.

Table 1: Required elements when translating and adapting the HIS into non-Latin-script alphabets.

Development stage	Target	Elements and challenges of the stage
1. Database language	Developed specifically for this HIS	# Labor intensive: all the names of the construction. elements—data elements, programs, option sets, etc.—had to be translated into Kurdish and Arabic. # The implementation team could define it without modifying the basic software platform.
2. Interface language	Developed to implement it in the software platform	# Usable only if the language is already present in the software platform. # Arabic was already present in the DHIS2 implementation, but not Kurdish-Sorani. # All interface strings had to be translated into Kurdish. # The DHIS2 team was requested to add Kurdish in both the software platform and the translation system. # After this intervention, a Kurdish personalized translation was created. # The Java developers were requested to add the Kurdish-Sorani language to the underlying Java library since it was not present. # After this intervention, all strings were translated for the interface and database elements.
3. Diagnosis coding into ICD-10 and translation	Developed to provide support to the local staff performing data entry	# Training of medical doctors about ICD-10 disease coding. # Training of data entry personnel about the electronic use of ICD-10 codes. # Setting upcoded disease lists tailored for each specific data entry unit.

2.6 Training and Interaction with the Local Personnel

Before a new center started working on the system, the local health authority had to equip the health centers with hardware and internet connections fully. This step was also essential to engage the local authorities to participate in the process, guaranteeing the future sustainability of the project. Subsequently, the local staff was trained in coding and data entry (Table 1). The training sessions were held in each KRI province through plenary meetings with presentations and frontal lessons. Afterwards, on-the-job training was conducted at each unit's facility (PHC or PH) for one week while upgrading the IT infrastructure. User's manuals were provided in the two languages used (Arabic and Kurdish-Sorani). The main activities for setting up a new data entry unit into the HIS are summarized in Figure 1.

2.7 Importing Bulk Data from Hospital Databases

Since PHC did not have any information system before this project, the system had to be built from scratch. Contrariwise, some PH already had their system, which was usually very basic (usually a simple spreadsheet) and without the possibility to interact with other systems beyond the hospital itself. An ad-hoc tools were developed to import data into the HIS since DHIS2 is interoperating with the databases and other systems using application programming interface (API) integration and bulk data import. Since PH's existing electronic systems were different from hospital to hospital, a single logic for API integration and data import process / flow could not be applied. Therefore, the process had to be dealt with each dataset as a separate case, and for each flow, a different logic had to be applied to accomplish the process. The main challenges of this process were quality of data received, records with blank fields (i.e., date, age, diagnosis), records with the wrong disease diagnosed field, lack of existing translations for Arabic and Kurdish diseases mapping of ICD-10 codes, and diseases diagnosed fields in Arabic and Kurdish having different spellings. The Pentaho Data Integration (PDI) tool was used to automate data integration and bulk import effectively (HITACHI, 2022).

PDI provides extract, transform and load (ETL) capabilities, making acquiring data using different sources and their subsequent cleaning and transformation faster and more consistent by using a uniform format. In our case, PDI was used for the data migration from different local application databases to our centralized KRI region Health Monitoring System, the KRG-HIS. A screenshot of a Pentaho elaboration processing for one hospital, with the different phases of data transformation and checks, is shown in Figure 2.

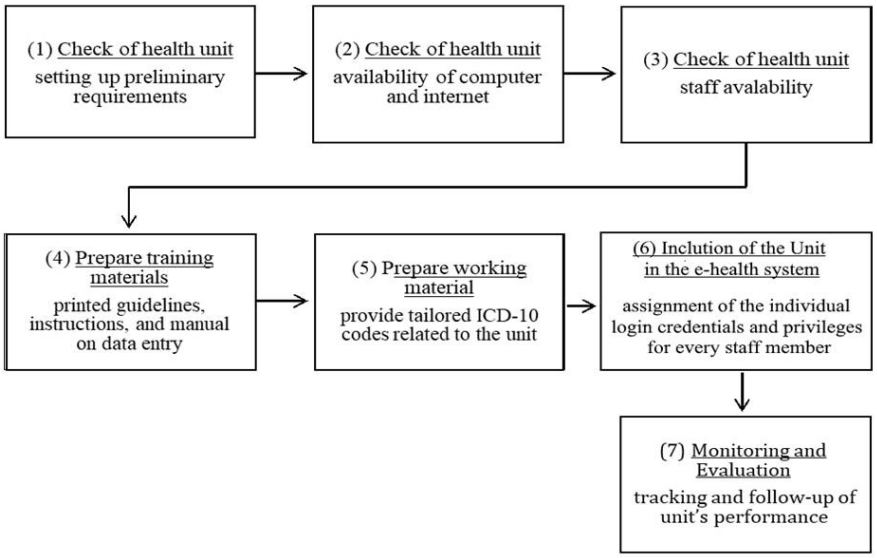


Figure 1: Flowchart for setting up a data entry unit into the HIS.

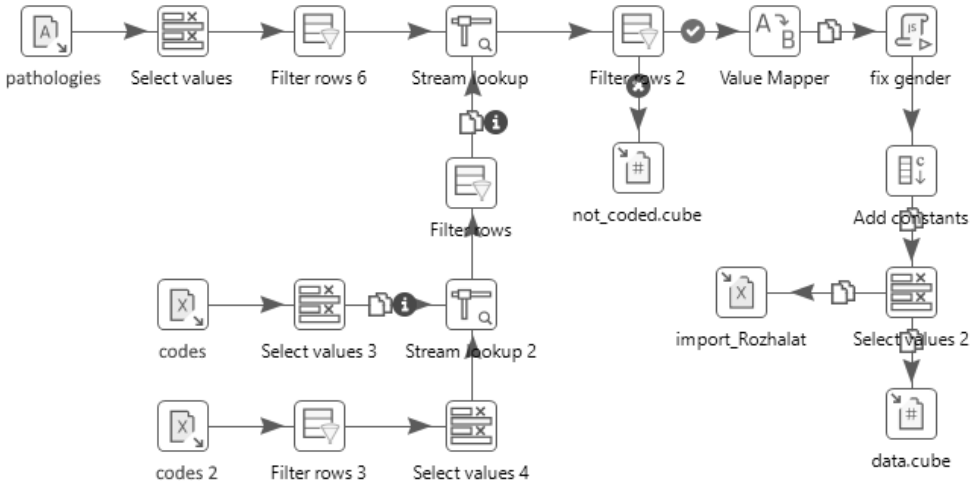


Figure 2: Data transformation and import screenshot from the Pentaho tool for the Rozhalat Hospital.

The steps taken for the data import from each hospital were: (1) Manual analysis, cleaning, and preparation of the data; (2) Creation of a program / job to perform data extraction, translation, mapping and importing into the DHIS2 platform (Figure 3).

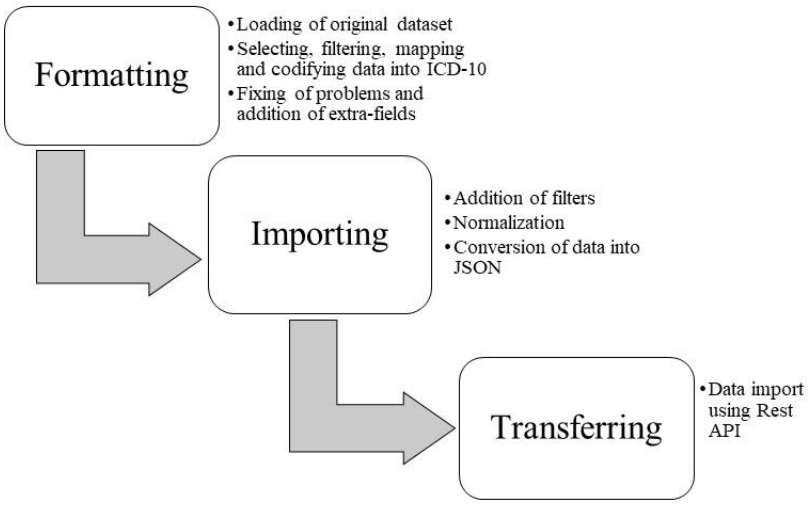


Figure 3: Data import activities flowchart.

2.8 Data Check and Statistics Generation

The caseload of each unit is checked daily. Any suspicious case or unexpected change in the regularity of cases entered can be rapidly detected, and inconsistent data is monitored. Each center's staff has to be trained to be accurate and focused on entering data regularly and correctly into the system. As in the DHIS2 software, the user statistics are limited, and there are no internal tools to evaluate the user's activities. The entered events were regularly downloaded, and statistics were performed using external data processing. A rapid question / answer mechanism was created by setting up communication groups between the operational centers, local teams, and international experts (using Viber, WhatsApp, and other messaging tools).

2.9 Request for the Features to the DHIS2 Platform Developers

DHIS2 constantly evolves thanks to a large developer community coordinated by HISP at the University of Oslo (Dias, 2020). During the implementation and use of the HIS, various challenges on the DHIS2 software were found and reported to the developers, and new features were requested. For example, a request has to be made to develop a dimension to create different groupings of ICD-10 diagnoses used for classification / aggregation of the event diagnosis by ICD-10 Chapters (I-XXII), ICD-10 letters (A-Z), WHO Global Health Estimates, (WHO, 2021b), or other aggregations based on specific disease groups.

2.10 Ethical Considerations

The HIS was developed following the national and regional laws. The *WHO Guidelines on Ethical Issues in Public Health Surveillance* (WHO, 2017a) was followed. At this stage, the institutional review board of the University of Rome Tor Vergata (Italy) waived any requirements for further ethical approval.

3. RESULTS

The pilot phase of the HIS started in 2016, while the operational phase was activated in 2018. Figure 4 shows the login user interface and system dashboard.

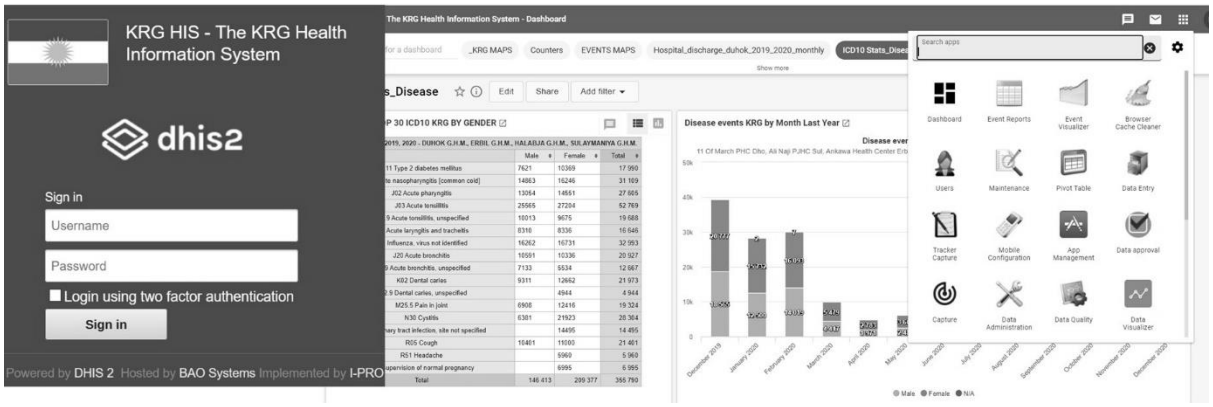


Figure 4: Login user interface and system dashboard.

By the end of December 2021, 128 health centers (94 PHC, 30 PH, and four Registration Bureau of Births and Deaths-RBBD have been included and are active in the HIS (Table 2), covering at present nearly the 50% of the overall public health facilities of the area.

Table 2: Cumulative distribution of the health units activated by year.

Governorate	2016	2017	2018	2019	2020	2021	Health Units Type		
							PHC	PH	RBBD
Duhok	7	10	11	27	35	58	46	11	1
Hawler	9	10	11	17	17	23	17	6	-
Halabja	1	1	1	3	4	8	3	3	2
Sulaimaniya	10	10	10	16	16	39	28	10	1
Totals	**27**	**31**	**33**	**63**	**72**	**128**	**94**	**30**	**4**

The system gathers more than 1,200,000 disease events from the PHC and about 370,000 from PH. The 15 most common diagnoses gathered within the PHC since the beginning of the project are shown in Table 3. They cover almost half of all accesses to the PHCs (Emberti Gialloreti *et al.*, 2020).

Table 3: Top 15 ICD-10 diagnoses gathered in the HIS within the PHC.

Diagnosis (ICD-10)	Number of events
Acute tonsillitis (J03)	79,384
Influenza, virus not identified (J11)	60,913
Acute nasopharyngitis -common cold (J00)	57,732
Cystitis (N30)	36,743
Acute pharyngitis (J02)	39,225
Dental caries (K02)	24,754
Cough (R05)	37,982
Acute bronchitis (J20)	26,228
Acute tonsillitis, unspecified (J03.9)	51,651
Pain in joint (M25.5)	27,957
Type 2 diabetes mellitus (E11)	35,732
Acute laryngitis and tracheitis (J04)	22,474
Urinary tract infection, site not specified (N39.0)	51,305
Acute bronchitis, unspecified (J20.9)	33,256
Supervision of normal pregnancy (Z34)	8,450
Totals	**593,786**

3.1 Tailoring Structural Elements Based on The Health Units' Feedback

As feedback from the units began to flow in, functional enhancement and additions to the initial programs were detected. For example, for some Emergency Departments, it was necessary to introduce an "Admission Mode" in the "Hospital Discharges" program. Overall, the program that received the most feedback for changes was the "Births" program. Before the project implementation, basic information was received from health managers about the required statistics and analyses. However, after starting the data acquisition, the features which were more meaningful and enforced could be evaluated.

3.2 Developing A Culture for Data Actions

Seminars and training sessions have been conducted to increase the overall process among the public health personnel. By 2020, 258 people were trained: 142 medical doctors, 53 administrative staff, 36 nurses, 12 statisticians, seven information scientists, six public health specialists, and two pharmacists.

Furthermore, during these years, 734 medical doctors, nurses, statisticians, and public health officials in the region have been trained on public health, epidemiological surveillance and the DHIS2 system. A key aspect of the program is to guarantee the project's continuity and sustainability by establishing a team of highly specialized experts to direct the entire system once the local authorities manage it. In order to broaden the local team of experts, for the academic year of 2018 / 2019, six PhD positions have been granted, related to the project itself: three in *Nursing Sciences and Public Health* and three in *Computer Science, Control and Geoinformation*, under the University of Rome Tor Vergata, Italy.

4. DISCUSSION

Reliable and timely health information is one of the six building blocks of a health system (WHO, 2007). It is essential for guiding public health decision-makers and policy formulation for allocating resources according to actual health priorities. Including data on all the vital statistics in the setting-up of the HIS is a priority for timely identification of health requirements (WHO, 2014a; ESCAP, 2019). In the era of big data, data collection is just the initial challenge, which should be followed by the need to make good use and sense of what has been collected (Hazel *et al*., 2018).

Alongside the direct and short-term effects of war, conflicts and terrorism, the long-term effects could have devastating consequences for public health, such as the erosion of countries' ability to target health interventions and to be prepared for future emergencies (Moramarco *et al*., 2020).

Being a war-torn country, one of the Iraq's main priorities is reconstructing a full-fledged public health system based on reliable and complete health data (WHO, 2017b). In recent years, the Iraqi Ministry of Health has endeavored to employ e-health to support the healthcare sector in the country. However, the implementation plan is still in the preliminary phase (Jaber *et al*., 2014). Nevertheless, some Iraqi health authorities declared that the vision of building an efficient healthcare system throughout the development of e-health is a key element for modern policy development, decision-making, and regulatory oversight (MoPKRG, 2013). Several studies have highlighted that there is an urgent need to develop an integrated HIS to support KRI policymakers in public health (Anthony *et al*., 2013).

The choice of the DHIS2 platform met most of the core requirements for the local settings. A highly specialized local team has been established to "train the trainers" and ensure the system's self-sufficiency in the future. The public health and engineering specialists trained under two PhD programs should guarantee the project's future continuity and sustainability with hundreds of other trained operators (RAND, 2014b). Continuous supervision of the centers and constant dialogue with key stakeholders, which consider the specific contextual requirements, are paramount elements to implement HIS successfully (Fennelly *et al*., 2020).

By the beginning of 2022, the system has covered at least half of all primary health centers, family health centers and public hospitals of the region. KRG-HIS has the potential to become one of the broadest sentinel surveillance systems in the Middle East, since it will provide high-quality information needed for national and sub-national planning, policy implementation, as well as monitoring health outcomes and services. The future scale-up to overall Iraq will be essential to support the rebuilding and reorganization of an efficient public health system.

However, it should be noted that in 2020, the COVID-19 pandemic partially disrupted the routine KRG-HIS delivering system. Almost all of the PHC clinical activities (including data entry) were put on hold. Some PHCs were converted into COVID-19 specialized centers, while PHs with Intensive Care Units were identified as COVID-19 hospitals (Stefania *et al*., 2020). Nevertheless, by December 2020, most of the PH and PHC have restarted collecting data.

At present, data in the system is currently captured only by event and cannot be linked to a specific person due to the lack of unique identity numbers (ID) for each citizen. Nevertheless, the software is already prearranged to include ID. Since the system has been set in an unstable geopolitical area, the future local and international scenarios will play essential roles in the sustainability and scaling up of the project.

5. CONCLUSION

Global health security is an international priority that must undergo planning and resource mobilization to address gaps, implement activities and scale-up programs to achieve impact. Within its scope, the rebuilding of an effective healthcare system in war-torn countries is a long-term process that requires multiple actions and actors, with epidemiological surveillance being the cornerstone. A functioning e-health system for epidemiological surveillance is a key instrument to managing complex and fluid health situations in the medium and long term while preparing to respond and cope with future emergencies.

The presented case study on KRG-HIS reports challenges and lessons learned in developing a tailored tool for e-health in a war-torn region, the Iraqi Kurdistan, offering suggestions to software developers and public health actors in other similar regions who aim to build data systems that support evidence-based decision-making.

REFERENCES

AbouZahr, C., Cleland, J., Coullare, F., Macfarlane, S.B., Notzon, F.C., Setel, P. & Szreter, S. (2007). The way forward. *Lancet*, **370:** 1791–1799.

Anthony, C.R., Hansen, M.L., Kumar, K.B., Shatz, H.J. & Vernez, G. (2013). *Building the Future: Summary of Four Studies to Develop the Private Sector, Education, Health Care, and Data for Decision making for the Kurdistan Region - Iraq*. RAND Corporation, California, US.

Asaad, A. M., Lami, F., Khaleel, H.A., Assi, W.S. & Ahmed, W. (2020). Results of the rapid assessment of civil registration and vital statistics in Iraq, 2012. *Can. Stud. Popul.*, **47**: 183–193.

DHIS2 (2019). *District Health Information Software 2 (DHIS2)*. Available online at: https://dhis2.org (Available online at: 20 January 2020).

Dias, B. (2020). *News: Roux Prize Awarded to Kristin and Jørn Braa*. Available online at: https://dhis2.org/roux-prize (Last access date: 17 July 2022).

Emberti Gialloreti, L., Basa, F.B., Moramarco, S., Salih, A.O., Alsilefanee, H.H., Qadir, S.A., Bezenchek, A., Incardona, F., di Giovanni, D., Khorany, R., Alhanabadi, L.H.H., Salih, S.O., Akhshirsh, G.S., Azeez, B.S., Tofiq, B.A. & Palombi, L. (2020). Supporting Iraqi Kurdistan health authorities in post-conflict recovery: The development of a health monitoring system. *Front. Public Health*, **8:** 2296-2565.

ESCAP (2019). *Scoping Mission on Improvement of Civil Registration System Based Vital Statistics*. Available online at: https://www.unescap.org/events/scoping-mission-improvement-civil-registration-system-based-vital-statistics (Last access date: 25 February 2020).

Fennelly, O., Cunningham, C., Grogan, L., Cronin, H., O'Shea, C., Roche, M., Lawlor, F. & O'Hare, N. (2020). Successfully implementing a national electronic health record: a rapid umbrella review. *Int. J. Med. Inform*., **144**: 104281.

Gialloreti, L.E., Moramarco, S. & Palombi, L. (2020). Investing in epidemiological surveillance for recovering health systems in war-torn countries. *Perspect. Public Health*, **140:** 25–26.

GHSA (Global Health Security Agenda). *Global Health Security Agenda*. Available at: https://ghsagenda.org (Last access date: 17 July 2022).

Hazel, E., Wilson, E., Anifalaje, A., Sawadogo-Lewis, T. & Heidkamp, R. (2018). Building integrated data systems for health and nutrition program evaluations: lessons learned from a multi-country implementation of a DHIS 2-based system. *J. Glob. Health*, **8:** 1-5.

HITACHI. (2022). Pentaho Data Integration. Hitachi Vantara Lumada and Pentaho Documentation. Available online at: https://help.hitachivantara.com/Documentation/Pentaho/7.1/0D0/Pentaho_Data_Integration (Last access date: 29 September 2022).

Jaber, M., Abd Ghani, M.K. & Suryana, N. (2014). A review of adoption of telemedicine in middle east countries: Toward building Iraqi telemedicine framework. *Sci. Int.*,26: 1795-1800.

Lopez, A.D. & Setel, P. W. (2015). Better health intelligence: a new era for civil registration and vital statistics? *BMC Med*, **13**: 73.

Mahapatra, P., Shibuya, K., Lopez, A.D., Coullare, F., Notzon, F.C., Rao, C. & Szreter, S. (2007). Civil registration systems and vital statistics: successes and missed opportunities. *Lancet*, **370:** 1653–1663.

Ministry of Planning Kurdistan Regional Government. (2013). *Kurdistan Region of Iraq 2020 A Vision for the Future*. Ministry of Planning, Kurdistan Regional Government, Iraq.

Moramarco, S., Palombi L., Basa F.B., Emberti Gialloreti L. (2020). The multidimensional impact of CBRNe events on health care in the Middle East: the role of epidemiological surveillance in the long-term recovery of public health systems. *Defence S&T Tech. Bull*; **13**: 162-165.

Murray, C. J. L., Alamro, N. M. S., Hwang, H., & Lee, U. (2020). Digital public health and COVID-19. *Lancet Public Health*, **5:** 469–470.

RAND Corporation (2014a). *Capacity Building at the Kurdistan Region Statistics Office Through Data Collection*. RAND Corporation, Santa Monica, California

RAND Corporation. (2014b). *The Future of Health Care in the Kurdistan Region — Iraq Toward an Effective, High-Quality System with an Emphasis on Primary Care*. RAND Corporation, Santa Monica, California

Rashidian, A. (2019). Effective health information systems for delivering the Sustainable Development Goals and the universal health coverage agenda. *East. Mediterr. Health J.*, **25:** 849–851.

EMRO (Regional Health Systems Observatory) (2006). *Regional Health Systems Observatory- EMRO, WHO*. Available online at: https://apps.who.int/medicinedocs (Available online at: 21 January 21, 2022).

Sahay, S., Rashidian, A. & Doctor, H.V. (2019). Challenges and opportunities of using DHIS2 to strengthen health information systems in the Eastern Mediterranean Region: A regional approach. *Electron. J. Inf*, **86**: e12108.

Setel, P.W., Macfarlane, S.B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S. & AbouZahr, C. (2007). A scandal of invisibility: making everyone count by counting everyone. *Lancet*, **370:** 1569–1577.

Shabila, N.P., Al-Tawil, N.G., Tahir, R., Shwani, F.H., Saleh, A.M. & Al-Hadithi, T.S. (2010). Iraqi health system in Kurdistan region: medical professionals' perspectives on challenges and priorities for improvement. *Confl. Health*, **4:** 19.

Stefania, M., Alsilefanee, H.H., Qadir, S.A., Salih, S.O., Alhanabadi, L.H., Basa, F.B., Leonardo, P. & Leonardo, E.G. (2020). COVID-19 and Iraqi Kurdistan: A regional case in the middle east. *Disaster Adv.*, **13:** 14-16.

Webster, P. C. (2011). lraq's health system yet to heal from ravages of war. *Lancet*, **378:** 863–866.

WHO (World Health Organization) (2007). Everybody's business -- strengthening health systems to improve health outcomes: WHO's framework for action. Available online at: https://apps.who.int/iris/handle/10665/43918 (Last access date: 17 July 2022).

WHO (World Health Organization) (2014a). *Providing Health Intelligence to Meet Local Needs: A Practical Guide to Serving Local and Urban Communities Through Public Health Observatories*. World Health Organization (WHO), Geneva, Switzerland.

WHO (World Health Organization) (2014b). *Regional Strategy for the Improvement of Civil Registration and Vital Statistics Systems 2014–2019*. Available online at: https://apps.who.int/iris/handle/10665/123413 (Last access date: 16 January 2022).

WHO (World Health Organization) (2016). *International Statistical Classification of Diseases and Related Health Problems, 10th Rev. (ICD-10)*. Available online at: https://icd.who.int/browse10/2016/en (Last access date: 20 January 2020).

World Health Organization. (2017a). *WHO Guidelines on Ethical Issues in Public Health Surveillance*. Available online at: https://apps.who.int/iris/handle/10665/255721 (Last access date: 17 July 2022).

WHO (World Health Organization) (2017b). *Iraq Health Profile 2015*. Available online at: https://apps.who.int/iris/handle/10665/254984 (Last access date: 17 July 2022).

WHO (2021a). Universal health coverage (UHC). Available online at: https://www.who.int/news-room/fact-sheets/detail/universal-health-coverage-(uhc) (Last access date: 16 January 16, 2022).

WHO (World Health Organization) (2021b). Health Statistics and Information Systems: Global Health Estimates (GHE). Available online at https://www.who.int/healthinfo/global_burden_disease/en (Last access date: 2 April 2021).

A COMPUTATIONAL MODEL OF HUMAN-ROBOT COLLABORATION TRUST AND ITS APPLICATION IN SIMULATED OPERATIVE DOMAIN

Wadhah A. Abdulhussain & Azizi Ab Aziz*

Relational Machines Group, Human-Centred Computing Lab, School of Computing, Universiti Utara Malaysia (UUM), Malaysia

*Email: aziziaziz@uum.edu.my

ABSTRACT

Many years back, robots were fundamentally designed and separated from the human environment to prevent humans from getting too close and to eliminate possible risks. However, now some technologies are being made available for humans to work closely in the same area as robots. These robots are likely to operate as substitute humans / experts, where they will be employed to increase the efforts of, or stand in for, humans. In order to accomplish this goal, one of the essential precursors is a robot to gain the trust of human team members. This article presents a computational analysis of human-robot collaboration trust within a simulated military operative domain. In this study, simulation experiments under various parameter settings indicate that the model can generate reasonable behaviour of distinct types of chosen cases. Moreover, through equilibria analysis, the model's stability state has been established, and by automated checking, the model's fundamental empirical-based properties have been confirmed.

Keywords: Collaborative robot simulation; human-robot-in-the-loop; computational cognitive modelling; military operative domain; trust modelling.

1. INTRODUCTION

In the past few years, computer scientists and engineers have developed special robots to work closely with humans (human-robot-in-the-loop) that have gained momentum in real-world applications. The main idea of human-robot-in-the-loop refers to the process of combining robot and human intelligence to obtain the best results in the long term. This impetus has encouraged major researches on human-robot collaboration worldwide. One of the underlying research themes deals with the challenging questions of trust for human-robot joint tasks. Trust is a substantial factor that needs to be taken into attention when robots are going to work together with humans as functional teammates (Lewis *et al*., 2018). This concept can be perceived as the main element defining how much a robot would be recognised and managed by the human. The wide range of perspectives within which trust has been examined leads to numerous descriptions and theories of trust. For instance, throughout the interpersonal (psychology) related literature, trust is also viewed as encompassing affective processes since trust development involves seeing others as personally motivated by care and concern to safeguard the trustor’s interests. When it comes to human-robot collaboration, trust plays an essential role in conserving trustworthiness, a requirement for such robots to be trusted by humans to perform the assigned tasks. The preservation of trust provides a factor of its motivational activities for humans to participate actively in collaborative tasks (Javaid *et al*., 2020). For instance, when individuals work closely with a collaborative robot to reach the intended goal, trust is vital in improving belief in following the robot's suggestions. While humans may trust a robot to replace brooms at home, there are still some concerns about certain personal aspects due to possible distrust in specific contexts. For example, a vacuum robot malfunctioning (e.g., a partially cleaned room) or being trapped under a sofa may not have an enormous impact on daily living. However, having a similar malfunction in high-risk tasks (e.g., self-driving cars or surgery robots) can produce catastrophic outcomes. Regardless of the

risk, robotic technologies and other autonomous systems offer possible advantages by supporting individuals in achieving their missions (Boos & Moshkina, 2019).

In general, trust has been investigated in various areas (including human factors, social psychology, and system automation) to identify possible connections between humans and machines. The wide range of contexts within which trust has been researched leads to several representations and perceptions of trust. Throughout this paper, our fundamental objective is to demonstrate the high-level concept of a computational human-robot collaboration trust model via examples / analyses in the literature with a specific focus on simulated scenarios in military-based operations. This paper is structured as follows: Section 2 explains theoretical constructs of trust in human-robot collaboration factors, while Section 3 describes some examples of practical and potential applications of human-robot collaboration in military domains. Section 4 covers important concepts related to human-robot collaboration and trust interplays. Next, Section 5 provides a formal representation of the obtained concepts. The simulation traces and results of the implemented model in military-related scenarios are explained in Section 6. In Section 7, the model is confirmed using mathematical analysis and automated logical checking properties of simulation traces. Lastly, Section 8 summarises this study.

2. FORMS OF HUMAN-ROBOT INTERACTION (HRI)

Human-robot interaction (HRI) is a multidisciplinary and problem-based field by nature and necessity. It brings together people from various domains, including engineers, psychologists, designers, anthropologists, sociologists and philosophers, into a single research domain (Honig & Oron-Gilad, 2018). There are three categories of HRI based on human-robot factors: human-related factors (i.e., human ability factors, performances and qualities) (Beer *et al*., 2014); robot-related factors (i.e., robot performance issues and attributes) (Chen, 2018); and environmental factors (i.e., teamwork, task assignment and collaboration) (Onyeulo & Gandhi, 2020). Among these issues, robot-related factors have the largest impact on the trust in HRI. For example, robot performance regulates the value or perception of an executed action staged by the robot from the human operator's viewpoint. These viewpoints could be in terms of reliability, faulty behaviour, fault occurrence, transparency and / or level of autonomy. Another spectrum to investigate forms of HRI is through their interaction behaviours. In general, there are three types of HRI, namely, *coexistence*, *cooperation* and *collaboration*, as shown in Figure 1. Based on these types, coexistence is the slightest form of human-machine interaction, whereby it is an episodic encounter of robots and humans, with the contact (e.g., interaction) being very restricted in time and space. In general, this form of interaction does not involve a common goal, and the main objective of the interaction is to impede mutual constraints, accidents and confrontations (Pickering *et al*., 2017).

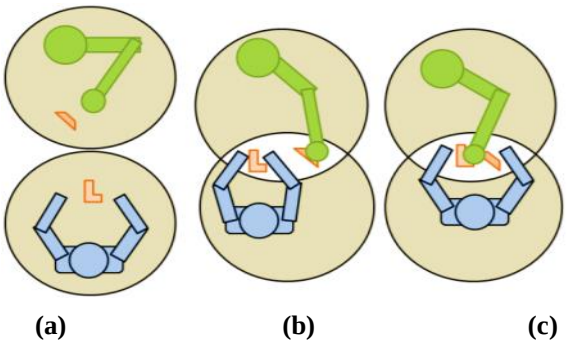


Figure 1: Forms of human-robot interaction: (a) Coexistence (b) Cooperation (c) Collaboration (adopted from Bauer *et al*., 2016).

Next, cooperation is involved when tasks are divided among the group members, and each person is accountable for solving a portion of the problems. In industrial contexts, humans and robots collaborate on a common task, but task execution occurs at distinct times. These operations include putting in place suitable safety precautions because people and robots share the same workstation (e.g., progressive padding, pressure mats and safety curtains). Finally, collaboration entails explicit contact and coordination concerning interaction between different entities (Pearson, 2019). It is also characterised by the establishment of task assignments and the use of synergies. In order to do this, collaboration entails sharing a common intention, preparing an action, and carrying it out as a unit. It is also important to note that cooperation calls for an equitable role distribution, which leads to more effective interactions and a less cognitive load for the human partner (Galin & Meshcheryakov, 2020). This can be particularly prevalent in physical interactions, i.e., scenarios with direct physical contact or physical contact mediated through an object of shared interest. For example, physical interactions often require interdependent actions based on a shared representation of the task and environment (Bütepage & Kragic, 2017).

3. HUMAN-ROBOT COLLABORATION IN MILITARY APPLICATIONS

In recent years, computer scientists and engineers have developed several working prototypes to demonstrate that intelligent machines can work together with humans. These prototypes range from robotic carriers that can assist infantry units in carrying ammunition and equipment to artificial intelligence (AI) enabled autonomous drones that affiliate with human fighter pilots to deliver assistance for aerial bombardment. The pragmatic notion of human-robot collaboration can be observed in manufacturing environment settings where these collaborative robots (cobots) are planned to work hand-in-hand with human operators. Thus, cobots focus on monotonous chores, such as detailing, inspection and picking heavy loads, to help human operators focus more on tasks that require intuitive rule-of-thumb and advanced problem-solving skills. In addition, it can speed up some procedures and adapt to unique requirements (e.g., special requests), leading to an increase in production and curbing the unpleasant, dull, repetitive and tedious work resulting in elevating the burden on humans (Marvel & Norcross, 2017). As a result, the improvement of the working environment can diminish potential occupational injuries such as slipped discs, eye strains or even mental-related issues (Perelman *et al*., 2018). From a military standpoint, the vital perspective to deploy human-robot collaboration-related applications is due to the increasing pace and complexity of decision-making during critical missions. Moreover, in future conflicts, military authorities will operate across numerous domains with several allies, considering multiple predicaments and alternatives. Therefore, human-robot collaboration will provide related decision cycles faster than the capacity of human-only cognition to process complex issues and problems (Au *et al*., 2018).

Currently, it is common to use robots in military operations (e.g., war zones) to investigate and defuse threats (e.g., land mines, booby-traps) or retrieve high-risk objects with the insight that the loss of a robot is a far more appropriate outcome than the death of a soldier. The implementation of human-robot collaboration for military purposes has been in combat tasks that are dull, dirty and dangerous (i.e., 3D jobs). For example, remotely controlled robots (e.g., telepresence robots) play a vital role in current military operations. These robots are used in disarming bombs and performing air / ground-based troop missions (Chen, 2018). It is expected that soon, more autonomous robotic systems with various functions, responsibilities and operating requirements will dominate most conventional military operations, particularly when it falls to the collective strategies between humans and robots. Future applications of this perspective are: i) strengthen information superiority (e.g., provide aerial photos of an adversary's position); ii) reduce operator's cognitive load (e.g., detect new targets in a firefight); iii) take on physical loads (e.g., carry a squad's heaviest gear); and iv) handle dangerous tasks (e.g., deliver supplies or conduct medical evacuation in a danger zone) (Ghute *et al*., 2018; Hidalgo *et al*., 2019; Boos *et al*., 2019).

Another potential example is the execution of automated negotiation skills by autonomous robots. Currently, many research endeavours have been conducted to develop and evaluate artificial intelligence techniques (e.g., machine learning, multi-agent models) to enrich possible negotiation algorithms to be used in robots / machines. As a result, these robots can perform optimal negotiation decisions for some situations, which can be presented as a human-robot collaborative negotiator. In this case, the instructions can be translated into moves by a negotiation counterpart robot to which the soldiers must respond and decide upon the action (Haring *et al*., 2019). For example, this type of technology could coach humans to negotiate or extend negotiation best practices during military missions. This creative standpoint will look at the future of human-robot teams as a critical component of future battlespaces, creating a complex but potentially increasing survivability while helping human counterparts (e.g., soldiers) to accomplish the assigned missions (Ullman & Malle, 2016).

4. TRUST IN HUMAN-ROBOT COLLABORATION

The computational model of human-robot trust was developed based on empirical and theoretical findings obtained in well-established literature and domain (e.g., robotics and automation). In the past, human-robot trust was dedicated primarily to collaboration / teamwork within groups of autonomous agents, machines or robots. On the other hand, robots differ from other forms of automation in several ways, including mobility, unfamiliarity with the general public and physical embodiment. As a result, issues influencing HRI trust should be explored independently. The influential factors found in our human-robot trust model are based on the literature on human-robot collaboration ranging from robot attributes (e.g., automation level, controllability, behaviours and embodiment) to human interaction factors (e.g., feedback, personality and openness towards interaction). Moreover, the latest trend shows an increasing interest in leveraging human involvement effectively by enhancing trust in human-robot collaboration (Chen *et al*., 2018; Langer *et al*., 2019; Aziz & Ghanimi, 2020). Contrary to autonomous systems, which is devised mainly to take humans out of the loop, collaborative robots require people and robots to work together in constant and comparatively long-term interaction. In collaborative tasks or activities, for robots to participate as team members with humans, they must have human-like capabilities that enable fluid and effective teamwork / collaboration with humans (Kwiatkowska & Lahijanian, 2016; Robert *et al*., 2020). Trust is also an essential factor to consider as trust changes may significantly affect teamwork outcomes, consequently affecting intended / assigned tasks (Lewis *et al*., 2018).

The attention on trust in human-robot working alliance emerges because several previous works show that humans tend to trust robots similarly to other humans (Bodala *et al*., 2020). Therefore, it is a major concern that people may misunderstand the possible risks of handing over decisions to a robot. Depending on the assigned tasks, the definition of trust is particularly appealing as different tasks deal with various levels of risk. An example of this risk is that a robot in a life-threatening condition is not attempting to reward humans for their conformity but rather to mitigate risk to save human life (Onyeulo & Ghandi, 2020). When it comes to collaborative tasks, a robot needs real-time feedback to maintain the trust that leads to positive outcomes in collaboration, otherwise the main risk will be potential rejections from team members. Several components are imperative to regulate trust in human-robot collaboration for specific assignments. First, human factors such as personality (based on personality traits), openness for interaction, perceived controllability and appraisal about progress (e.g., achievement in solving the assigned task) are vital components to trust the robot. A considerable amount of literature has been published on these components (Ososky *et al*., 2014; Pickering *et al*., 2017; Satterfield *et al*., 2017; Pearson, 2019; Ajenaghughrure *et al*., 2021). In general, openness and agreeableness are often associated with success in HRI. Perceived controllability shows conditions where the robot is under the team member's control. It means the robots' actions are predictable when performed by humans (Xu *et al*., 2019). Thus, appraisal of progress enables team members to determine their expressions and take more active involvement in the collaborative activity. It has been found that positive feedback decreases dropout rates and enhances team members' experiences (Perelman *et al*., 2018).

Other important robot-based concepts, such as automation level, physical embodiment and social behaviours, provide a realistic idea of how robot-based collaboration will work. Automation level refers to the degree of an automated task from Levels 1 to 10. These levels are a set of ranges from complete human control to complete computer / robot control. For instance, Level 1 indicates that the user does the task and turns it over to the robot / computer to execute the rest. In contrast, Level 10 demonstrates that the robot / computer does the action if it determines it should be done. In addition, the automation level of a system changes an agent's capacity to make some critical decisions based on evidence/knowledge on its own without any other external control requests (Beer *et al*., 2014; Lewis *et al*., 2018; Razin & Feigh, 2020). Thus, the robot's automation level modifies the robot's ecosystem, whether humans can intervene in the robot's control loop and the tasks that can be given. The robot / computer informs the human operator only if it decides that the operator should be informed. When it comes to HRI, social behaviour capabilities allow humans to use sophisticated social cues and interaction. This concept includes human-centred multi-modal communication and teamwork. This factor is a unifying feature because social robots communicate and synchronise their behaviour through verbal, non-verbal or affective-based sensory systems. Other components such as transparency, reliable behaviours, perceived competency and perception about robots drive positive perceptions from users' perspectives about trust-related perceptions with human-robot collaborative alliances (Xu *et al*., 2019; Robert *et al*., 2020; Javaid *et al*., 2020).

First, transparency is operationalised as the user's understanding of why a robot behaves the way it does. It allows some technical understanding of how the technology's operating rules and logic are apparent to users. Furthermore, researchers have demonstrated that for robots intended to serve as autonomous squad members in military contexts, transparent communication from the robot (e.g., updates on the robot's status, surroundings, projected course and uncertainty) improve self-reported trust and assist operators in calibrating trust in the capabilities of the robot (Baker *et al*., 2018). This suggests that transparency will increase in importance as a system's autonomous capabilities increase as it allows users to calibrate their trust in the systems during information uncertainty (Bodenhagen *et al*., 2017; Phillips *et al*., 2018). Reliable behaviour relates to whether the technology exhibits the same expected behaviour over time (through the understanding of technological constructs and perceived level of automation). It indicates how human (in terms of human-like subjects) judges the behaviours or possible outcomes of the robot regarding its perceived competency. Through this idea, the robot should be able to convey social behaviours (e.g., prosody, tone, turn-taking, and emotional expressions) to be considered trustworthy (Lewis *et al*., 2018; Khavas *et al*., 2020).

Perception about robots deals with how humans behave towards robots as collaborators / teams and hence correspondingly decide on robots as "equal" co-workers. This notion is related to the perceived intelligence that asserts how intelligent and human-like subjects pass judgment on the robot's behaviours or performances. Nonetheless, humans' pre-conceived perceptions about robots can be manipulated by science fiction concepts that may lead to highly overstated beliefs. Therefore, humans should be aware of the robot's abilities and what signifies suitable interaction. Other concepts, such as perceived risk and distrust, deal with the destruction of trust (Salem *et al*., 2015; Pearson, 2019). Perceived risk helps to explain why users frequently do not progress from the desired stage to the action stage, relying on the robot's decision. This could be viewed as either the robot is highly adaptive but tends to be risky or not very adaptive and conservative (Schaefer et al., 2016). Therefore, to avoid this, the robot should be attentive to every detail the user wants and reassure them by answering all their questions. Another concept that eliminates trust is distrust. Distrust in robots could be the most prominent dividing force in implementing robot-based collaboration. In general, several findings suggest that users might trust a robot more if they had more experience with it and have control over how it is used rather than being told to follow orders from a mysterious robotic system (Hoff *et al*., 2015; Ullman & Malle, 2016; Lazanyi & Hajdu, 2017; Guo *et al*., 2019).

5. FORMAL MODEL

Several traditional AI-based modelling approaches, such as formal logic (e.g., temporal / modal logic), predicate calculus and dynamic systems, have been used in cognitive modelling to address the complexity of processes through some substantial form of assumption in decomposition. Nevertheless, within these human-directed sciences (e.g., cognitive sciences, computational psychology and intelligent ambient systems), serious debates or disputes have occurred repeatedly on such kinds of assumptions. These separation assumptions are difficult to address in conventional AI-based modelling due to limited knowledge representation abilities. Separation assumptions used to address human complexity concerns include: 1) mind versus body; 2) cognition versus emotion; 3) individual processes versus collective processes; 4) non-adaptive processes versus adaptive processes; and 5) earlier versus later (temporal separation). In addition, it is debatable whether consciousness can be studied while ignoring the body, cognition while neglecting emotion, sensory processing besides action preparation, or how it contributes to social processes. In order to overcome these issues, the network-oriented modelling approach was introduced. This approach incorporates a temporal dimension to overcome these challenges, allowing connections to be interpreted as temporal-causal connections (Treur, 2020). The mental concept of a network also goes some way to explaining the dynamics of the modelled processes. This section defines how the network-oriented modelling method (based on temporal causal network) is used to formulate and construct a computational model of human-robot collaboration trust. Consequently, all the identified factors from the related literature (as discussed in Section 4) were utilised to conceptualise the model, as described in Figure 2.

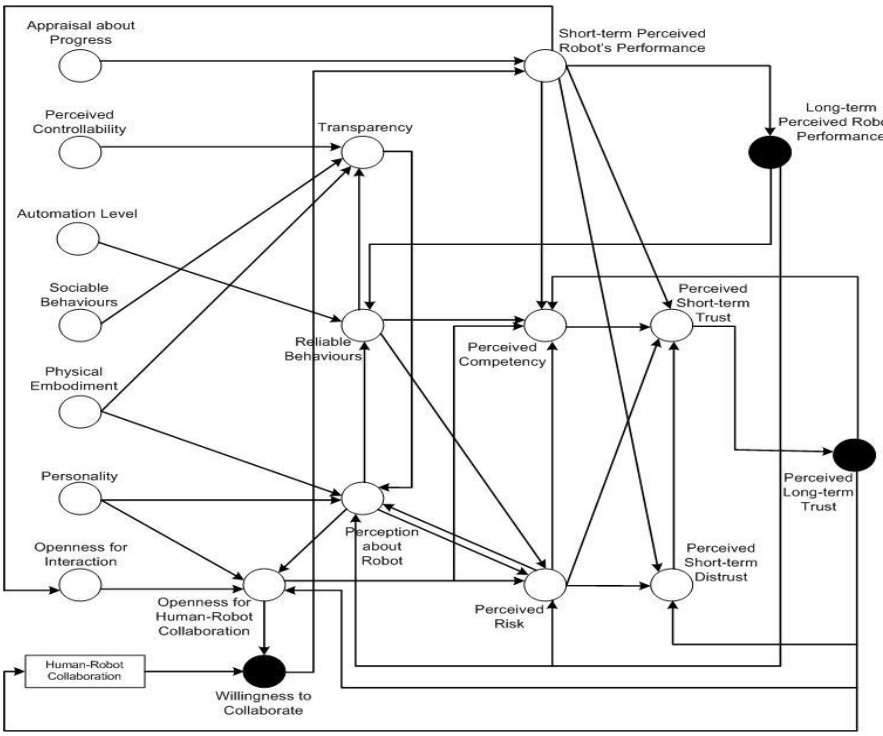


Figure 2: Conceptual network-oriented modelling of trust in human-robot collaboration.

Furthermore, by integrating a temporal dimension, it is explicitly modelled to have intense and circular causal interaction, including how the processes' timings are. This temporal dimension allows causal reasoning and simulation for cyclic causal graphs or networks, such as networks modelling mental or brain operations, social interaction mechanisms, and the duration of such observed processes (Treur, 2020). With this ability, the network-oriented modelling approach can model a complex cognitive process that requires the strength of a connection, multiple impacts on a state and speed of change of a temporal effect.

5.1 Modelling Approach

In this paper, a set of differential equations (function with its derivatives) to represent interplays between related concepts is specified. In applications, these differential equation functions usually cover physical quantities, with the derivatives indicating their rates of change. These equations determine an interplay or connection between two or more states. There are two central relationships to represent the observed phenomena, namely instantaneous and temporal relationships (Aziz & Ghanimi, 2020). An instantaneous relationship explains the direct impact on states and their connections. For example, given *f*:

$$f(t) = \alpha.z(t) + (1-\alpha).h(t) \quad (1)$$

From this equation (Equation 1), parameter α is used to regulate the possible contribution rate between any *z* and *h* functions or variables. In this case, if $\alpha = 0.7$, it means that the function *z* will contribute up to 70% towards the overall value (with 30% contribution from function *h*). The temporal relationship is often related to the accumulated effect from the previous contribution of the same function. This form of contribution can be considered as a delay condition (regardless of accumulating or decaying contributions). A description of the temporal representation of *y* function can be presented as follows:

$$y(t+\Delta t) = y(t)+\tau\ .<change\_\ expression>.\Delta t \quad (2)$$

and assuming $\tau \neq 0$, this is equivalent to $<change\_expression> = 0$ for all variables *y*. Moreover, as;

$$<change\_expression> = (1-y(t)).(<basic\_change>).y(t) \quad (3)$$

the criterion for equilibrium is:

$$(1-y(t)).(<basic\_change>).y(t) = 0 \quad (4)$$

It is also worth noting that the change process (Equation 2) is measured in time intervals between *t* and $t+\Delta t$. A flexibility rate (τ), in addition to all of this, determines the rate of change for all the temporal specifications. In order to construct a quantifiable model of trust, a formal specifications trust measurement within a human-to-robot perspective must be designed. Both relationships (instantaneous and temporal) are related to these three main causal relations, namely: 1) strength of a causal relation; 2) combining causal impacts on a state; and 3) speed of change of a state. These three are essential to represent the conceptual constructs in the real world as not all relationships are equally important (e.g., equal weightage) in representing the notion of states and connections among them (Treur, 2020). Next, the formal specifications will be discussed for simulation purposes.

5.2 Formal Specifications Dynamics

In general, trust can be considered as a dynamic process. As a result, in order to investigate these dynamics, it is necessary to formalise and study the dynamic specifications of such processes. For example: How does a trust level at one point in time compare to trust levels at other points in time? (Treur, 2020) To study different trust levels for various scenarios, we develop a set of formal specifications to perform input and parameter manipulation actions to simulate possible outcomes. In this section, we introduce the description of our formal specifications of the human-robot collaboration trust model. The formalisation introduced in this part has been based on the differential equations technique for both instantaneous and temporal relations.

5.2.1 Transparency, Reliable Behaviours, Perception of Robot

Transparency (*Tr*) is related to the weighted contributions of physical embodiments (*Pe*), social behaviours (*Sb*), perceived controllability (*Cl*) and reliable behaviours (*Rb*) (Equation 1). This concept illustrates how higher transparency (via the perception of embodiment, visible automation, and behaviours) may mitigate the cry wolf effect, a phenomenon commonly observed in high-risk decision making where the threshold to invoke an alarm is regularly set very low in order to accept all conditions as critical (Ajenaghughrure *et al*., 2020). Next, reliability (*Rb*) is shown to affect trust and participants' self-assessment of performance. It relates to the combination of a robot's performances (*Lp*), perceived controllability and perceived automation level (*Al*) (Equation 6). Perception about a robot (*Pc*) is calculated using the positive contribution of current long-term performance, personality (*Ps*) (openness and agreeableness), physical embodiment and transparency. However, perceived risk (*Pr*) reduces positive perceptions about the robot (Equation 7).

$$Tr(t)= \alpha_{Tr}(w_{r1}.Pe(t)+w_{r2}.Sb(t)) + (1- \alpha_{Tr}).(w_{r3}.Cl(t) + w_{r4}.Rb(t)) \quad (5)$$

$$Rb(t)= \beta_{Rb}.Lp(t)+(1- \beta_{Rb}).(w_{b1}.Pc(t) + w_{b2}.Al(t)) \quad (6)$$

$$Pc(t)=(1-Pr(t)).(\lambda_{Pc}.(w_{p1}.Lp(t)+ w_{p2}.Ps(t))+(1-\lambda_{Pc}).(w_{p3}.Pe(t) + w_{p4}.Tr(t))) \quad (7)$$

5.2.2 Openness for Human-Robot Collaboration, Openness for Interaction, Perceived Competency and Risk

Openness to collaborate is one of the vital core components of the success of any teamwork. It denotes receptivity to new ideas and experiences. For the openness to human-robot collaboration (*Oc*), this component can be measured by assessing the openness in the intervention (*Oi*), perception about a robot, individual's personality and accumulated (long-term) trust (*Ls*) (Equation 8) (Baker *et al.*, 2018). *Oi* is influenced by the perception of the short-term robot's performance (*Sp*) and the individual's openness norm (Oi_{norm}) (Equation 9). When the robot has particular skills to support solving assigned tasks or processes, the competency (*Cy*) levels are crucial as it will improve trust towards a human-robot collaboration. Components such as long-term trust, reliable behaviours, openness for human-robot collaboration and short-term performance increase perceived competency while perceived risk negates the level (Equation 10). The effect of perceived risk (*Pr*) is determined through the contrast sum contribution of performance, reliable behaviour, perception of robots and openness to human-robot collaboration (Equation 11).

$$Oc(t)= \gamma_{Oc}.(w_{o1}.Oi(t)+ w_{o2}.Pc(t) + w_{o3}.Ps(t))+(1-\gamma_{Oc}).Ls(t) \quad (8)$$

$$Oi(t)= \beta_{Oi}.Oi_{norm}(t) + (1-\beta_{Oi}).Sp(t) \quad (9)$$

$$Cy(t)=(\lambda_{Cy}.Ls(t)+(1-\lambda_{Cy}).(w_{y1}.Rb(t)+ w_{y2}.Oc(t)+ w_{y3}.Sp(t)).(1-Pr(t))) \quad (10)$$

$$Pr(t)= (1-(\alpha_{Pr}.Lp(t)+(1-\alpha_{Pr}).(w_{d1}.Rb(t)+ w_{d2}.Pc(t)+ w_{d3}.Oc(t)))) \quad (11)$$

5.2.3 Perceived Short-term Trust, Distrust and Performance

Perceived short-term trust (*Ss*) (Equation 12) is essential in shaping human interactions with one another and with robots. Within human-robot collaboration scenarios, the combination of perceived risk and distrust (*Sd*) reduces short-term trust, while perceived performance and competency always provide positive feedback towards trust. *Sd* (Equation 13) creates negative feedback towards many aspects of human-robot collaborative efforts. The sum contribution of *Sp*, *Ls* and competency mitigate the formation of human-robot distrust. *Sp* (Equation 10) is generated when an individual has appraised the robot's performance (*Ap*) and the accumulated effects of willingness to collaborate (*Wc*) are positive.

$$Ss(t)=(1-(Pr(t).Sd(t))).(\gamma_{Ss}.Sp(t)+(1-\gamma_{Ss}).Cy(t)) \quad (12)$$

$$Sd(t)=(1-(\beta_{Sd}.Sp(t)+(1-\beta_{Sd}).Ls(t))).Pr(t) \quad (13)$$

$$Sp(t) = \beta_{Sp}.Ap(t) +(1-\beta_{Sp}).Wc(t) \quad (14)$$

5.2.4 Willingness to Collaborate, Long-term Perceived Robot's Performance and Perceived Long-Term Trust

Here, willingness to collaborate (*Wc*) (Equation 15) increases or reduces over time. When the weightage combination (w_{cb}) between collaboration task component (*Ck*) and openness on human-robot collaboration (*Oc*) is higher than the previous willingness to collaborate and decay (λ_{decay}) in belief in human-robot collaboration multiplied with the contribution factor (λ_{Cb}), then the belief in collaboration increases. Otherwise, it declines based on its preceding level and influencing component. This circumstance also can be applied to explain the related phenomenon for all temporal relations (e.g., long-term perceived robot's performance (*Lp*) and perceived long-term trust (*Ls*)) based on their corresponding parameters and attributes. It is also essential to address that the change process is assessed in a time interval between *t* and $t+\Delta t$.

$$Wc(t+\Delta t) = Wc(t)+ \lambda_{Cb}.(((w_{cb1}.Ck(t)+ w_{cb2} .Oc(t))-Wc(t))-\lambda_{decay}).Wc(t).(1-Wc(t)).\Delta t \quad (15)$$

$$Lp(t+\Delta t)=Lp(t)+ \lambda_{Lp}.(Sp(t)-Lp(t)).(1-Lp(t)).\Delta t \quad (16)$$

$$Ls(t+\Delta t)=Ls(t)+ \lambda_{Ls}.(Ss(t)-Ls(t)).(1-Ls(t)).\Delta t \quad (17)$$

6. SIMULATION RESULTS

In this study, the proposed model was executed using a numerical programming platform. We conducted several simulations based on selected cases using the developed formal specifications to determine the human-robot collaboration trust level to demonstrate how the model can be used. Several unique patterns of trust have been discovered. These different types are accomplished by setting the levels to range from 0 to 1 (as shown in Table 1). These weights, contributions and regulation rates can also be adapted to simulate unique individual traits. In order to simulate the formation of trust between humans and robots (as depicted in Figure 3), consider this scenario:

> *"A team of military negotiators that consists of a collaborative robot and soldiers has been deployed to an area of high risk for terrorist attacks. The collaborative robot is programmed with human-like negotiation skills, with some capabilities to understand human emotions and gestures. The robot's negotiation skills were programmed based on past experiences of human negotiator experts using deep learning models. For this mission, both soldiers and robot need to collaborate to convince the terrorist to disarm his / her weapon and surrender. During the negotiation process, the robot will analyse the conditions and prepare for any possible encounter to assist the soldiers on what are the best actions to be executed."*

In general, negotiation is a process by which two or more parties make a joint decision. Generally, each party begins the negotiation by proposing the most feasible choice from the particular area of interest. If the other parties are not completely comfortable with an offer, they will make counter-offers to make them relevant to an agreement. As machines can effectively deal with computational complexity; the negotiation scenario was chosen because existing automated negotiating agents / robots could significantly improve if the negotiation space is well-understood. On the other hand, the negotiation space can only be properly developed if the human parties jointly explore their interests. In some complex scenarios, negotiation cannot be managed by AI alone, owing to the inherent semantic problem and emotional issues involved, implying the use of a human-robot collaborative system. For the simulation purpose, the successfulness of the negotiation advice and tasks can be perceived as an appraisal of negotiation progress. The discussion of negotiation algorithms and models is beyond the scope of this paper.

6.1 Baseline Setting Results

Based on this scenario, three baseline fictional soldiers have been specified in Table 1. These fictional characters are labelled as #1 (individual with a positive experience with collaborative robot), #2 (moderate experience) and #3 (no experience). In addition, both soldiers #2 and #3 have sceptical perceptions towards the implementation of collaborative robots in military-related operations.

This paper only shows the simulation runs for three fictional soldiers with different personality profiles due to the extreme possible sequences. For this section, the simulations demonstrate some key human-robot collaboration trust behaviours, including: (i) risk; (ii) distrust; (iii) individual openness; and (iv) perceived robot performance. The simulation results (Figure 4) show that the soldier in Case #1 has a lower perceived risk than the other soldiers in Cases #2 and #3. Thus, it mitigates possible negative impacts in perceived distrust, and hence improves both perceived competency and reliable behaviours. From these graphs (Figures 4(a) and (b)), the initial perceived risk and distrust levels are slightly high and later decrease as the trust and openness towards human-robot collaboration improve. These findings are consistent with Salem *et al.* (2015), Satterfield *et al.* (2017) and Ajenaghughrure *et al.* (2020). It is also important to support the success of following the negotiation advice and tasks assigned by the collaborative robot. The overall human-robot collaboration trust remains low for Case #3 compared to Case #2.

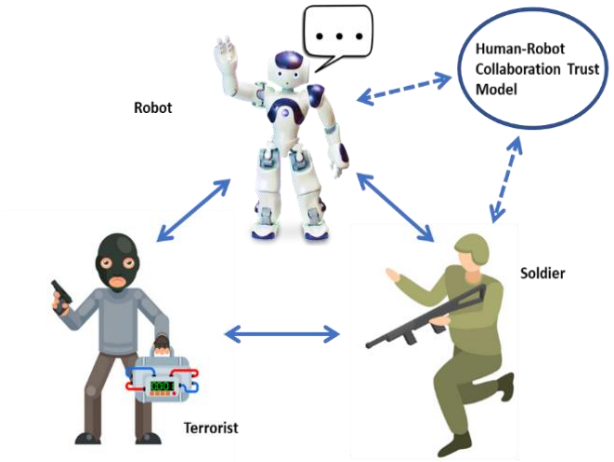


Figure 3: Conceptualisation of the scenario for simulated tasks.

Table 1: Settings of the simulations.

Factors / Cases	**#1**	**#2**	**#3**
Appraisal about Progress (Ap)	0.8	0.5	0.1
Perceived Controllability (Cl)	0.7	0.6	0.2
Perceived Automation Level (Al)	0.8	0.5	0.2
Social Behaviours (Sb)	0.9	0.4	0.1
Physical Embodiment (Pe)	0.7	0.5	0.2
Personlity (Ps)	0.7	0.4	0.1
Openness Towards Interaction (Oi)	0.8	0.5	0.3
Individual Openness Norm (Oi_{norm})	0.9	0.5	0.1
Collaborative Task Components Norm (Ck_{norm})	0.9	0.5	0.1

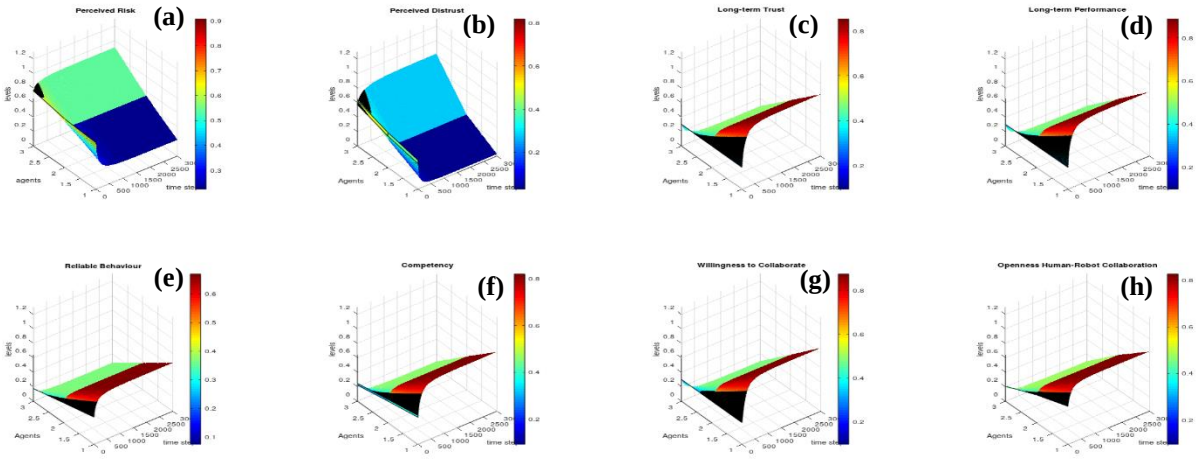


Figure 4: Simulation results for baseline condition: (a) Perceived risk (b) Perceived distrust (c) Long-term trust (d) Long-term performance (e) Reliable behaviours (f) Competency (g) Willingness to collaborate (h) Openness towards human-robot collaboration.

In this experiment, an improvement in the competency and reliability levels (and other concepts) will influence the development of long-term trust, performance, and perception in human-robot collaboration (as in Figures 4 (c), (d) and (g)). These simulation results reflect those of Beer *et al.* (2014), Salem *et al.* (2015), Pearson (2019) and Khavas *et al.* (2020), who also found similar behaviours in their empirical studies. In addition to the initialised experimental settings, as depicted in Table 1, we have explored the impact of the norms for both openness to interaction and collaboration task components. Regardless of different variations of the initialised values, the overall baseline results for all temporal conditions show almost similar patterns except at the initial convergence points.. Therefore, the following section moves on to discuss the different variations of initial conditions to show the behaviour of human-robot collaboration behaviours during unpredictable conditions (e.g. cyclic patterns). These behaviours provide us with some insights into how erratic recommended advice will influence trust formation throughout time.

6.2 Various Performance Appraisal Setting Results

We implemented the appraisal uncertainty related to the robot's advice. This is important to investigate potential impacts on overall conditions within the model (e.g., progression on willingness to collaborate or changes in openness for human-robot collaboration). For this condition, a repeated sinusoidal-like (oscillation) behaviour (based on some specific intervals) was used to simulate changes in beliefs on appraisal about the progress (performance) of the robot. For example, if the robot recommends specific actions and proven to expose some risk to the mission, the human's appraisal of progress will decline. In contrast, a good recommendation will improve the appraisal of the performance of the robot. In order to do this, we change the dynamics of appraisal (*Ap*) based on selected variations, with all the other variables in Table 1 being constant, with the simulation results shown in Figure 5.

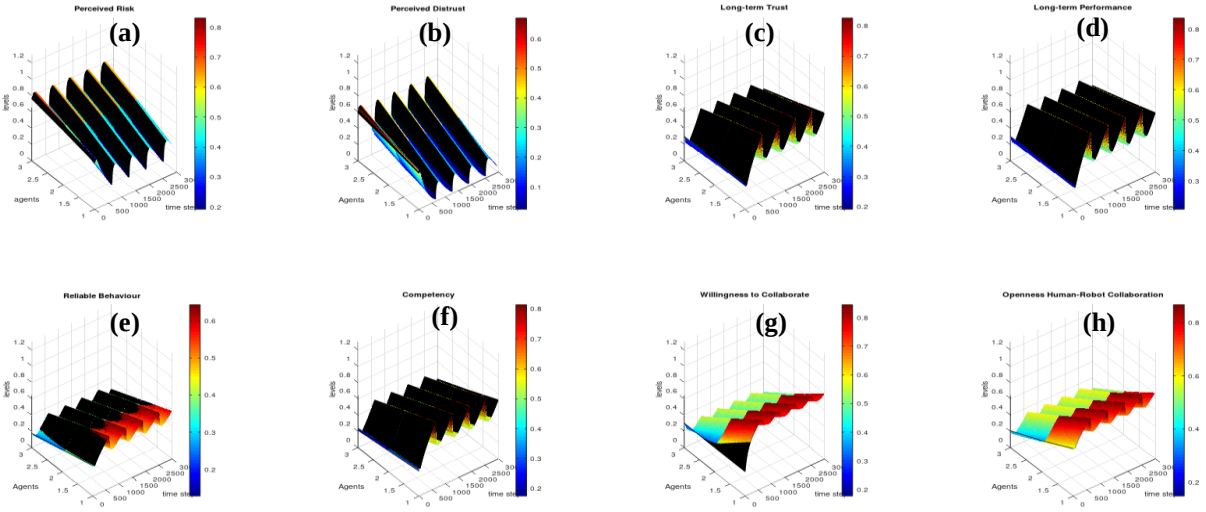


Figure 5: Simulation results for the oscillated condition: (a) Perceived risk (b) Perceived distrust (c) Long-term trust (d) Long-term performance (e) Reliable behaviours (f) Competency (g) Willingness to collaborate (h) Openness towards human-robot collaboration.

From Figures 5(d) and 4(g), the dynamic nature of willingness to collaborate can be linked to trends in appraisal of the robot's performance (e.g., capable of providing the right advice for optimal decision-making process throughout negotiation processes). It is noteworthy that the model also shows that the trust level is changing according to the robot's performance. Even though the robot sometimes has some low trust levels, the trust recovery will occur as the results are still within the acceptable range. This condition is also seen in many automation systems, where if automation reliability is significantly higher, intermittent failures within the appropriate limits do not substantially reduce trust in the automation unless failures are sustained far beyond the human acceptance threshold (Baker *et al*., 2018). In addition, it is believed that human intervention with a robot-suggested decision is crucial since robots in the future will not be completely compatible. However, humans will still use them to perform various high-value tasks. In this case, human interference in robot tasks may be a solution to lower out-of-the-loop negative impacts. Moreover, the simulation results of this model reflect the empirical findings described by Lazanyi & Hajdu (2017) and Bodala *et al*. (2020).

The second experiment was conducted to simulate the impact of long-term declining appraisal of progress in human-robot collaboration trust, with the results shown in Figure 6. Herein, we describe this condition as "the robot begins to fail dramatically in providing good advice or recommendation to reduce the risk of the terrorist attack." In this condition, all human-robot trust levels are decreased drastically (as in Figure 6(c)), resulting from higher perceived risk and distrust. One plausible explanation is that trust appears to be subject to inertia, with trust after loss being especially difficult and beyond repair.

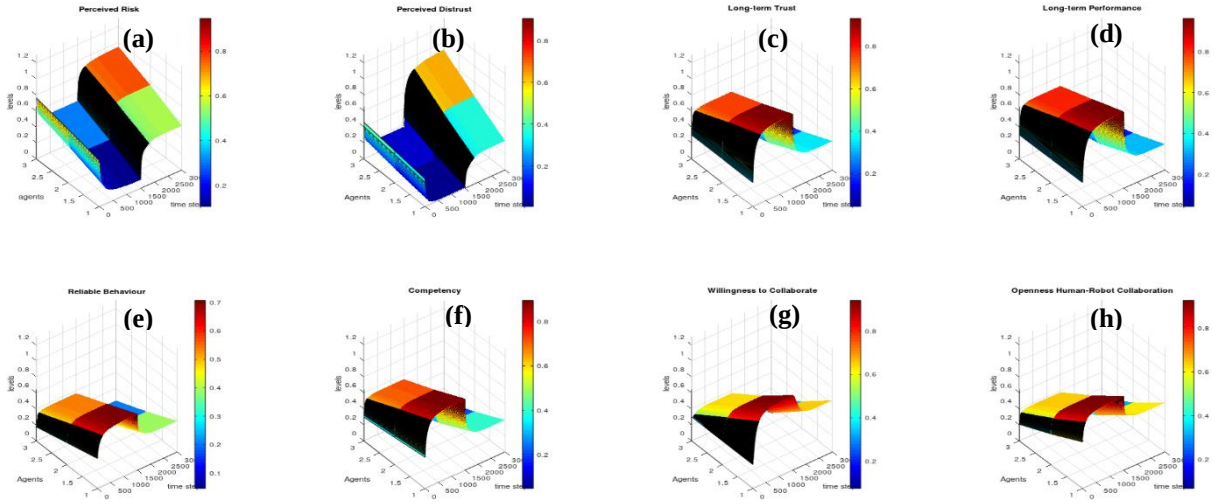


Figure 6: Simulation results for the fluctuated condition: (a) Perceived risk (b) Perceived distrust (c) Long-term trust (d) Long-term performance (e) Reliable behaviours (f) Competency (g) Willingness to collaborate (h) Openness towards human-robot collaboration.

In general, a robot's wrong judgement can cause significant performance risks as decision-making unreliability will intensify the out-of-the-loop problem since the robot task must abruptly transfer total decision-making to the human team (Baker *et al.,* 2018). Without any trust recovery mechanism, it will inhibit the openness in human-robot collaboration and diminish the overall willingness to collaborate. The simulation traces results obtained to confirm this condition are presented in Figures 6(g) and 6(h). These results are in parallel with the empirical findings in Beer *et al*. (2014), Lazanyi & Hadju (2017), Baker *et al.* (2018), Lewis *et al*. (2018) and Bodala *et al.* (2020).

7. EVALUATION

In our study, we employed two evaluation methods (mathematical analysis and temporal trace analysis) to determine that the model implementation and its associated results accurately represent the conceptual description and specifications as in the literature.

7.1 Mathematical Analysis

For mathematical verification, equilibria analysis describes situations in models where the values (continuous) approach and stabilise at a boundary within certain conditions. It implies that if a differential equation defines a system's dynamics, then equilibria can be approximated by setting a derivative (or all derivatives) to zero. One thing to consider is that an equilibria response is considered stable if the system always returns to it after minimal fluctuations. For example, using this autonomous equation:

$$f(t) = y(t+\Delta t) = y(t) + \beta[q(t)-y(t)].\Delta t \quad (18)$$

Therefore, assuming β is nonzero:

$$df(t)/dt = q-y \quad (19)$$

The equilibria or constant solutions of this differential equation are the roots of the equation:

$$df(t)/dt = 0 \quad (20)$$

Hence the equilibria point can be found when $q = y$.

As a result, the presence of reasonable equilibria indicates the model's correctness. It can also be observed that when a specific state is increasing or decreasing when the state is not among the equilibria points. For example:

- f has an equilibria point at t if $df(t)/dt = 0$
- f is increasing if at t if $df(t)/dt > 0$
- f is decreasing if at t if $df(t)/dt < 0$

These equilibria conditions are interesting to be explored, as it is possible to explain them using knowledge from the theory or problem that is modelled. In this section, the stationary points and equilibria occurring in the model have been analysed. Based on this assumption, we will have an equilibrium stage when:

$$dWc/dt = 0,\ dLp/dt = 0,\ dLs/dt = 0 \quad (21)$$

Thus the following equations are found:

$$\lambda_{Cb}.(((w_{cb1}.Ck + w_{cb2}.Oc) - Wc) - \lambda_{decay}).Wc.(1-Wc) \quad (22)$$
$$\lambda_{Lp}.(Sp-Lp).(1-Lp) = 0 \quad (23)$$
$$\lambda_{Ls}.(Ss-Ls).(1-Ls) = 0 \quad (24)$$

Theoretically, it can be presumed that the equilibrium phase exists when the difference between the current accumulated impact for all temporal specifications and short-term instantaneous specifications is equal to zero. Thus, the equilibria points of these specifications can be made by distinguishing cases from all instantaneous equations (from Equation 1 to Equation 14). In this paper, only three equilibria points are considered. These equilibria points are Wc=0, Lp=1, and Ls=0. For these cases, it can be obtained that the subsequent cases can determine the values for the equilibria:

Case #1: *Wc=0*
For this case, from Equation 14, it follows that:

$$Sp = \beta_{Sp}.Ap, \text{ assuming } \beta_{Sp} \neq 0 \quad (25)$$

Case #2: *Lp=1*
From Equations 6, 7 and 11, it follows that:

$$Rb = \beta_{Rb} + (1 - \beta_{Rb}).(w_{b1}.Pc + w_{b2}.Al) \quad (26)$$
$$Pc = (1-Pr).(\lambda_{Pc}.(w_{p1} + w_{p2}Ps) + (1-\lambda_{Pc}).(w_{p3}.Pe + w_{p4}.Tr)) \quad (27)$$

From Equation 11, this is equivalent to:

$$Pr = (1-(\alpha_{Pr} + (1-\alpha_{Pr}).(w_{d1}.Rb + w_{d2}.Pc + w_{d3}.Oc))) \quad (28)$$

Case #3: *Ls=0*
For this case, from Equation 9, it follows that;

$$Oc = \gamma_{Oc}.(w_{o1}.Oi + w_{o2}.Pc + w_{o3}.Ps) \quad (29)$$

Moreover, from Equation 13, it follows that,

$$Sd=1-\beta_{Sd}.Sp. \quad (30)$$

In addition to the equilibria analysis, we performed another analysis to confirm the stability point that exists in our proposed model. Figure 7 shows the differences in equilibria results over time under three scenarios: baseline, fluctuated and oscillated for all three concepts; willingness to collaborate, long-term performance and long-term trust. Except for the baseline condition, the values of the model's states may not always end up in a simple equilibrium value but instead fluctuate according to a predefined sequence, widely recognised as a limit cycle. The figure depicts the experimental results for all the states (willingness to collaborate, long-term performance and long-term trust) that exhibit such behaviours. According to the analysis of the results, stationary point conditions are met when the average absolute deviation overall temporal states is 0.00014195 (which is $< 10^{-2}$), with the absolute maximal and minimal deviation values being 0.001297 and 0.0185, respectively. This provides computational indications that the implemented model corresponds to the model description (Treur, 2020).

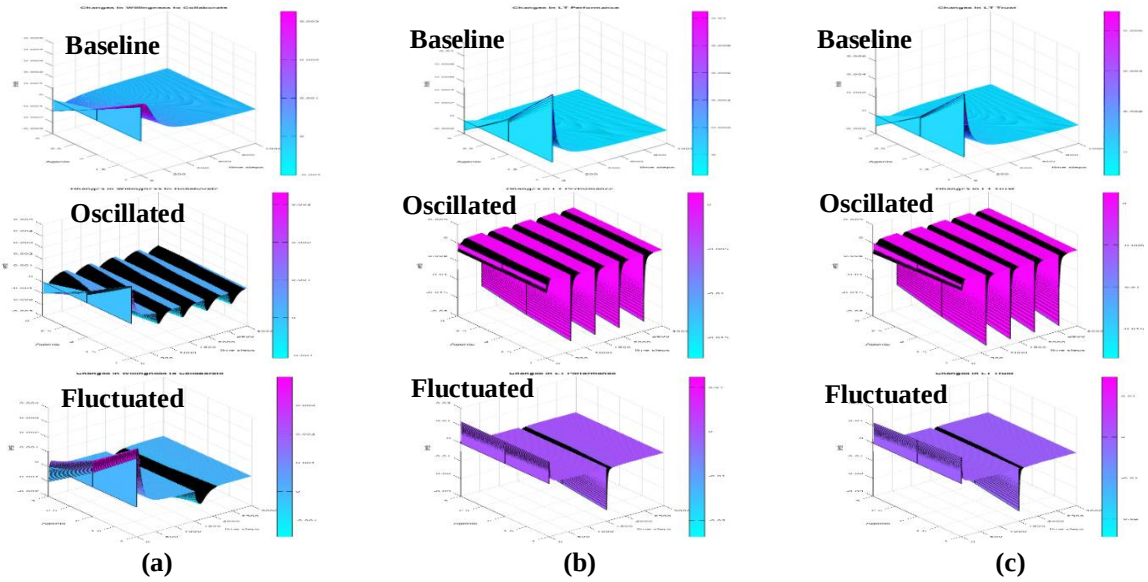


Figure 7: Equilibria results for conditions (baseline / oscillated / fluctuated): (a) Willingness to collaborate (b) Long-term performance (c) Long-term trust.

7.2 Temporal Trace Analysis

In this section, the temporal trace language (TTL) representations are used by specifying a set of dynamic statements that should (or should not) hold. To express dynamic properties precisely, explicit references to time points and traces are required. TTL, like the approach in situation calculus, is built on atoms that refer to, for example, traces, time and state properties. For example, the output state of X in trace γ at time t property p is formalised by:

$$\text{state}(\gamma, t, \text{output}(X)) \models p \quad (31)$$

Throughout the remainder of this paper, these kinds of atoms will be referred to as *Holds* atoms. Based on such *Holds* atoms, dynamic properties can be built using the usual logical connectives and quantification (for example, over traces, time and state properties). These expressions are used to verify these statements based on generated temporal traces. The TTL implementation allows for the formal specification and assessment of dynamic properties, including qualitative and quantitative properties.

This type of verification aims to determine whether or not the simulation model performs as anticipated. A common application of a property that can be verified is whether no unforeseen circumstances occur, such as a variable exceeding its limits. In this paper, several dynamic properties have been formalised in TTL for logical verification purposes. Below, a number of them are shown, both in semi-formal and formal notations.

VP1 ≡ Physical embodiment and social behaviours of a robot improve trust in human-robot collaboration.

If a physical robot is located within the visible range of a human and shows some social behaviours understood by a human, then it will increase the level of trust in human-robot collaboration:

VP1 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀R1,R2, D1,D2:REAL
[state(γ,t1)|= has_value(physical_embodiment, R1) &
state(γ,t1)|= has_value(sociable_behaviours, R2) &
state(γ,t1)|= has_value(LT_trust, D1) &
state(γ,t2)|= has_value(LT_trust, D2) &
t2 > t1 +d & R1 > 0.8 & R2 > 0.7 & D1 > 0.6] ⇒ D2 ≥ D1

VP2 ≡ Robot's performance increases humans' perception of the robot's ability to collaborate with them.

Once a robot shows the interaction behaviours and results as expected by humans at any time point t1, it will increase the perception about humans' perception of the ability of the robot to collaborate with them in the future:

VP2≡ ∀γ:TRACE, ∀t1, t2:TIME, ∀M1, M2, H1, H2:REAL
[state(γ, t1)|= has_value(performance, M1) &
state(γ, t1)|= has_value(perception_robot, H1) &
state(γ, t2)|= has_value(performance, M2) &
state(γ, t2)|= has_value(perception_robot, H2) &
M1 ≥ 0.5 & t2 > t1 & H1 > 0.5] ⇒ H2 > H1

VP3≡Stability of Variable *x*.

For all time points t1 and t2 between tb and ta in trace γ, if at t1 the value of x is J1, then at t2, the value of x is between x-α and x+α (where α is constant):

VP3≡ ∀γ: TRACE, ∀t1, t2: TIME, tb, te :TIME, ∀J1,J2:REAL
[state(γ,t1)|= has_value(x, J1) &
state(γ,t2) |= has_value(x, J2) &
tb < t1 < te & tb < t2 < te] ⇒ J1 – α ≤ J2 ≤ J1 + α

This property can be used to evaluate which factors do not vary significantly or change after a series of time steps (stable point).

VP4 ≡ Perceived risk results in the low perceived performance.

If humans experience negative results based on the previous encounter with human-robot collaboration at a later time point, it reduces the perceived performance of the robot:

VP4 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀J1,J2, L1,L2:REAL
[state(γ,t1)|= has_value(perceived_risk, J1) &
state(γ,t2)|= has_value(perceived_risk, J2) &
state(γ,t1)|= has_value(performance, L1) &
state(γ,t2)|= has_value(performance, L2) &
t2 > t1 +d & J1 > 0.5] ⇒ L1 ≥ L2

VP5 ≡ Variable *v* between boundaries.
For all time points t between tb and ta in trace γ, if at t, the variable v value is x, then the overall result will be within this boundary min ≤ x ≤ max:
VP5 ≡ ∀γ: TRACE, ∀t, tb, te:TIME, ∀v:VAR, ∀x, max, min:REAL
state(γ,t)|= has_value(v,x) &
tb ≤ t ≤ te ⇒ min ≤ x ≤ max

This property can be used to decide whether a variable continues to remain within the boundaries. For example, the perceived risk should never fall below 0 or continue to climb above 1.

VP6 ≡ Robot behaviours transparency improves human trust.
When robot behaviours are understood by humans (in terms of mechanism, algorithms and operation), humans will trust a robot more than a robot without this attribute.:
VP6 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀H1,H2, K1,K2:REAL
[state(γ,t1)|= has_value(transparency, H1) &
state(γ,t2)|= has_value(transparency, H2) &
state(γ,t1)|= has_value(LT_trust, K1) &
state(γ,t2)|= has_value(LT_trust, K2) &
t2 > t1 +d & H1 > 0.5] ⇒ K2 ≥ K1

VP7 ≡ Reliable behaviour influences human willingness to collaborate with the robot.
If the robot provides consistent and reliable behaviour throughout the interaction time frame, it will improve human willingness to work closely with the robot:
VP7 ≡ ∀γ: TRACE, ∀t1, t2:TIME, ∀x1,x2, y1,y2:REAL
[state(γ,t1)|= has_value(reliable_behaviour, x1) &
state(γ,t2)|= has_value(reliable_behaviour, x2)
state(γ,t1)|= has_value(willingness_collaborate, y1) &
state(γ,t2)|= has_value(willingness_collaborate, y2) &
t2 > t1 +d ⇒[x2 ≥ x1 ⇒ L2 ≥ L1] & [x2 ≤ x1 ⇒ L2 ≤ L1]

8. CONCLUSION

In this paper, the computational representation of trust in human-robot collaboration was introduced. Several simulation experiments under related cases and distinctive parameter settings have been executed using a numerical programming environment. Although extensive empirical validation is left for upcoming work, these experimental results have demonstrated that the model can generate several trust circumstances when humans collaborate with robots for designated tasks. Our method has resulted in two interesting simulation results. First, a computational trust model developer can evaluate possible trustworthiness from the user's perspective in the first place by incorporating correct considerations (e.g., personalised parameters) into the computational human-robot collaboration trust function. Furthermore, using this model, we can analyse possible scenarios and make them available to improve interaction design and condition-action-support repertoire. In addition, we performed a mathematical analysis to identify potential equilibria points. Several expected simulation traces of the model have also been verified. These traces demonstrate that all the variables remained within their boundaries, and the predefined equilibria points were confirmed. The automated temporal trace language was also used to verify against simulation traces that ensued from the chosen scenario. Based on several parameter settings, these logical properties succeed, giving formal evidence that the model performs as anticipated. Besides that, it allows us to get more computational and theoretical insights into the temporal dynamics of trust formation in human-robot collaboration processes. As our concluding remark, we would like to highlight that our human-robot collaboration trust modelling method can be expanded to other formal modelling approaches where various integration and parameter settings are involved.

ACKNOWLEDGEMENT
This project is partially funded by Universiti Utara Malaysia (UUM) as part of a doctoral scholarship.

REFERENCES

Ajenaghughrure, I.B., Sousa, S. & Lamas, D. (2021). Psychophysiological modeling of trust in technology. *Proc. ACM on Human-Comp. Interaction*, **5**: 1 - 25.

Au, T., Hoek, P.J. & Lo, E. (2018). Combat analysis of joint force options using agent-based simulation. *2018 Military Commun. Inform. Syst. Conf. (MilCIS 2018)*, pp. 1-7.

Aziz, A.A. & Ghanimi, M.H.A. (2020). Reading with robots: A personalised robot-based learning companion for solving cognitively demanding tasks, *Int. J. Adv. Sci. Eng. Inform. Tech.*, **10**: 1489-1496.

Baker, A.L., Phillips, E., Ullman, D. & Keebler, J. (2018). Toward an understanding of trust repair in human-robot interaction. *ACM T. Interactive Intell. Syst.*, **8**:1 - 30.

Bauer, W., Bender, M., Rally, M., & Scholz, O. (2016). *Leichtbauroboter in der Manuellen Montage – Einfach Einfach Anfangen*, Fraunhofer IAO, Germany

Beer, J., Fisk, A.D. & Rogers, W.A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *J. Hum. Robot Interact*., **3**: 74-99.

Bodala, I.P., Kok, B.C., Sng, W. & Soh, H. (2020). Modelling the interplay of trust and attention in hri: an autonomous vehicle study. *2020 ACM/IEEE Int. Conf. Human-Robot Interaction*, pp. 145-147.

Bodenhagen, L., Fischer, K. & Weigelin, H.M. (2017). The influence of transparency and adaptability on trust in human-robot medical interactions. *2nd Worksh. Behav. Adaptation, Interaction Learn. Assist. Robot*, 28 August 2017, Lisbon, Portugal.

Boos, L. & Moshkina, L.V. (2019). Conveying robot state and intent nonverbally in military-relevant situations: an exploratory survey. *In* Chen, J. (Eds), *Advances in Human Factors in Robots and Unmanned Systems*. Springer, Washington DC, pp. 181-193.

Bütepage, J. & Kragic, D. (2017). *Human-Robot Collaboration: From Psychology to Social Robotics*. Available online at: *ArXiv, abs/1705.10146* (Last access date: 19 March 2022).

Chen, J.Y.C., Lakhmani, S.G., Stowers, K., Selkowitz, A.R., Wright, J.L. & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theor. Issues Ergonomics Sci.*, **19**: 259–282.

Chen, J.Y. (2018). Human-autonomy teaming in military settings. *Theor. Issues Ergonomics Sci.*, **19**: 255 - 258.

Phillips, E., Zhao, X., Ullman, D. & Malle, B.F. (2018). What is human-like? Decomposing robots' human-like appearance using the anthropomorphic robot (ABOT) database. 2018 *ACM/IEEE Int. Conf. Human-Robot Interaction* (HRI 2018), pp. 105–113

Galin, R. & Meshcheryakov, R. (2020). Human-robot interaction efficiency and human-robot collaboration. *Robot.: Ind. 4.0 Issues & New Intell. Contr. Paradigms*, pp. 55-63.

Ghute, M., Kamble, K., & Korde, M. (2018). Design of military surveillance robot. *2018 First Int. Conf. Secure Cyber Comput. Comm. (ICSCCC 2018)*, pp. 270-272.

Guo, X., Huang, Y., Gamborino, E., Tseng, S., Fu, L. & Yeh, S. (2019). Inferring human feelings and desires for human-robot trust promotion. *Cross-Cultural Design. Methods, Tools User Experience (HCII 2019)*, pp. 365-375.

Haring, K., Tobias, J., Waligora, J., Phillips, E., Tenhundfeld, N.L., Lucas, G.M., Visser, E., Gratch, J. & Tossell, C.C. (2019). Conflict mediation in human-machine teaming: using a virtual agent to support mission planning and debriefing. *28th IEEE Int. Conf. Robot Human Interactive Comm. (RO-MAN 2019)*, pp. 1-7.

Hidalgo, M., Reinerman-Jones, L. & Barber, D. (2019). Spatial ability in military human-robot interaction: A state-of-the-art assessment. *Int. Conf. Human-Computer Interaction (HCI 2019)*, pp. 363-380.

Hoff, K. A. & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, **57**: 407–434.

Honig, S. & Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: literature review and model development. *Front. Psy.*, **15**: 1-21.

Javaid, M., Estivill-Castro, V. & Hexel, R. (2020). Enhancing human trust and perception of robots through explanations. *13th Int. Conf. Adv. Comp. Human Interaction, 21-25 November 2020,* Valencia, Spain.

Khavas, Z., Ahmadzadeh, R. & Robinette, P. (2020). Modeling trust in human-robot interaction: A survey, *Int. Conf. Soc. Robot. (ICSR 2020)*, Madrid, Spain, pp. 529-541.

Kwiatkowska, M. & Lahijanian, M. (2016). Social trust: a major challenge for the future of autonomous systems, *AAAI Fall Symposia 2016*, Virginia, USA.

Langer, A. Feingold-Polak, R., Mueller, O., Kellmeyer, P., & Levy-Tzedek, S. (2019). Trust in socially assistive robots: Considerations for use in rehabilitation. *Neuroscience Bio-behavioural Rev.*, **104**: 231–239.

Lazanyi, K. & Hajdu, B. (2017). Trust in human-robot interactions. *2017 IEEE 14th Int. Conf. Informatics*, pp. 216-220.

Lewis, M. Sycara, K., & Walker, P.M. (2018). The role of trust in human-robot interaction. *In* Abbass, H.A., Scholz, J. & Reid, D.J. (Eds.), *Foundations of Trusted Autonomy*, Springer, Switzerland., pp. 135–159.

Marvel, J. & Norcross, R. (2017). Implementing speed and separation monitoring in collaborative robot workcells. *Robot. Comp. Integrated Manuf.*, **44**: 144-155.

Onyeulo, E.B. & Gandhi, V. (2020). What makes a social robot good at interacting with humans? *Inform.*, **11**: 43-56.

Ososky, S., Sanders, T., Jentsch, F., Hancock, P. & Chen, J. Y. C. (2014). Determinants of system transparency and its influence on trust in and reliance on unmanned robotic systems. *Unmanned Syst. Tech. XVI*, 90840E.

Pearson, C.J. (2019). *How Cognitive Risk Types Influence Trust and Reliance Across Automation Stages*. PhD Thesis, North Carolina State University, North Carolina.

Perelman, B., Evans, A., Schaefer, K. & Hill, S. (2018). Attitudes toward risk and effort trade-offs in human-robot heterogeneous team operations. *Proc. Human Factors Ergonomics Soc. Annual Meeting*, **62**: 1098 - 1102.

Pickering, J. B., Engen, V. & Walland, P. (2017). The interplay between human and machine agency. *Int. Conf. Hum.-Comp. Inter. (CHI 2019)*, pp. 47–59.

Razin, Y. & Feigh, K. (2020). Hitting the road: Exploring human-robot trust for self-driving vehicles. *2020 IEEE Int. Conf. on Human-Machine Sys. (ICHMS 2020)*, pp. 1-6.

Robert, L., Alahmad, R. & Esterwood, C. (2020). A review of personality in human-robot interactions. *Foundations and Trends in Info. Sys.*, **4**: 107–212.

Salem, M., Lakatos, G., Amirabdollahian, F.& Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. *10th ACM/IEEE Int. Conf. Human-Robot Interaction (HRI 2015)*, pp. 1–8.

Satterfield, K., Baldwin, C.L., Visser, E.D. & Shaw, T. (2017). The influence of risky conditions in trust in autonomous systems. *Human Factors Ergonomics Soc. Annual Meeting*, **61**: 324 - 328.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L. & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors*, **58**: 377–400.

Treur, J. (2020). *Network-Oriented Modeling for Adaptive Networks: Designing Higher-Order Adaptive Biological, Mental and Social Network Models*. Springer, Switzerland.

Ullman, D. & Malle, B. (2016). The effect of perceived involvement on trust in human-robot interaction. *2016 11th ACM/IEEE Int. Conf. Human-Robot Interaction (HRI 2016)*, pp. 641–642.

Xu, P., Zeng, Q., Zhang, G., Zhu, C. & Zhu, Z. (2019). Design of control system and human-robot-interaction system of teleoperation underwater robot. *Int. Conf. Intell. Robot. Appl.* (*ICIRA 2019)*, pp. 649-660.

A FRAMEWORK FOR ASSESSING THE IMPACTS OF POTENTIALLY DISRUPTIVE MILITARY TECHNOLOGIES

José Paulo Silva Bartolomeu[1] & Pedro B. Água[2]

[1]Centro de Investigação e Desenvolvimento, Instituto Universitário Militar, Portugal
[2]CINAV, Escola Naval, Instituto Universitário Militar, Portugal

*Corresponding author: pedroagua@escolanaval.pt

ABSTRACT

This paper proposes an alternative framework for assessing the impacts of potentially disruptive military technologies across all relevant dimensions. The research follows a critical thinking methodology supported by alternative analysis techniques. The proposed framework includes strategic, operational, tactical, technical and organizational dimensions. Political, economic, military, cultural, and legal factors are the variables for the strategic dimension. The variables to assess the operational dimension are performance, congruence, and opportunity. Secrecy and tactics, techniques, and procedures are the tactical variables. The technical dimension includes performance, maturity, and interconnectedness. Internal support, pacing gap and cost are the variables within the organizational dimension. The convergence of the assessed impact on variables and dimensions reveals the impact of a specific technology as null, moderate, high or revolutionary. The proposed framework for assessing the impact of potentially disruptive military technologies informs policymakers and industry leaders, as well as supports decisions about technology investment, defense capabilities and related strategies.

Keywords: *Disruptive technologies; framework; impact; military; warfare*.

1. INTRODUCTION

Today's context may be characterized as "an age of chaos", in a brittle, anxious, nonlinear and incomprehensible (BANI) world (Cascio, 2020). Technological advancements in data, artificial intelligence (AI), space, hypersonic technology, quantum computing, materials and biotechnology are transforming society and weaponry at a very high pace (Future of Defense Task Force, 2020, NATO STO, 2020).

Oftentimes, when a technology first emerges, its disruptive potential is not obvious. The disruption only occurs later on, once it has been applied or combined in an innovative way. However, in some other cases, a scientific breakthrough can lead to one or a series of disruptions (NAS, 2010). Disruptions often present challenges and may have negative connotations, yet moments of dramatic change also provide opportunities. The effects of disruptive technologies depend on the perspective and their distribution across the affected groups (Boucher *et al*., 2020).

Several authors, such as Liotta & Lloyd (2005) and Stojkovic & Dahl (2007), highlighted the potential impacts of future technologies and innovation on defense planning. In view of this, the NATO Science and Technology Council has requested the NATO Science and Technology Organization Group - a network of nearly 5,000 scientists, engineers and analysts - to watch for technology changes (*Technology Watch*), as well as identify and document its disruptive potential for the Alliance (NATO, 2018). The impact of disruptive technologies is also a concern for the European Union (EU) as evidenced from a conference held in Lisbon in 2021 under the theme of *The Impact of Disruptive Technologies on Defense* (Portuguese Presidency of the Council of the

EU, 2021). Therefore, while complicated, assessing the impact of disruptive technologies is critical. The Five Whys is a diagnostic technique that aids in identifying the root cause(s) of problems. It starts with the definition of the problem to focus on, after which initial causes to the problem are identified, followed by the question "Why is this a problem?" for each initial cause and "why" as many times as necessary (NATO, 2017).

The selected research focus for this paper is disruptive military technologies. Table 1 presents the application of the Five Whys technique to the problem of assessing the impacts of potentially disruptive military technologies. This paper is centered on technologies suitable for military applications and discusses possible solutions to tackle the root causes of the problem. As they already focus on the problem, no further limitations are specified.

Table 1: Assessing the impacts of potentially disruptive military technologies: problems and possible root causes.

Focus problem	**Assessing the impacts of potentially disruptive military technologies is complicated.**	
Initial causes	Some disruptive military technologies are unknown or unexpected	Existing assessments focus on some factors, but others are not considered (e.g., NATO STO, 2020)
Why is this a problem?	It triggers sudden and unexpected effects or inability to respond to potential threats and attacks	Assessments are not the most appropriate to support planning
Why?	Lack of capacity and / or inability to adjust operational concepts to face it in time or properly	Critical thinking, innovation, and "out of the box" factors are not included in the analysis
Why?	Lack of technological foresight, intelligence (and / or good counterintelligence by the other part), training, will, doctrine, and/or equipment	Assessments, as well as planning, tend to follow traditional patterns
Why?	Lack of / bad defense planning processes, lack of motivational objectives, lack of investment in research and development, defense programs, and / or military capabilities	It is easier to follow well known and tested procedures than to create new ones
Why?	Policymakers and industry leaders are not informed or sensibilized about the impact of potentially disruptive military technology across all relevant dimensions (Root cause)	Lack of academic and organizational frameworks to make adequate assessments (Root cause)

The purpose of this research is to develop a framework for assessing the impacts of potentially disruptive military technologies across all relevant dimensions. The specific objectives (SO) to operationalize it are: 1) Deduct the common features of disruptive military technologies; 2) Assess the main enablers and constraints for exploiting disruptive military technologies. The research question is: *Which framework should be used for assessing the impacts of potentially disruptive military technologies across all relevant dimensions?*

In addition to the introduction and conclusion, this paper has two additional sections. The second section reviews the literature and explains the methodology. The third section assesses the common features, enablers and constraints for exploiting potentially disruptive military technologies, and proposes a framework for assessing its impact.

2. LITERATURE REVIEW AND METHODOLOGY

This section clarifies concepts and approaches relevant to the undertaken study.

2.1 Key Concepts

From literature, it is possible to find several definitions of disruptive technologies. NAS (2010) defined it as "*An innovative (although not necessarily new) technology that triggers sudden and unexpected effects*". Brimley *et al*. (2013, p. 4) argued that "*What makes a technology "game changing," "revolutionary," "disruptive" or a "killer application" is that it both offers capabilities that were not available – and were in many ways previously unimaginable – a generation earlier and in so doing provokes deep questions whose answers are not readily available.*"

In the context of defense and security, Harald Andås (2020) proposed the following definition: "*Disruptive technologies are technological developments that change the conduct of conflict and the rules of engagement.*" NATO STO (2020) described it as "*Technologies that are expected to have a major, or perhaps revolutionary, effect on defense, security, or enterprise functions.*" The disruptive effects have significance within a limited time frame and force the planning process to adapt and change long-term goals for concepts, strategy and planning (Andås, 2020).

Disruptive technologies are hard to foresee or identify, and occur infrequently (NAS, 2010). However, Pierce (2004) argued that "*Militaries do not succumb to disruptive innovations because of a lack of foresight, but a lack of insight. Disruptive innovation (including technological) is one that appears to sneak onto the battlefield – because senior military leaders failed to recognize the threat it posed – and then out-performs the established way of fighting and defeats the unsuspecting military.*" A variety of factors such as a scientific breakthrough or a new manufacturing method, power source, weapons system or platform provide potential for game-changing technologies (Brimley *et al*., 2013).

Successfully assessing the potential impact of disruptive technologies requires consideration for current and future threats, legal and policy constraints, political factors, investment decisions, as well as estimating the potential for organizational entrepreneurial drive and risk tolerance (NATO STO, 2020). However, other factors are relevant as well, such as congruence, perspectives, societal values, organizational culture and time, with the synergies among these factors considered as "convergence" (Brimley *et al*., 2013). On the other hand, Andås (2020) used convergence to define the merging of existing technologies to create new and better possibilities, enabling further development and maturity. The first approach is to successfully create and implement game-changing technologies and the later to drive the technological development cycle.

2.2 Methodology

Critical thinking and alternative analysis methodologies are crucial for coping with a BANI environment together with emerging and disruptive technologies. Therefore, critical thinking is the adopted research methodology. This paper follows a structured reasoning, as proposed by Paul & Elder (2009): purpose, key questions, assumptions, key concepts, facts and experiences to support conclusions, own points of view, as well as conclusions and implications. Furthermore, the Five Whys technique is used to determine the root causes of the problem by asking "why" several times, while the Concept Mapping technique is used to produce tables depicting suggested relationships between concepts (NATO, 2017).

3. BUILDING THE FRAMEWORK

This section starts with the Concept Mapping technique to identify the relevant dimensions and factors. Facts and experiences from disruptive military technologies consolidate the dimensions and factors and allow the assessment of the indicators for the framework.

3.1 Concept Mapping

Tables 2 and 3 relate concepts from the literature review and concepts suggested by the authors to identify factors and dimensions for the framework. Focus Issue 1, in Table 2, is based on SO 1. After constructing, revising and interpreting the concepts, several factors and dimensions emerged from this focus issue:

1) Performance and congruence: Operational dimension
2) Secrecy and TTP: Tactical dimension
3) Performance, maturity and interconnectedness: Technical dimension.

Table 2: Concept mapping for Focus Issue 1.

Focus Issue 1: What are the main features of disruptive military technologies?			
Concepts		**Emerging Factors**	**Emerging Dimensions**
From Literature Review	**Own Suggestions**		
Hard to foresee or identify; never seen before; unimaginable; sudden and unexpected effects	Surprise, lack of countermeasures	Secrecy	Tactical
Used in a different way		Tactics, techniques and procedures (TTP)	Tactical
Revolutionary effect; change the conduct of conflict and the rules of engagement	Improve performance of operational capabilities	Operational performance	Operational
Innovative (although not necessarily new) technology; merging of existing technologies; congruence		Congruence and interconnectedness	Operational and technical
A scientific breakthrough or a new manufacturing method, power source, weapons system or platform		Technical performance	Technical

Focus Issue 2, in Table 3, is based on SO 2. The factors and dimensions that emerged from this focus issue are as follows:

1) Political, economic, military, cultural and legal: Strategic dimension
2) Opportunity: Operational dimension
3) TTP: tactical dimension
4) Internal support, pacing gap and cost: Organizational dimension
5) Maturity: Technical

Table 3: Concept mapping for Focus Issue 2.

Focus Issue 2: What are the main enablers and constraints for using disruptive military technologies?			
Concepts		**Emerging Factors**	**Emerging Dimensions**
From Literature Review	**Own Suggestions**		
Current and future threats; limited time frame	Operational environment	Opportunity	Operational
	Current tactics and counter-tactics; organization, doctrine and training	TTP	Tactical
Policy constraints; perspectives; potential for organizational entrepreneurial drive and risk tolerance; organizational culture; and time	Oversight mechanisms; support; resources and infrastructure	Pacing gap, internal support and cost	Organizational
Political factors; military strategies; investment decisions; societal values; ethics; and legal constraints		Political, economic, military, cultural and legal	Strategic
	Maturity of technology	Maturity	Technical

3.2 Facts and Experiences from Disruptive Technologies

The strategic, operational, tactical, technical and organizational dimensions, as well as the identified factors, provide the main structure for the analysis and framework. The strategic dimension focuses on political, economic, military, cultural and legal factors.

The atomic bombings of Hiroshima and Nagasaki in World War II, the Russian cyber-attacks on Estonia in 2007 (Aday *et al*., 2019), the *Stuxnet* worm attack on Iranian nuclear facilities in 2010 (Holloway, 2015), as well as AI, are examples of disruptive military technologies with political impact because they managed to achieve strategic objectives or change the perceived offense-defense balance. The stability of requisite institutions to sustain innovation and being (or not) a regional or global superpower are also relevant indicators for assessing the political impact.

Railroad, Global Positioning System (GPS) and unmanned aerial vehicles (UAV) are examples of the economic impact from disruptive military technologies due to their diffusivity and adoption rate both in the military and civilian realms (Patterson, 2018; SmartSense, 2019; Wagner, 2019; Rodriguez, 2020). The financial and human resources; scientific, technical and engineering capabilities; infrastructure capacity; and level of investment in science and technology are also very critical. For instance, in 1846, the French pioneered the adoption of steam propulsion and screw propellers on auxiliary ships (Naval Revolution), but the British's economic strength gave them the ability to take the lead in applying such technologies (Krepinevich, 1994).

Disruptive technologies with high impact in military genetic, structural and operational strategies, such as the crossbow (Infantry Revolution), gunpowder (Artillery Revolution), railroad (Land Warfare Revolution), nuclear weapons (Nuclear Revolution), and computers and information-related technologies (Information Revolution) have a revolutionary impact on military strategy. The level of acceptability or resistance to certain technologies and applications for cultural, religious or ethical reasons is also relevant. For instance, the railroad was widely accepted by society (Rodriguez, 2020), while on the other hand, technologies such as autonomous lethal

weapons have raised a wide array of serious ethical, legal, operational, proliferation, moral and technological concerns (Stauffer, 2020).

The legal factor is related to the ineffectiveness or effectiveness of limitations from international conventions or norms. In 1139 AD, the Second Lateran Council (Council Fathers, 1139) prohibited the use of bows and crossbows against Christians, but they proved to be "*useful in the Crusades and, once introduced, could not be eradicated in any event*" (Guilmartin, 2020). The St. Petersburg Declaration of 1868 prohibited the use of exploding bullets (Rodriguez, 2020), while the 1997 Convention banned antipersonnel landmines (United Nations, 1997). However, the Treaty on the Non-Proliferation of Nuclear Weapons of 1968 and Comprehensive Nuclear-Test-Ban Treaty of 1996, as well as security guarantees, troop deployments, arms sales, nuclear umbrellas and sanctions threats, did not prevent India, Pakistan, North Korea and Iran from developing, acquiring or testing nuclear weapons (Rodriguez, 2020).

The operational dimension focuses on performance, congruence and opportunity. The performance level of the operational capabilities (prepare, project, engage, C3 [command, control, and consult], sustain, protect, and inform) associated with a specific technology is highly relevant to assess its operational impacts. Crossbows became game-changing in feudal Europe because the topography was unfavorable for mounted shock action, and they were easy to master and capable of killing powerful mounted warriors (Guilmartin, 2020), e.g., battles of Laupen in 1339 and Crecy in 1346 (Krepinevich, 1994). Unmanned combat aerial vehicles (UCAV) are very effective for targeting infantry and provide less collateral damage than missiles or aerial bombing deployed from manned aircrafts (Research and Markets, 2021). These examples provide high or revolutionary impact on the operational capability of engage.

By the mid-nineteenth century, railroads enabled quick mobilization and sustainment of large armies by moving soldiers, artillery shells and supplies at an unprecedented scale (Imperial War Museums, 2020; Rodriguez, 2020), thus was significant or revolutionary for the operational capabilities of project and sustain. Metamaterial adaptive camouflage allows the user to totally conceal itself from plain sight, and has the potential to disrupt the entire detection and intelligence paradigm (Kosal & Stayton, 2020), which is highly relevant to the protect capability.

Autonomous weapon systems act as a force multiplier that reaches into areas that were previously inaccessible, allowing humans to engage in the battlefield from remote locations, removing them from dangerous missions and increasing their effectiveness (Etzioni & Etzioni, 2017), thus being significant or revolutionary for the protect and engage operational capabilities. The same rational applies to the telegraph for the C3 capability; to GPS and satellites for the C3 and inform capabilities; and to computers and the Internet for the engage, C3, protect and inform capabilities.

Congruence allows the assessment of the integration of the technology itself with a concept for its use in a timely relevant situation. One of the best examples of congruence is the *Blitzkrieg* warfare (World War II). The integration of fast tanks, aircrafts and two-way radios into an operational concept of advanced maneuver warfare created synergies that produced a discontinuous shift in the balance of military power in Europe (Brimley *et al*., 2013).

Technology adoption in high operational tempo expresses opportunity. For instance, the machine gun was invented by the Americans, but its tactical value was first fully exploited by the Germans in 1914 (Pierce, 2004).

The tactical dimension focuses on secrecy and TTP. Secrecy can be measured by the level of surprise of the disruption, or lack of countermeasures and / or counter-countermeasures. For instance, there is still no specific evidence of who developed the original cyber weapon used in the Stuxnet Worm Attack at Iranian Nuclear Facilities (Holloway, 2015). In September 2019, Houthi rebels from Yemen surprised with the first known coordinated massive swarm UAV strike (10

UAVs) on two key oil installations inside Saudi Arabia, after defeating the Saudi air defense systems (Hubbard *et al*., 2020).

The impact motivated by technology can also be revealed by changes in TTP, including changes in size, organization and training of the units; boosting of research and development; and contribution for the effects at the operational level. For example, in World War II, the Gee-7000 radio navigation system, developed by the Royal Air Force, allowed bombers to fly in a long, tight formation in the dark, a tactic called single stream bomber saturation (Air Ministry, 1947). Tactics like this and radar countermeasures (e.g., chaff) swung the balance in favor of the Allies (Kosal & Stayton, 2020). However, the emergence of nuclear weapons and other highly destructive payloads changed low-sophistication and high-volume air tactics to high-sophistication and low-volume tactics. The new air campaign was and still is avoiding detection in order to deliver a few carefully selected highly precise and destructive weapons at critical command and control targets (Kosal & Stayton, 2020). Thereafter, projects such as RAINBOW and OXCART sought to explore new methods of achieving stealth to avoid radar detection (National Security Archive, 2013).

The technical dimension includes performance, maturity and interconnectedness. Technical performance can be measured in terms of speed, range, accuracy, lethality or survivability, among other possible indicators. For example, the hollow point bullet made firearms easier to clean, generally more accurate with greater range, and deadlier than other bullets (Rodriguez, 2020). Similar assessments can be made in revolutionary technological innovations, such as the English longbow (1340-1415), Japanese Type 93 "Long Lance" torpedo (1933–1945), atomic bomb, GPS, or stealth technologies.

The nine technology readiness levels (TRL) pioneered by John C. Mankins at National Aeronautics and Space Administration (NASA) in the 1980s were incorporated in the NASA Management Instruction 7100 and are commonly used to measure the maturity of a particular technology (Mankins, 1995). For instance, blockchain technology can enhance supply chain processes and strengthen cybersecurity across military services, but the requirements are not yet clearly defined (Croft, 2019), which means that it is on TRL 1. In 2020, the US Navy successfully test-fired, in the Pacific, a high energy laser weapon that can destroy aircrafts mid-flights (The Economic Times, 2020; Tanwar, 2020); this is categorized as TRL 6. In 2017, Russia tested and evaluated its Ural-9 unmanned ground vehicle (UGV) in Syria (Tanwar, 2020); this is categorized as TRL 7. The success of the Houthi rebels in the coordinated massive swarm UAV strike inside Saudi Arabia is an example of a TRL 9 product / system.

Interconnectedness enables the assessment of the potential for integration of two or more well-understood technologies where no correlation had previously been identified. Some examples include the adaptation of the method used to cast church bells for casting artillery weapons fuelled the Artillery Revolution (Krepinevich, 1994); the Predator system first flew in 1995, but only became a game-changer for US counterterrorism when integrated with GPS technology (Brimley *et al*., 2013); the concept of AI dates to the early 1950s but the main changes to boost its use only occurred in the past decade, including miniaturization of processors, spread of mobile and connected devices; and application of new types of algorithms exploiting leaps forward in machine learning (Missiroli, 2020).

Internal support, pacing gap and cost are the relevant factors within the organizational dimension. The level of internal support can be assessed with several indicators, such as senior leader top cover, small team participation, junior personnel promotion pathways and disguising disruptive innovations as sustaining ones (Scott *et al*., 2019). For example, the British introduced the tank in 1915 but lacked a coherent effort to pursue armor's potential due to the resistance of senior military leaders. The Germans supported innovators, learned from some the British's experiences, created the *Panzer* force, and successfully exploited armored warfare (Pierce, 2004).

Pacing gap refers to the time required to establish laws, regulations and oversight mechanisms for the safe development and effective implementation of a new technology. This is especially evident in international organizations. For instance, NATO has been slow in leveraging communication solutions (e.g., tactical assault kit communication system), which can interconnect its respective special operations forces (Gojowsky *et al*., 2018).

The impact indicators for assessing the cost of a particular technology are the size and type of investment (initial and support), human capital, infrastructure required, and replicability of a product once developed. For example, crossbows could kill the most powerful of mounted warriors, yet they could be attained and used by those with considerably less financing and less training (Guilmartin, 2020; Rodriguez, 2020). Software development and computational biology requires virtually no investment in infrastructure beyond computing power (NAS, 2010); while UCAV are low-cost alternatives to combat aircraft (Research and Markets, 2021). On the other hand, nanotechnology and biotechnology developments require significant investment in laboratory equipment (NAS, 2010).

3.3 Outputs and Assessments

Table 4 summarizes the findings of the study in terms of the individual indicators level for one variable, or their total, reflecting the impact of technology on that variable. Assessing the impact of all variables along a specific dimension reveals the impacts of a technology in that dimension. The convergence of the impacts from the various dimensions allows the assessment of a specific technology's impacts. Each variable and dimension should have the impact assessed as null, moderate, high or revolutionary.

Table 4: Framework for assessing the impacts of potentially disruptive military technologies.

Dimensions	Variables	Indicators	Impact
Strategic	Political	Strategic objectives partially attained, attained or overcame; perceived offense-defense balance altered; stability (in requisite institutions) to sustain innovation; and being a regional or global superpower	Null Moderate High Revolutionary
	Economic	Level of development of defense industries; private sector defense industrial base; number of new markets; new industries or technology sectors emerging; financial and human resources; scientific, technical, and engineering capabilities; infrastructure capacity; level of investment in science & technology; and diffusivity / adoption rate in military and civilian domains	
	Military	Level of changes in genetic, structural, and operational strategy	
	Cultural	Level of acceptability or resistance to certain technologies, and applications for cultural, religious or ethical reasons	
	Legal	Ineffectiveness or effectiveness of limitations from international conventions or norms	
Operational	Performance	Limited and secondary in nature to the associated operational capability; moderate overall, or of significant relevance to a limited subset only; significant to an operational capability; revolutionary to one operational capability or significant to more than one operational capability	
	Congruence	Level of integration of the technology itself with a concept for its use in a timely relevant situation	
	Opportunity	Operational tempo; failure into adoption; and/or adopted first by the opposer	

Tactical	Secrecy	Level of surprise of the disrupted / lack of countermeasures and / or counter-countermeasures	
	TTP	Level of changes in tactics, counter-tactics, techniques and procedures; changes in size, organization, and training of the units; boosting of research and development; and contribution for the effects on the operational level	
Technical	Performance	Speed, range, accuracy, lethality and / or survivability	
	Maturity	TRL 1 - Basic principles observed and reported; TRL 2 - Technology concept and / or application formulated; TRL 3 - Analytical and experimental critical function and / or characteristic proof-of-concept; TRL 4 - Component and / or breadboard validation in laboratory environment; TRL 5 - Component and / or breadboard validation in relevant environment; TRL 6 – System / subsystem model or prototype demonstration in a relevant environment; TRL 7 - System prototype demonstration in operational environment; TRL 8 - Actual system completed and qualified through test and demonstration; TRL 9 - Actual system proven through successful mission operations (Mankins, 1995)	
	Interconnectedness	Potential for integration with other technologies of two or more well-understood technologies where no correlation had previously been identified	
Organizational	Internal support	Credible senior leader top cover; small team participation; junior personnel promotion pathways; disguising disruptive innovations as sustaining ones (Scott *et al*., 2019)	
	Pacing gap	Time required to establish laws, regulations and oversight mechanisms for the safe development or implementation of a new technology	
	Cost	Size and type of investment (initial and maintenance); human capital required; infrastructure required; replication viability of a product once is developed	

4. CONCLUSION

The brittleness or illusory strength of systems and the anxiety enhanced by the media in a nonlinear and incomprehensible world provide the framework for the innovation and proliferation of current and future threats faced by nations. Policymakers and industry leaders must be aware of the impacts of disruptive military technology to make informed decisions about technology investment, defense capabilities, acquisition and strategies.

In order to develop a framework for assessing the impacts of potentially disruptive military technologies across all relevant dimensions, this research followed a critical thinking methodology supported by alternative analysis techniques. The evidence extracted from factual disruptive technologies, in addition to the verification of the dimensions and variables extracted from the literature and concept mapping, were instrumental in identifying and supporting the framework indicators.

The proposed framework includes strategic, operational, tactical, technical and organizational dimensions. Political, economic, military, cultural and legal factors are the variables for the strategic dimension. The variables to assess the operational dimension are performance, congruence and opportunity. Secrecy and TTP are the tactical variables. The technical dimension includes performance, maturity and interconnectedness. Internal support, pacing gap and cost are

the variables within the organizational dimension. The convergence of the assessed impact on the variables and dimensions, supported in the framework indicators from Table 4, reveals the impact of a specific technology as null, moderate, high or revolutionary.

To conclude, the proposed framework for assessing the impact of potentially disruptive military technologies informs policymakers and industry leaders, as well as supports decisions about technology investments, defense capabilities and related strategies. Further research is recommended to support decision-makers in defining criteria and weights for the variables and dimensions.

REFERENCES

Aday, S., Andžāns, M., Bērziņa-Čerenkova U., et al (2019). Hybrid threats: 2007 cyber-attacks on Estonia. *In Hybrid Threats. A Strategic Communications Perspective*. NATO Strategic Communications Centre of Excellence, Riga, Latvia, pp. 51-70.

Air Ministry (1947). *Introductory Survey of Radar: Part II*. Air Ministry, UK.

Andås, H. (2020). *Emerging Technology Trends for Defence and Security*. Norwegian Defence Research Establishment, Kjeller, Norway.

Boucher, P., Bentzen, N., Laţici, T., Madiega, T., Schmertzing, L. & Szczepański, M. (2020). *Disruption by Technology: Impacts on Politics, Economics and Society*. European Union, Brussels, Belgium.

Brimley, S., FitzGerald, B. & Sayler, K. (2013). *Game Changers: Disruptive Technology and US Defense Strategy*. Center for a New American Security, Washington, DC.

Cascio, J. (2020). *Facing the Age of Chaos*. Available online at: https://medium.com/@cascio/facing-the-age-of-chaos-b00687b1f51d (Last access date: 6 August 2022).

Council Fathers. (1139). Second Lateran Council – 1139 A.D. *In Papal Encyclicals Online*. Available online at: http://www.papalencyclicals. net/Councils/ecum10.htm (Last access date: 10 August 2022).

Croft, H. (2019). 10 Disruptive Technologies of the 2010s. *Defence iQ*. Available online at: https://www.defenceiq.com/defence-technology/editorials/10-technologies-in-the-2010s-that-disrupted-defence (Last access date: 10 August 2022).

Etzioni, A. & Etzioni, O. (2017). Pros and cons of autonomous weapons systems. *Milit Rev.*, May-June2017: 72–81.

Future of Defense Task Force. (2020). *Future of Defense Task Force Report 2020*. House Armed Services Committee, Washington, DC.

Gojowsky, T., Koegler, S., Haspels, B., Haar, F. & Wetteland, S. (2018). *Resistance to Innovation in NATO*. Available online at: https://thestrategybridge.org/the-bridge/2018/8/16/resistance-to-innovation-in-nato (Last access date: 10 August 2022).

Guilmartin, J. F. (2020). *Military Technology*. Available online at: https://www.britannica.com/technology/military-technology (Last access date: 10 August 2022).

Holloway, M. (2015). *Stuxnet Worm Attack on Iranian Nuclear Facilities*. Available online at: http://large.stanford.edu/courses/2015/ph241/holloway1 (Last access date: 10 August 2022).

Hubbard, B., , P. & Reed, S. (2019). *Stuxnet Worm Attack on Iranian Nuclear Facilities*. Available online at: https://www.nytimes.com/2019/09/14/world/middleeast/saudi-arabia-refineries-drone-attack.html (Last access date: 10 August 2022).

Imperial War Museums. (2020). *Transport and Supply During the First World War*. Available online at: https://www.iwm.org.uk/history/transport-and-supply-during-the-first-world-war (Last access date: 10 August 2022).

Kosal, M. E. & Stayton, J. W. (2020). Meta-materials: Threat to the global status quo? *In* Kosal, M.E. (Ed.), *Disruptive and Game Changing Technologies in Modern Warfare. Development, Use, and Proliferation.* Springer, Berlin, Germany, pp. 135-154.

Krepinevich, A.F. (1994). *Cavalry to Computer: The pattern of Military Revolutions*. Available online at: https://nationalinterest.org/article/cavalry-to-computer-the-pattern-of-military-revolutions-848 (Last access date: 10 August 2022).

Liotta, P.H. & Lloyd, R.M. (2005). From here to there: the strategy and force planning framework. *Naval War Coll.,* **58**: article 7.

Mankins, J.C. (1995). *Technology Readiness Levels*. Advanced Concepts Office, Office of Space Access and Technology, National Aeronautics and Space Administration (NASA), Washington, DC, US.

Missiroli, A. (2020). Game of drones? How New technologies Affect Deterrence, Defence and Security. Available online at: https://www.nato.int/docu/review/articles/2020/05/05/game-of-drones-how-new-technologies-affect-deterrence-defence-and-security/index.html (Last access date: 10 August 2022).

NAS (National Academy of Sciences) (2010). *Persistent Forecasting of Disruptive Technologies*. Available online at: https://www.nap.edu/catalog/12557/persistent-forecasting-of-disruptive-technologies (Last access date: 10 August 2022).

National Security Archive (2013). *The U-2's Intended Successor: Project OXCART, 1956-1968*. Available online at: https://nsarchive2.gwu.edu//NSAEBB/NSAEBB434/docs/U2%20-%20Chapter%206.pdf (Last access date: 10 August 2022).

NATO (North Atlantic Treaty Organization) (2017). *AltA: The NATO Alternative Analysis Handbook*, *2nd Ed.* North Atlantic Treaty Organization (NATO), Brussels, Belgium.

NATO (North Atlantic Treaty Organization) (2018). *Framework for Future Alliance Operations*. NATO, Brussels, Belgium.

NATO STO (North Atlantic Treaty Organization Science & Technology Organization) (2020). *Science & Technology Trends 2020-2040*. North Atlantic Treaty Organization (NATO), Science & Technology Organization (STO), Brussels, Belgium.

Patterson, J. (2018). *An Aerial View of the Future – Drones in Construction*. *Geospatial World*. Available online at: https://www.geospatialworld.net/blogs/an-aerial-view-of-the-future-drones-in-construction/ (Last access date: 10 August 2022).

Paul, R. & Elder, L. (2009). *The Miniature Guide to Critical Thinking: Concepts and Tools, 6th Ed*. The foundation for Critical Thinking, Foundation for Critical Thinking, Santa Barbara, California

Pierce, T.R. (2004). *Warfighting and Disruptive Technologies: Disguising Innovation*. Taylor & Francis, New York.

Portuguese Presidency of the Council of the EU (2021). *R&T Conference: Impact of Disruptive Technologies on Defence*. Available online at: https://eda.europa.eu/news-and-events/events/2021/04/20/default-calendar/impact-of-disruptive-technologies-on-defence (Last access date: 10 August 2022).

Research and Markets (2021). *Global Unmanned Aerial Vehicles Market (2020 to 2025) - Growth, Trends, and Forecasts*. Available online at: https://www.globenewswire.com/news-release/2021/01/20/2161392/0/en/Global-Unmanned-Aerial-Vehicles-Market-2020-to-2025-Growth-Trends-and-Forecasts.html (Last access date: 10 August 2022).

Rodriguez, R. (2020). Game-changing military technologies: adoption and governance. *In* Kosal, M.E. (Ed.), *Disruptive and Game Changing Technologies in Modern Warfare. Development, Use, and Proliferation.* Springer, Berlin, Germany, pp. 13-29.

Scott, B. Kaahaaina, N. & Stock, C. (2019). *Innovation in the Military*. Available online at: https://smallwarsjournal.com/jrnl/art/innovation-military (Last access date: 10 August 2022).

SmartSense. (2019). *Global Positioning System (GPS): The Past, Present, and Future*. Available online at: https://blog.smartsense.co/gps-past-present-future (Last access date: 10 August 2022).

Stauffer, B. (2020). Stopping Killer Robots: Country Positions on Banning Fully Autonomous Weapons and Retaining Human Control. Available online at: https://www.hrw.org/report/2020/08/10/stopping-killer-robots/country-positions-banning-fully-autonomous-weapons-and (Last access date: 10 August 2022).

Stojkovic, D. & Dahl, B. R. (2007). *Methodology for Long Term Defence Planning*. Norwegian Defence Research Establishment, Kjeller, Norway.

Tanwar, S. S. (2020). *Disruptive Technologies: Impact on Warfare & Their Future in Conflicts of 21st Century*. Available online at: https://www.claws.in/disruptive-technologies-impact-on-warfare-their-future-in-conflicts-of-21st-century/ (Last access date: 10 August 2022).

The Economic Times. (2020). US Navy successfully tests a laser weapon that can destroy aircraft mid-flight. Available online at: https://economictimes.indiatimes.com/news/defence/us-navy-successfully-tests-a-laser-weapon-that-can-destroy-aircraft-mid-flight/articleshow/75919159.cms?from=mdr (Last access date: 10 August 2022).

United Nations. (1997). *Convention on the Prohibition of the Use, Stockpiling, Production and Transfer of Anti-Personnel Mines and on Their Destruction*. United Nations, Oslo, Norway.

Wagner, I. (2019). Commercial UAVs - Statistics & Facts. Available online at: https://www.statista.com/topics/3601/commercial-uavs/ (Last access date: 10 August 2022).