

Exergy graph-based fault detection and isolation of a gas-to-liquids process

Sarita Greyling* George van Schoor**
Kenneth Richard Uren* Henri Marais*

* *School of Electrical, Electronic, and Computer Engineering, Faculty of Engineering, North-West University, Potchefstroom 2531, South Africa (e-mail: sarita.svl@gmail.com/*

kenny.uren@nwu.ac.za/henri.marais@nwu.ac.za).

** *Unit for Energy and Technology Systems, Faculty of Engineering, North-West University, Potchefstroom 2531, South Africa (e-mail: george.vanschoor@nwu.ac.za)*

Abstract: With the sheer size of modern process plants, the Fault Detection and Isolation (FDI) field continues to gain popularity. FDI is a sophisticated concept which aims to detect and isolate anomalies that occur within a plant to avoid losses of personnel, damages to the environment, and financial implications. It does so in a way which is more direct, efficient and safer than what human operators are capable of. One approach to FDI is to consider the exergy characterisation of a system. By describing the exergy of the system units and streams, in this case a gas-to-liquids (GTL) process plant, the various process variables are encapsulated under a universal energy-domain parameter. The advantage of this being that it can describe the physical states as well as the chemical characteristics of the process. Previous work which utilised exergy characterisation along with a fixed-threshold approach showed promise. This study however, shows that the approach falls short when presented with 3 % faults. These results motivated the investigation of utilising attributed graphs, which package exergy data into a framework that preserves structural information of the plant. The usefulness of finding similarities (called graph matching) between the graphs constructed of operational conditions and pre-collected fault conditions to detect and isolate faulty conditions, is demonstrated. The technique performs well when considering fault magnitudes bigger than 8 % but deteriorates when applied to smaller, 3 % faults. The poor performance could be ascribed to the graph matching aspect, which is described by a single distance value that discards dimensionality. Future work will therefore look into the graph matching technique specifically, aiming to retain more informative dimensions.

Keywords: Fault detection, Fault isolation, Exergy, Gas-to-liquids, Graph matching, Attributed graphs

1. INTRODUCTION

In most industrial process plants, operators are tasked with the monitoring and management of operations. This means overseeing a large number of units and associated process variables. When anomalies occur within the plant, the operators are expected to detect, diagnose and rectify the situation in the shortest possible time. As technology progresses these industrial processes are enhanced, resulting in even more complex systems. Consequently, the operators' responsibilities could escalate beyond their capabilities. Sometimes the mishandling of events by operators result in costly incidents, not only risking human life and the environment, but also causing detrimental financial situations. Two well-known incidents that illustrate this, is the methyl isocyanate (MIC) leak in Bhopal, India which claimed thousands of lives, according to Kletz (1998). The second incident, as highlighted by Venkatasubramanian et al. (2003), was the Kuwait Petrochemical Mina al Ahmadi oil refinery explosion

which resulted in \$100 million in damages. This is where Fault Detection and Diagnosis (FDD) schemes are of interest. By employing an appropriate FDD approach, the efforts required of operators are reduced, the rectification of anomalies are more efficient and the associated cost and health risks are limited. The advancement of FDD approaches would therefore specifically benefit volatile and expensive processes such as seen in the petrochemical industries (PCIs). FDD approaches are generally categorised as being either model-based or data-driven; the main difference being whether an analytical model is present. Model-based methods utilises analytical or structural models of a process and its behaviours, encapsulating both normal and faulty aspects. Data-driven techniques do not make use of explicit models; but rather derive mathematical relations, based on provided historical process data, between faults and the effects thereof. When surveying literature pertaining to chemical processes, the approaches seem to lean towards being data-driven. The most prominent approaches being either

statistical, as applied by Choi et al. (2005), Xie et al. (2013), Ghosh et al. (2014), Fezai et al. (2018), and Dong and Qin (2018); or machine learning as demonstrated by Watanabe and Hirota (1991), and Sorsa et al. (1991), to name but a few. Venkatasubramanian et al. (2003) states that it would not be impossible to develop analytical models of petrochemical (PC) processes but that it would be exceptionally challenging. Recent research, such as done by Chiang and Braatz (2003), Maki et al. (2004) and Chiang et al. (2015), show hybrid approaches taking the forefront. The hybridisation, which is usually a combination of model-based and data-driven techniques, endeavours to leverage the advantages afforded by the different approaches whilst minimising the drawbacks. Of noticeable interest is the approach proposed by Chiang and Braatz (2003) which aimed at combining causal maps and Partial Least Square (PLS) methods in order to include process connectivity information. Much in the same direction of thinking, Marais et al. (2019), proposed a hybrid approach which makes use of energy properties of the system. Not only is the energy description a unifying parameter across different domains, but it is also a way of abstracting data. The energy properties are then packaged in such a manner that the physical structural information of the system is retained. In the work of Greyling et al. (2019) the same approach was applied to a gas-to-liquids (GTL) process, incorporating exergy rather than energy. By monitoring the exergy, additional information is encapsulated, specifically regarding the chemical characteristics of the system. The results recorded in the work done by Greyling et al. (2019) showed promise, however the question that arose was whether a fixed-threshold approach would still work if the system faults' magnitudes were smaller.

This paper is divided into two parts. The first part focusses on evaluating the threshold approach performance when presented with 3 % faults. The results proved to be less than stellar, which meant some alterations were required. It must however be emphasised that the exergy and structural information concepts still hold promise; the issue seemed to be the fixed-threshold applied. Therefore, keeping with the exergy characterisation and wanting to preserve the structural information, Ould-Bouamama et al. (2014) suggest that a graphical method would allow for both. Such graphical approaches would also provide different mathematical schemes of detecting and isolating considered faults. Most of the graphical approaches reviewed by Ould-Bouamama et al. (2014) make use of graphs to describe system properties and relevant causal relations. For this study the most suitable graphical approach was chosen to be attributed graphs along with graph matching, a popular technique that quantifies the dissimilarities of compared graphs. The second part of the paper therefore demonstrates the usefulness of comparing operational graphs to faulty graphs (stored in a database) in order to detect and isolate faulty conditions.

The paper starts off with briefly detailing the GTL model and exergy characterisation thereof. Section 3 goes on to describe the considered faults' detail and their locations. The threshold approach as applied to 3 % faults and the results obtained is summarised in Section 4. Section 5 gives a quick overview of attributed graphs and graph matching

and goes on to detail the methodology as these are applied to the GTL process. The fault detection and isolation results obtained are given in Section 6 with the paper being concluded in Section 7.

2. THE GAS-TO-LIQUIDS PROCESS

A gas-to-liquids (GTL) process is used to transform gaseous feedstock, such as natural gas, into hydrocarbon liquids. A GTL process usually comprises of three principal sections as shown in Fig. 1. In the first section the natural gas is reformed to obtain synthesis gas, also referred to as syngas. The syngas is made up of a certain ratio of hydrogen (H_2) and carbon monoxide (CO), depending on the intended end-products. The produced syngas is then introduced to a Fischer-Tropsch reactor which converts the syngas into an array of hydrocarbons (syncrude). The upgrading section is used to rework the syncrude to hydrocarbon products of particular chain lengths. Since the upgrading section is quite complex only the first two sections of the GTL process, shown boxed in Fig. 1, are considered in this study. Interested readers are referred to the works of Rafiee and Hillestad (2010), Panahi et al. (2011), De Klerk (2011), and Knutsen (2013) for comprehensive information on the GTL process and the modelling thereof.

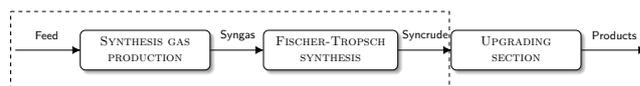


Fig. 1. A process diagram of a gas-to-liquids (GTL) process

2.1 Simulation model

In order to have a representative system to work with, a **steady-state** simulation model was constructed in the commercial process simulator, Aspen HYSYS[®]. No process variations were intentionally included in this study. The exact particulars on how the model was developed, the operating points and validation of the model are comprehensively documented in Greyling et al. (2019). However, some of the most important modelling aspects are highlighted as:

- (1) Autothermal reformer (ATR)
 - (a) No pre-reformer was included as there was no recycling to the ATR.
 - (b) The feedstocks used were pure methane (CH_4), steam ($H_2O(g)$), oxygen (O_2) and carbon dioxide (CO_2).
- (2) Fischer-Tropsch reactor (FTR)
 - (a) A plug flow reactor in conjunction with a separator was used to represent a multi-tubular fixed bed (MTFB) reactor.
 - (b) The syngas that was fed into the FTR was at a temperature of 210 °C.
- (3) Recycle
 - (a) 76.8 % of the unreacted syngas in stream 16 was recycled back to the FTR.
 - (b) The remaining 23.2 % was purged (stream 22).

Fig. 2 shows the Aspen HYSYS[®] process flow diagram of the developed GTL process. Note that the validation of

the simulation model comprised of comparing the attained product distribution to the theoretical distributions seen in literature.

2.2 Exergy characterisation

According to Dincer and Rosen (2013), exergy is defined as being a quantitative measure of an energy quantity's usefulness to perform work. Unlike energy which is based only on the first law of thermodynamics, exergy also takes into account the second law of thermodynamics. The second law states that entropy cannot decrease in any real process, therefore the ability to deliver valuable work is eventually lost. In other words exergy is not conserved and some exergy losses would occur which could be quantified by using the process' exergy balance (Magnanelli et al. (2018)). Therefore the most prominent advantage of using exergy is that it enables a manner of quantifying the quality of an energy stream or (more importantly) the efficiency of various elements. Consequently, any deviation of the known efficiencies could be indicative of an anomaly within the system. To characterise the GTL system the intrinsic exergy of each stream was calculated. It is important to note that exergy is always evaluated relative to a reference environment (RE). This means that the RE's intensive properties will determine the exergy. For physical exergy, these include temperature and pressure and are typically $T_0 = 25 \text{ }^\circ\text{C}$ and $P_0 = 101.325 \text{ kPa}$. However, the chemical exergy is based on an environment consisting of certain reference elements. Various methods exist for selecting and calculating the standard chemical exergy (b_{ch}^0) of these reference elements with the RE proposed by Szargut (2007) being the most distinguished one. In order to automatically calculate the physical and chemical exergy within Aspen HYSYS[®], *user variables* were developed. A *user variables* is a section of program code that the user creates, which can access various elements of the Aspen HYSYS[®] model. To calculate the physical exergy per mole

$$b_{ph} = (h - h_0) - T_0(s - s_0), \quad (1)$$

is used where h and s are the current enthalpy and entropy, respectively, and h_0 and s_0 the enthalpy and entropy at RE state. The total physical exergy (B_{ph}) is obtained by multiplying the stream's molar flow rate with the computed intrinsic physical exergy. The chemical exergy is calculated by making use of

$$b_{ch} = \sum x_{(i)} b_{ch(i)}^0, \quad (2)$$

with $x_{(i)}$ being the mole fraction and $b_{ch(i)}^0$ the standard molar chemical exergy of substance i . The utilised $b_{ch(i)}^0$ values were defined by Szargut's (2007). In order to utilise (2), the standard molar chemical exergies of all the relevant substances were firstly made available to the *simulation basis* by creating a *user property* tabulating the corresponding values. Not all substances' standard chemical exergies were readily available but were pre-calculated (Greyling et al. (2019)). Seeing as the GTL process would inevitably have multi-phase streams and since some substances have different standard chemical exergies based on their phase, Equation (2) was modified to account for this. Thus the total chemical exergy is the sum of the vapour, liquid and aqueous phases mathematically expresses as

$$b_{ch} = \sum x_{(i)v} b_{ch(i)v}^0 + \sum x_{(i)\ell} b_{ch(i)\ell}^0 + \sum x_{(i)a} b_{ch(i)a}^0. \quad (3)$$

Whenever a certain phase was not present, the phase exergy was assumed to be zero. The complete details on how the physical and chemical exergy *user variables* were developed is also discussed in Greyling et al. (2019). The assumption was made that the physical exergy (B_{ph}) and chemical exergy (B_{ch}) of all the streams and units were available to utilise. The next iteration of the research will look into using fewer data points that carry more importance.

3. SYSTEM FAULT SPECIFICATIONS

Before the considered faults are introduced, a recap of the terminology is crucial. According to Shah (2011), a *failure* is a permanent disruption of the system's operations. A *disturbance* is an unknown input that negatively affects the system's performance where a *fault* is any unintentional deviation of a parameter from its normal behaviour. A fault can be further classified based on its physical location or the effects on the system operation. Faults classified based on their location are system faults, actuator faults, and sensor faults. The different fault effects can be seen as additive, multiplicative, abrupt, incipient, intermittent, permanent, or transient. In order to evaluate the fault detection capabilities of the proposed approaches, eleven relevant fault conditions were identified. These eleven faults include system faults and actuator faults. For this study however, the proposed system need not distinguish between the two fault-categories. The reasoning behind how the faults were chosen, was established in Greyling et al. (2019). The details of faults and their location are summarised in Table 1. The locations of the faults are also visually depicted using danger triangles, as shown in Fig. 2. In order to keep track of the magnitude variation of the faults, fault IDs were assigned to each. The general form of the fault ID is given as F_{pq_r} , where p denotes the relevant GTL section, q the type of fault within the section and r the magnitude deviation considered; the magnitude variations being 3 %, 8 %, 9 %, 10 %, 11 %, 12 %, and 25 %. The most important datasets to take note of are:

- F_{pq_1} are the eleven faults that deviated with a magnitude of 3 %.
- F_{pq_4} are the eleven faults that deviated with a magnitude of 10 %.
- F_{pq_R} is a random selection of various magnitude deviations, *excluding* 3 and 10 % magnitudes, of each one of the eleven faults.

The specified datasets are shown highlighted in Table 1 for easy identification. The GTL model was modified individually and simulated for every fault tabulated, each time recording all the streams' physical exergy (B_{ph}) and chemical exergy (B_{ch}) to use as the exergy characterisation information.

4. A THRESHOLD APPROACH

It is important to note that the threshold approach developed, evaluated and the results documented in Greyling et al. (2019) were considering the 10 %

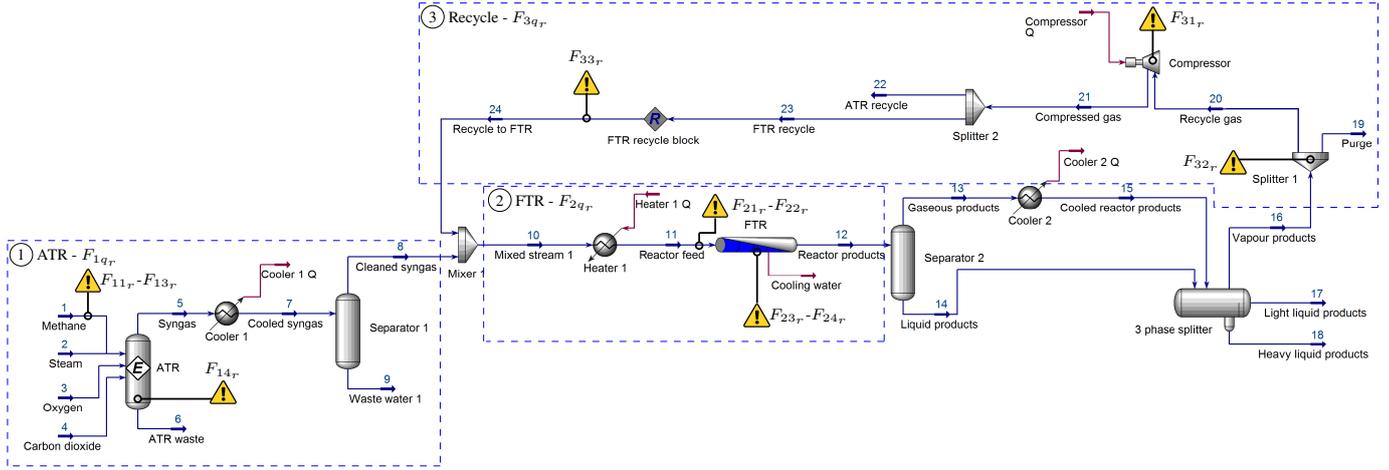


Fig. 2. The GTL process as developed in Aspen HYSYS[®] with fault locations indicated

Table 1. The faults' details and denotation

Fault ID [†]	Description	r							Location
		1	2	3	4	5	6	7	
F_{1q_r}									
ATR section									
F _{11_r}	Molar flow	+ 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Methane stream
F _{12_r}	Molar flow	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Methane stream
F _{13_r}	Pressure	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Methane stream
F _{14_r}	Pressure	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	ATR
F_{2q_r}									
FTR section									
F _{21_r}	Temperature	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Reactor feed stream
F _{22_r}	Leakage	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Reactor feed stream
F _{23_r}	Pressure	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	FTR
F _{24_r}	Temperature	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	FTR
F_{3q_r}									
Recycle section									
F _{31_r}	Pressure	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Compressor
F _{32_r}	Lower split ratio	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Splitter 1
F _{33_r}	Leakage	- 3 %	8 %	9 %	10 %	11 %	12 %	25 %	Recycle to FTR

[†] F_{pq_r} - p represents the section, q the fault number and r the magnitude deviation

(F_{pq₄}) deviations only. The approach performed very well, successfully detecting all eleven faults and the resultant isolability being 100 %. *Detection* being able to correctly indicate that a fault was present and *isolability* specifically referring to whether the faults were uniquely distinguishable from one another. Subsequently the question arose as to how well the approach would perform when small fault magnitudes are evaluated. To determine this, the following methodology was applied to the 3 % dataset (F_{pq₁}):

- (1) Firstly the collected exergy data, per stream, was normalised with respect to the normal condition.
- (2) Next a simple threshold function was applied to the normalised values in order to obtain a qualitative value for each entry. The threshold function used is described by:

$$y = \begin{cases} -1 & \text{if } z < (1 - \kappa) \\ 1 & \text{if } z > (1 + \kappa) \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

In (4), z represents the normalised exergy value being considered and y the magnitude of the resultant fault element. In order to assign an appropriate value to κ , the solver deviations seen within the Aspen HYSYS[®] environment were utilised. As

the simulation model was rerun - under identical conditions - small solver variations were noticed. To ensure that these simulation variations were distinguishable from the faults, the variances were quantified. This was achieved by calculating the statistical experimental error between 10 simulation runs. The threshold κ -value was found to be 0.00635; the precise calculation hereof shown in Greyling et al. (2019).

- (3) After applying the threshold function to the normalised data, a 20×2 qualitative matrix is obtained with the form

$$F_{pq_r} = \begin{bmatrix} y_{B_{ph}(stream1)} & y_{B_{ch}(stream1)} \\ \vdots & \vdots \\ y_{B_{ph}(stream24)} & y_{B_{ch}(stream24)} \end{bmatrix}. \quad (5)$$

- (4) When evaluating the detection and isolation performance
 - (a) any non-zero matrix would indicate a fault condition
 - (b) any identical matrices, for different fault conditions, would signify unisolability

Table 2 shows the qualitative matrices obtained for dataset F_{pq₁} after applying the threshold function. When evaluating the matrices, a shortcoming in terms of detection is evident. Seeing that the qualitative matrix of fault F_{23₁} is zero, the fault condition was not successfully detected. To calculate the detection rates of a proposed approach, a confusion matrix is drawn up. The idea behind a confusion matrix is to determine the number of instances where the decision of the approach:

- resulted in a *true negative* (TN) - the approach detected a fault-free condition and the true condition was indeed fault-free (value assigned to a)
- resulted in a *false negative* (FN) - the approach detected a fault-free condition and the true condition was faulty (value assigned to b)
- resulted in a *false positive* (FP) - the approach detected a fault condition and the true condition was actually fault-free (value assigned to c)

Table 2. Threshold results for dataset F_{pq_1}

Stream	F_{1q_1}				F_{2q_1}				F_{3q_1}											
	F_{11_1}		F_{12_1}		F_{13_1}		F_{14_1}		F_{22_1}		F_{23_1}		F_{24_1}		F_{31_1}		F_{32_1}		F_{33_1}	
	B_{ph}	B_{ch}																		
1	1	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	1	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	1	-1	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	-1	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	-1	-1	1	0	0	0	0	-1	-1	0	0	-1	0	-1	-1	-1	-1	-1
11	1	1	-1	-1	1	0	0	0	-1	0	0	0	0	0	0	-1	-1	-1	-1	-1
12	1	1	-1	-1	1	0	0	0	-1	0	0	0	0	0	0	-1	-1	-1	-1	-1
13	1	1	-1	-1	1	0	0	0	-1	0	0	0	0	0	0	-1	-1	-1	-1	-1
14	1	-1	-1	1	1	-1	1	-1	-1	1	1	0	0	1	-1	0	0	1	1	1
15	1	1	-1	-1	0	0	0	0	0	-1	-1	0	0	0	1	0	0	-1	-1	-1
16	1	1	-1	-1	0	0	0	0	0	-1	-1	0	0	0	0	0	0	-1	-1	-1
17	-1	-1	0	0	1	1	0	0	-1	0	-1	0	0	0	0	0	0	0	0	1
18	1	1	-1	-1	1	1	0	0	-1	0	1	0	0	0	0	0	1	0	1	0
19	1	1	-1	-1	0	0	0	0	0	1	1	0	0	0	0	0	1	1	-1	-1
20	1	1	-1	-1	0	0	0	0	0	1	-1	0	0	0	0	0	-1	-1	-1	-1
21	1	1	-1	-1	0	0	0	0	0	-1	-1	0	0	0	0	-1	0	-1	-1	-1
24	1	1	-1	-1	0	0	0	0	0	1	1	0	0	0	0	-1	0	-1	-1	-1

- resulted in a *true positive* (TP) - the approach detected a fault condition and the true condition was faulty (value assigned to d)

Subsequently, the detection rates r_{FP} , r_{FN} , r_{TP} , and accuracy are calculated by making use of these assigned values. Ideally, the false positive rate (r_{FP}) and false negative rate (r_{FN}) should be 0 % and the true positive rate (r_{TP}) and accuracy 100 %. A confusion matrix is completed for the threshold approach and is shown in Fig. 3. Evaluating these obtained rates, it is clear that the threshold approach did not perform flawlessly, therefore motivating the development of a different approach.

CONFUSION MATRIX				DETECTION RATES		
		TRUE CONDITION		Rate	Formula	%
		Fault-free	Fault			
DETECTION CONDITION	Fault-free	True negative a	False negative b	$r_{FP} = \frac{b}{(a+c)} \times 100$		0
	Fault	False positive c	True positive d	$r_{FN} = \frac{c}{(a+d)} \times 100$		9.09
				$r_{TP} = \frac{d}{(b+d)} \times 100$		90.91
				Accuracy = $\frac{(a+d)}{(a+b+c+d)} \times 100$		90.91

Fig. 3. Confusion matrix and detection rates of threshold approach applied to dataset F_{pq_1}

5. GRAPH THEORETICAL APPROACH

5.1 Background

Graph theory has been in use since the 1730's and became very popular in the 1930's. It is mathematical in nature and the concepts thereof have diverse capabilities. A broad spectrum of applications are seen throughout literature, including pattern recognition, transportation and even economics. A graph essentially consists of an ordered pair $G = (V, E)$, where V is the set of vertices (also called nodes) and E the set of edges (sometimes referred to as links or arcs). Usually vertices represent certain properties of a system, whereas the edges are used to describe the incidence relation of the vertices to themselves or other vertices within the graph, as stated by Bondy

and Murty (1976). Furthermore, the graph vertices and edges can contain information. If the information is simply a name or number, the graph is called a *labelled graph*. Should additional information in the form of attributes be available, the graph is called an *attributed graph*. The edges can also be either directional or have no direction related to it. From the definition it is evident then why graph theory can be utilised in so many fields, notwithstanding FDD.

To show how one would go about constructing an attributed graph of the GTL process, the ATR unit will be used as an illustrative example. Fig. 4 (a) shows the ATR unit with its feed streams and syngas product stream. Firstly, each feed stream is represented by a node (nodes 1-4). These nodes are then connected to the ATR unit node (node 5) via directed edges, just as the process flow diagram depicts. The attributes of the nodes and edges are described by the energy characteristics calculated of the process. The completed attributed graph of the ATR unit is shown in Fig. 4 (b). An invaluable aspect of graph theory, called graph matching, is the determination of how similar one graph is to another. As summarised in Wilson and Martinez (1997), many different matching methods exist and the technique applied greatly depends on the type of graphs, their sizes, and the relevant information (symbolic, numeric,

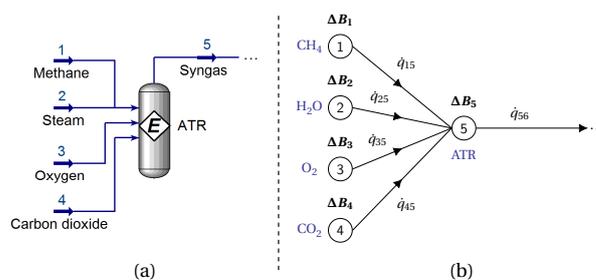


Fig. 4. (a) The ATR process unit (b) The constructed graph showing the nodes, edges and energy attributes of the ATR

etc.) being considered. For this study the Heterogeneous Euclidean Overlapping Metric (HEOM) proposed by Wilson and Martinez (1997), and reiterated by Jouli et al. (2009), was used. The technique works for both numerical and symbolic attributes, although the attributes in this case are numerical only. By utilising the HEOM instead of just calculating the Euclidean distance the following aspects are addressed:

- Should symbolic attributes be included in the future, the HEOM approach will be able to adequately handle the additional information.
- The Euclidean distance function does not include any normalisation, therefore, according to Wilson and Martinez (1997), attributes with large ranges would diminish smaller attributes' inputs.

5.2 Methodology

This section details the methodology of applying graph matching as a means to fault detection and isolation. To ensure a repeatable procedure, the following steps were determined and applied:

- (1) An attributed graph, as depicted in Fig. 5, was constructed based on the GTL process flow diagram, where the:
 - (a) *nodes* represent the process units
 - (b) *edges* convey the flow direction and connection of the units
- (2) The attributes of the:
 - (a) *nodes* are the changes in physical and chemical exergy (ΔB) over each process unit
 - (b) *edges* are the energy flows ($\dot{q}_{\nu\gamma}$) between connected process units ν and γ
- (3) Utilising the graph, a node signature matrix \mathbf{G} is constructed in the form given in (6); describing the change in physical exergy (ΔB_{ph}), chemical exergy (ΔB_{ch}) and the energy flow ($\dot{q}_{\nu\gamma}$) of each node. Should there be no energy flowing between two nodes, a 0 is added to that entry. Matrices were developed for each fault in datasets F_{pq_4} , F_{pq_R} , and F_{pq_1} .

$$\mathbf{G} = \begin{bmatrix} \Delta B_{ph_1} & \Delta B_{ch_1} & \dot{q}_{11} & \dots & \dot{q}_{118} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Delta B_{ph_{18}} & \Delta B_{ch_{18}} & \dot{q}_{181} & \dots & \dot{q}_{1818} \end{bmatrix} \quad (6)$$

- (4) Next a database was developed containing the graphs (\mathbf{G}_d) of every fault of dataset F_{pq_4} . No graph information (\mathbf{G}_o) pertaining to the operational faults

to be evaluated (F_{pq_R} and F_{pq_1}) are included in the database.

- (5) A cost matrix \mathbf{C}_{od} is used to determine how dissimilar two graphs, \mathbf{G}_o and \mathbf{G}_d , are when compared to one another. To calculate this

$$\mathbf{C}_{od} = \sqrt{\sum_{\alpha=1}^A \frac{|\mathbf{G}_{o_{i\alpha}} - \mathbf{G}_{d_{j\alpha}}|^2}{range_{\alpha}}}, \quad (7)$$

is used, giving an ($i \times j$) matrix. A is the number of columns of the graphs, j the number of rows in graph \mathbf{G}_d and i the number of rows in graph \mathbf{G}_o . To normalise the data, the $range_{\alpha}$ of each column of graph \mathbf{G}_o is obtained and calculated by:

$$range_{\alpha} = max_{\alpha} - min_{\alpha}, \quad (8)$$

where max_{α} is the largest numerical value and min_{α} the smallest in column α .

- (6) In order to determine a single distance (D_C) parameter between the two considered graphs,

$$D_C = \frac{\sum_{k=1}^i C_{kk}}{i}, \quad (9)$$

is utilised. The calculation basically comes down to summing the cost matrix's diagonal entries and dividing it by the number of rows, i , in the cost matrix.

- (7) The smaller the D_C -value the smaller the dissimilarities are between the compared graphs.

The detection and isolation will therefore work on the premise that given a known fault type of unknown magnitude, the above described method should match the operational condition to the corresponding fault - or more specifically - the graph within the database, by means of the smallest D_C -value.

6. RESULTS

The above-mentioned methodology was firstly applied to fault dataset F_{pq_R} . The D_C -values of each one of the eleven faults as compared to the database stored graphs were recorded and summarised in Table 3 (a). The smallest value, shown in bold, indicates the likeliest match. When evaluating the D_C -values it is seen that the proposed graph matching approach correctly matched all considered operational faults in dataset F_{pq_R} to their corresponding database faults. A confusion matrix was completed and is shown in Fig. 6 (a). The approach performed quite well as there were no false negatives (FN) or false positives

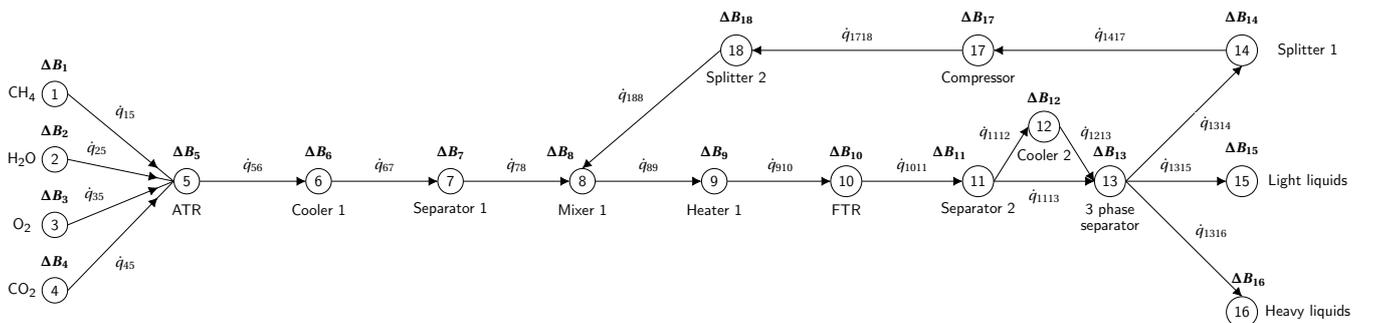


Fig. 5. The graph of the GTL process showing all of the nodes, edges and energy attributes

(FP), with both the true positives (TP) and accuracy being 100 %. However, when the approach was applied to dataset F_{pq_1} the performance drastically deteriorated. The D_C -values, shown in Table 3 (b), show poor matchings of the smaller magnitude faults. The confusion matrix for dataset F_{pq_1} is depicted in Fig. 6 (b). The fact that some faults graphs (F_{14_1} , F_{21_1} , F_{23_1} , F_{24_1} , and F_{31_1}) were matched to the normal graph seems to emphasise an issue

regarding the sensitivity of the proposed approach. The false negative rate (r_{FN}) of 45.5 % and accuracy of 54.5 %, clearly indicate the poor performance.

7. CONCLUSION

As the fixed-threshold approach, proposed in the work of Greyling et al. (2019), failed to detect all the considered 3 % faults, a different detection and isolation method

Table 3. Detectability and isolability of fault dataset (a) F_{pq_R} and (b) F_{pq_1}

(a)

FAULT ID		DATABASE STORED FAULTS													
		Normal	F_{11_4}	F_{12_4}	F_{13_4}	F_{14_4}	F_{21_4}	F_{22_4}	F_{23_4}	F_{24_4}	F_{31_4}	F_{32_4}	F_{33_4}		
F_{pq_R}	Normal ₁	✓	✓	0.00126	0.05044	0.06362	0.00362	0.00355	0.00260	0.09191	0.00169	0.01641	0.00142	0.05366	0.05086
	F_{11_6}	✓	✓	0.05018	0.00556	0.10589	0.04919	0.04929	0.04998	0.12818	0.05010	0.04997	0.05070	0.09310	0.08899
	F_{12_3}	✓	✓	0.05929	0.10922	0.01049	0.06119	0.06117	0.05901	0.07902	0.05885	0.06326	0.05881	0.05638	0.05315
	F_{13_7}	✓	✓	0.00726	0.04812	0.07038	0.00458	0.00458	0.00890	0.09882	0.00822	0.02095	0.00834	0.06064	0.05793
	F_{14_2}	✓	✓	0.00206	0.04839	0.06624	0.00059	0.00055	0.00405	0.09447	0.00310	0.01725	0.00332	0.05631	0.05355
	F_{21_3}	✓	✓	0.00211	0.04906	0.06420	0.00434	0.00436	0.00017	0.09232	0.00175	0.01659	0.00309	0.05434	0.05148
	F_{22_5}	✓	✓	0.11739	0.16831	0.08116	0.11989	0.11980	0.11688	0.01518	0.11671	0.11340	0.11684	0.06004	0.06787
	F_{23_6}	✓	✓	0.00146	0.04898	0.06414	0.00382	0.00384	0.00206	0.09228	0.00037	0.01571	0.00259	0.05416	0.05132
	F_{24_2}	✓	✓	0.00567	0.05032	0.06647	0.00668	0.00674	0.00767	0.09370	0.00667	0.00456	0.00661	0.05512	0.05257
	F_{31_7}	✓	✓	0.00311	0.04901	0.06532	0.00547	0.00551	0.00511	0.09350	0.00407	0.01204	0.00238	0.05524	0.05250
	F_{32_2}	✓	✓	0.04880	0.09490	0.05312	0.05110	0.05103	0.04844	0.05979	0.04810	0.04990	0.04827	0.01427	0.02527
	F_{33_3}	✓	✓	0.04776	0.09958	0.05292	0.04955	0.04939	0.04811	0.06359	0.04774	0.04595	0.04715	0.02314	0.02178

Yes = ✓ No = ✗

(b)

FAULT ID		DATABASE STORED FAULTS													
		Normal	F_{11_4}	F_{12_4}	F_{13_4}	F_{14_4}	F_{21_4}	F_{22_4}	F_{23_4}	F_{24_4}	F_{31_4}	F_{32_4}	F_{33_4}		
F_{pq_1}	F_{11_1}	✓	✗	0.01370	0.03447	0.07613	0.01284	0.01296	0.01385	0.10273	0.01386	0.02045	0.01438	0.06511	0.06207
	F_{12_1}	✓	✗	0.01406	0.06389	0.05221	0.01589	0.01589	0.01434	0.09127	0.01397	0.01833	0.01375	0.05321	0.05049
	F_{13_1}	✓	✓	0.00778	0.04812	0.07087	0.00508	0.00510	0.00939	0.09930	0.00872	0.01341	0.00886	0.06109	0.05837
	F_{14_1}	✗	✗	0.00074	0.04884	0.06515	0.00188	0.00186	0.00287	0.09338	0.00182	0.00977	0.00209	0.05523	0.05245
	F_{21_1}	✗	✗	0.00079	0.04932	0.06427	0.00316	0.00315	0.00150	0.09248	0.00107	0.00976	0.00185	0.05437	0.05156
	F_{22_1}	✓	✗	0.16198	0.18545	0.17694	0.16355	0.16353	0.16160	0.17489	0.16209	0.16144	0.16190	0.16488	0.16488
	F_{23_1}	✗	✗	0.00012	0.04928	0.06446	0.00268	0.00265	0.00223	0.09270	0.00111	0.00941	0.00143	0.05454	0.05177
	F_{24_1}	✗	✗	0.00206	0.04886	0.06506	0.00368	0.00378	0.00373	0.09293	0.00270	0.00842	0.00307	0.05450	0.05179
	F_{31_1}	✗	✗	0.00071	0.04961	0.06416	0.00316	0.00313	0.00234	0.09241	0.00131	0.00938	0.00096	0.05419	0.05139
	F_{32_1}	✓	✗	0.01701	0.06643	0.05301	0.01891	0.01877	0.01752	0.08161	0.01703	0.01907	0.01653	0.04127	0.04062
	F_{33_1}	✓	✗	0.01620	0.06601	0.05405	0.01802	0.01787	0.01673	0.08251	0.01629	0.01728	0.01570	0.04341	0.03985

Yes = ✓ No = ✗

CONFUSION MATRIX				DETECTION RATES		
		TRUE CONDITION		Rate	Formula	%
		Fault-free	Fault			
DETECTION CONDITION	Fault-free	True negative a	False negative b	r_{FP}	$= \frac{c}{(a+c)} \times 100$	0
	Fault	False positive c	True positive d	r_{FN}	$= \frac{b}{(b+d)} \times 100$	0
				r_{TP}	$= \frac{d}{(b+d)} \times 100$	100
				Accuracy	$= \frac{(a+d)}{(a+b+c+d)} \times 100$	100

(a)

CONFUSION MATRIX				DETECTION RATES		
		TRUE CONDITION		Rate	Formula	%
		Fault-free	Fault			
DETECTION CONDITION	Fault-free	True negative a	False negative b	r_{FP}	$= \frac{c}{(a+c)} \times 100$	0
	Fault	False positive c	True positive d	r_{FN}	$= \frac{b}{(b+d)} \times 100$	45.5
				r_{TP}	$= \frac{d}{(b+d)} \times 100$	54.5
				Accuracy	$= \frac{(a+d)}{(a+b+c+d)} \times 100$	54.5

(b)

Fig. 6. Confusion matrix and detection rates of the graph approach applied to dataset (a) F_{pq_R} and (b) F_{pq_1}

was required. By developing an attributed graph of the GTL process, information regarding the physical structure as well as stream composition and physical properties (described by exergy), are encapsulated. Graph theory then provides an array of methods in which to detect and isolate faults. The proposed approach presumed that the graphs of the 10 % faults were available as a database. The graphs of unknown operational faults are then compared to the database faults and their dissimilarities quantified by means of a distance parameter (D_C). The most likely fault being identified by the smallest D_C -value obtained. When evaluating the detection and isolation performance, all faults in dataset F_{pq_R} were successfully detected and isolated. Unfortunately, the performance declines when presented with the smaller faults as in dataset F_{pq_1} . This would suggest a similar issue with sensitivity such as the threshold approach displayed. One reason for this could be that the chosen D_C metric discards useful information contained within the cost matrix, seeing as the 18×18 matrix is reduced to a single distance parameter. Hence, the positive performance of the F_{pq_R} dataset warrants further investigation into the benefits more degrees of freedom would bring about. Additionally, a next phase of the research would assess the proposed approach's performance when completely unanticipated faults are considered.

ACKNOWLEDGEMENTS

This work is based on the research supported wholly/in part by the National Research Foundation of South Africa (Grant Number 127483). This work is based on the research supported by Sasol (Pty) Ltd. Opinions expressed and conclusions arrived at are those of the authors and are not necessarily to be attributed to Sasol.

REFERENCES

- Bondy, J. and Murty, U. (1976). *Graph Theory with Applications*. Elsevier Science Publishing Co., Inc.
- Chiang, L.H. and Braatz, R.D. (2003). Process monitoring using causal map and multivariate statistics: fault detection and identification. *Chemometrics and intelligent laboratory systems*, 65(2), 159–178.
- Chiang, L.H., Jiang, B., Zhu, X., Huang, D., and Braatz, R.D. (2015). Diagnosis of multiple and unknown faults using the causal map and multivariate statistics. *Journal of Process Control*, 28, 27–39.
- Choi, S.W., Lee, C., Lee, J.M., Park, J.H., and Lee, I.B. (2005). Fault detection and identification of nonlinear processes based on kernel PCA. *Chemometrics and intelligent laboratory systems*, 75(1), 55–67.
- De Klerk, A. (2011). *Fischer-Tropsch Refining*. John Wiley & Sons, first edition.
- Dong, Y. and Qin, S.J. (2018). A novel dynamic PCA algorithm for dynamic data modeling and process monitoring. *Journal of Process Control*, 67, 1–11.
- Fezai, R., Mansouri, M., Taouali, O., Harkat, M.F., and Bouguila, N. (2018). Online reduced kernel principal component analysis for process monitoring. *Journal of Process Control*, 61, 1–11.
- Ghosh, K., Ramteke, M., and Srinivasan, R. (2014). Optimal variable selection for effective statistical process monitoring. *Computers & Chemical Engineering*, 60, 260–276.
- Greyling, S., Marais, H., van Schoor, G., and Uren, K.R. (2019). Application of exergy-based fault detection in a gas-to-liquids process plant. *Entropy*, 21(6), 565.
- Jouili, S., Mili, I., and Tabbone, S. (2009). Attributed graph matching using local descriptions. In *International Conference on Advanced Concepts for Intelligent Vision Systems*, 89–99. Springer.
- Kletz, T. (1998). *What Went Wrong? Case Histories of Process Plant Disasters*. Elsevier Science, 4th edition.
- Knutsen, K.T. (2013). *Modelling and optimization of a Gas-to-Liquid plant*. Master's thesis, Institutt for kjemisk prosesssteknologi.
- Magnanelli, E., Berglihn, O.T., and Kjelstrup, S. (2018). Exergy-based performance indicators for industrial practice. *International Journal of Energy Research*, 42(13), 3989–4007.
- Maki, M., Jiang, J., and Hagino, K. (2004). A stability guaranteed active fault-tolerant control system against actuator failures. *International Journal of Robust and Nonlinear Control: IFAC-Affiliated Journal*, 14(12), 1061–1077.
- Marais, H., van Schoor, G., and Uren, K.R. (2019). The merits of exergy-based fault detection in petrochemical processes. *Journal of Process Control*, 74, 110–119.
- Ould-Bouamama, B., Biswas, G., Loureiro, R., and Merzouki, R. (2014). Graphical methods for diagnosis of dynamic systems: Review. *Annual Reviews in Control*, 38(2), 199–219.
- Panahi, M., Rafiee, A., Skogestad, S., and Hillestad, M. (2011). A natural gas to liquids process model for optimal operation. *Industrial & Engineering Chemistry Research*, 51(1), 425–433.
- Rafiee, A. and Hillestad, M. (2010). Optimal design and operation of a gas-to-liquid process. *Chemical Engineering Transactions*, 21, 1393–1398.
- Shah, M.D. (2011). Fault detection and diagnosis in nuclear power plant—a brief introduction. In *2011 Nirma University International Conference on Engineering*, 1–5. IEEE.
- Sorsa, T., Koivo, H.N., and Koivisto, H. (1991). Neural networks in process fault diagnosis. *IEEE Transactions on systems, man, and cybernetics*, 21(4), 815–825.
- Szargut, J. (2007). Egzergia. poradnik obliczenia i stosowania. *Wydawnictwo Politechniki Slaskiej*.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., and Kavuri, S.N. (2003). A review of process fault detection and diagnosis: Part I: Quantitative model-based methods. *Computers & chemical engineering*, 27(3), 293–311.
- Watanabe, K. and Hirota, S. (1991). Incipient diagnosis of multiple faults in chemical processes via hierarchical artificial neural network. In *Proceedings IECON'91: 1991 International Conference on Industrial Electronics, Control and Instrumentation*, 1500–1505. IEEE.
- Wilson, D.R. and Martinez, T.R. (1997). Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 6, 1–34.
- Xie, L., Lin, X., and Zeng, J. (2013). Shrinking principal component analysis for enhanced process monitoring and fault isolation. *Industrial & Engineering Chemistry Research*, 52(49), 17475–17486.