A Two-Step Framework for Predictive Maintenance of Cryogenic Pumps in Semiconductor Manufacturing

Sanjoy Kumar Saha[1], M.M. Manjurul Islam[1], Shaun McFadden[1], Saugat Bhattacharyya[1], Mark Gorman[2] and Girijesh Prasad[1]

[1]*Intelligent Systems Research Centre, Ulster University, Londonderry, BT48 7JL, UK*

saha-sk@ulster.ac.uk
g.prasad@ulster.ac.uk

[2]*Seagate Technology, Londonderry, Northern Ireland, BT48 0BF, UK*

mark.gorman@seagate.com

ABSTRACT

Semiconductor manufacturing involves many critical steps, wherein maintaining an ultra-high vacuum is mandatory. To this end, cryogenic pumps are used to create a controlled ultra-low-pressure environment through the use of cryogenic cooling. However, a sudden pump malfunction leads to contamination in the processing chamber, disrupting production. The primary focus of this study is preventing unplanned shutdowns of cryogenic pumps. The data was collected from various pump sensors also known as status variable identification (SVID) that reveals current behavior of the pump. A comprehensive framework is presented here to develop a condition monitoring and fault detection. In the proposed framework, a drift detection method is used for condition monitoring of the pump to locate gradual and abrupt drifts. Additionally, during regeneration (or maintenance) phase, intrinsic features are extracted to distinguish between normal and abnormal regeneration, achieving an accuracy of 90.91% and a precision of 66.67%. Utilizing the proposed system, cryo-pump operators can be given maintenance guidelines and warnings about potential health degradation of the pumps.

1. INTRODUCTION

Continuous operation of pumps ensures precise and high-quality production of semiconductors. A typical fabrication facility employs hundreds of pumps, from cryogenic to water pumps, within the clean room. Cryogenic pumps operate at fast pumping rates and very low pressure (10^{-10} torr). These pumps use helium as a coolant gas in two refrigeration stages. The first stage usually operates at 100 K(-173°C) followed by a second stage, operating roughly at 10 K (-263°C) to 12 K (265°C). In the first stage, water vapor is trapped, while in the second stage, gases such as nitrogen (N_2) and hydrogen (H_2) are cryogenically condensed from the process chamber (Collart et al., 2022a).

Within a process tool a group of four to six on-board cryopumps for each chamber are connected to a common compressor that services the connected cryopumps. Two types of operations—etching and deposition—are commonly performed within these chambers. The cryogenic pump requires a maintenance procedure known as regeneration that routinely removes the solid deposits from the cryogenically cooled stages and allows the pump to operate optimally. Additionally, a regeneration process typically consists of three stages: i) the warm-up stage, where the temperature is raised to approximately 320K, ii) the steady stage, during which gases are extracted using a roughing pump, and iii) the cooldown stage, where a cooling medium, such as liquid helium, is used to bring the system down to operational temperatures. Moreover, a regeneration can be classified into short or long regeneration based on time.

2. PROBLEM STATEMENT

In semiconductor manufacturing facilities, process tools are kept isolated from the surrounding environment. In such facilities, smart manufacturing can reduce risk and uncertainty by setting priorities and coordinating maintenance (Abdallah et al., 2023). Consider a typical scenario within a cryo-pump, where from captured gases escape, due to an unforeseen malfunction, leading to contamination in the process chamber. Additionally, wafer scrap could happen when implant dose mistakes and ion beam neutralization are brought on by elevated chamber pressure. This elevated pressure is a consequence of a malfunctioning cryo-pump. This situation leads to an unscheduled downtime that causes financial loss. The utilization of health monitoring and the conversion of unplanned incidents into planned ones can reduce or eliminate extra expenses.

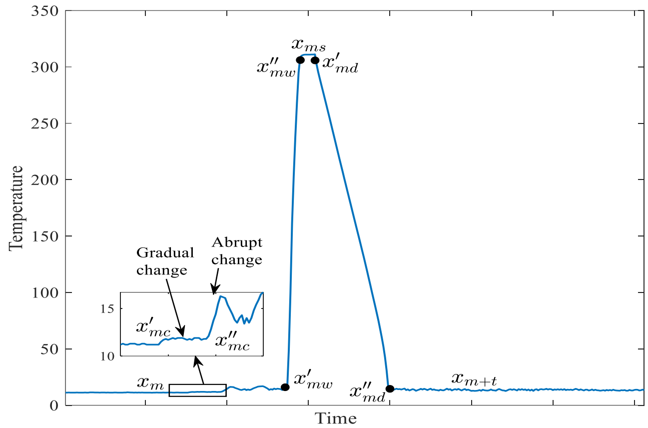

Figure 1. Dynamics of cryogenic pump's regeneration process; where, x'_{mw} to x''_{mw}, x'_{md} to x''_{md}, and x_{ms} represent warm-up, cooldown and steady stage respectively.

In the studied semiconductor fab, pump operators mainly monitor the second-stage temperature. Figure 1 illustrates the entire temperature monitoring scenario. When the temperature rises, operators observe the temperature (x'_{mc}) for some hours until it decreases. In the event of persistent high-temperature (x''_{mc}), regeneration process is started at point x'_{mw} as a part of maintenance. After regeneration, it is expected that the second stage temperature (x_{m+t}) will decline below 12 K. Subsequently, after regeneration, the operator waits and identifies if the second stage temperature is not attaining its temperature target, then they restart regeneration. After repeated regeneration, they declare that a fault occurred and replace the pump. In this scenario, semiconductor production is postponed due to the cleanliness protocols of the chamber.

The primary objective of the current research is to mitigate the sudden failure of the cryogenic pump and provide warning about maintenance. For this reason, the temperatures of both the first and second stage are crucial along with helium consumption, where mostly a fault marker is discerned. Moreover, the decision to start the regeneration process is taken by an operator.

The novel contributions of this research are outlined below:

i. Creation of a benchmark dataset, incorporating internal and add-on sensors, which will be made publicly available to the research community.

ii. Development of an online drift detection algorithm for multivariate status variable identification (SVID) for health monitoring.

iii. Performing feature engineering from regeneration and developing a novelty detection algorithm to identify abnormal regeneration patterns based on the extracted features.

iv. Enabling the operator to cease manual inspections by implementing automatic health monitoring of the cryogenic pump and automatic initiation of regeneration processes.

This study focuses on two-step solution for predictive maintenance. In the first stage, in-line data from pumps will be monitored and slow drift from a normal state will be identified. Then a change score will be calculated and if this score crosses a threshold a warning for regeneration will be sent to the pump operator. Here it is noteworthy that the drift will be normal when pumps accumulate a large volume of residuals and after regeneration pumps operate normally. In the second stage, regeneration cycle will be identified as normal or abnormal. Based on drift score and abnormal regeneration, an operator can make fast decisions about faultiness and they can swap the faulty pump with a new one.

3. PROPOSED FRAMEWORK

3.1. Health monitoring based on drift detection

The operational procedure of cryogenic pumps differs significantly from other types of pumps. Consequently, single point and multipoint anomaly detection suitable for other types of IoT data, may not be appropriate for cryopumps, as these anomalies may be considered normal phenomena (Blázquez-García et al., 2022; Rostami et al., 2021). Moreover, no single anomaly detection model performs well across all types of datasets. Therefore, for monitoring the state of cryogenic pump, continuous drift detection may be more appropriate. Drift is identified when there is a sudden or gradual change in data distribution (Ma et al., 2018). Abrupt change typically occurs due to sudden failure or change in equipment while gradual drift indicates slow degradation of equipment. Several methods have been proposed based-on dissimilarity score calculations of two distributions (Alippi et al., 2017; Yamada & Sugiyama, 2009). In this study, a two-stage drift detection framework is proposed. In the first stage, two distributions of data x_m and x_{m+n} are tested by hypothesis testing. Here x_m and x_{m+n} are data points at time m and $m+n$ separated by n number of samples and each distribution has k number of data points. When hypothesis test identifies any drift, second layer becomes active. In drift validation layer two distributions are tested again by directly estimating dissimilarity score of two probability density functions. This two-stage approach helps to reduce false positives (FP). To determine the threshold, a cumulative sum of all dissimilarity scores within a specific time range is calculated. When the cumulative sum exceeds the threshold, a maintenance warning is triggered. In the fab, this situation is handled by making several attempts to do regenerations.

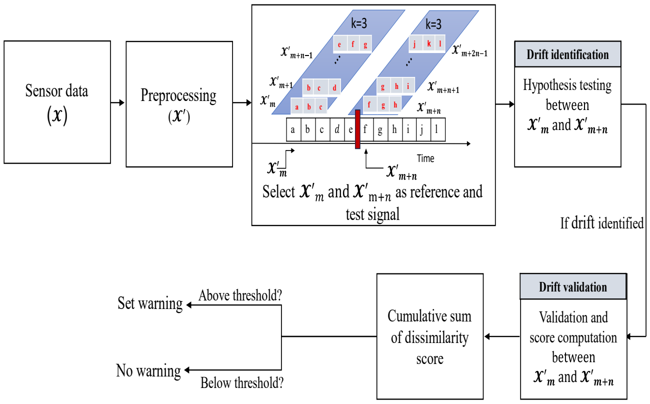

Figure 2. Block diagram of two-stage drift detection method (method 1); where $\mathcal{X}$ is the input IoT data from built-in sensor of cryogenic pump.

3.2. Abnormal regeneration identification

This stage focuses on abnormal regeneration identification based on extracted features from regenerations. Accurate identification of abnormal regeneration can save 12-15 hours in a real manufacturing hub, in making decisions about replacing an old pump. Potential problems with a cryogenic pump, compressor, and/or helium circulation circuit can arise in a cryogenic system. First or short regeneration is crucial for monitoring the condition of the cryogenic pump. However, the second stage temperature is the most critical metric monitored by an operator (Collart et al., 2022b). A divergent pattern among fast regenerations is a sign of abnormal pump behaviour. For instance, the N_2 purge problem is an indication of slower rate-of-rise. As a result, an appropriate method can effectively monitor and identify these abnormalities. Figure 3. illustrates how abnormal regeneration can be identified.

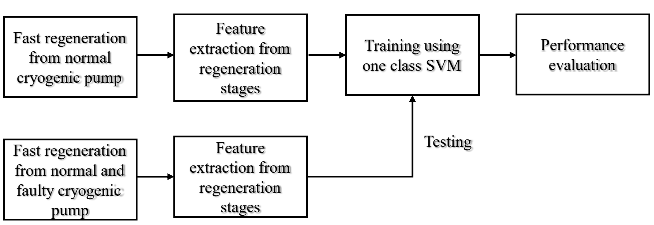

Figure 3. Block diagram of abnormal regeneration detection (method 2); where only short regenerations are utilized.

4. PRELIMINARY RESULTS AND DISCUSSION

4.1. Data Collection

Preliminary data for this study are collected from a real semiconductor manufacturing industry. Sensors inside the cryogenic pump record data at variable sampling rates. The data collection involved four different tools, named Tool-1, Tool-2, Tool-7, and Tool-8, each has four to six modules. Additionally, 2nd stage temperature, 1st stage temperature and helium consumption are studied here and data is collected from March 2023 to June 2023. In this article, units are omitted to maintain data confidentiality.

4.2. Drift Detection and Score Computation

Due to the inconsistent sampling rate among various SVIDs of sample data, some preprocessing is needed. Thus, the root mean square (RMS) is calculated per hour from SVID. Figure 4 illustrates average RMS value of 39 days. Afterwards, in module-4 (M04) a fault occurred. While module-4 has higher 2nd stage temperature, it remains within range although helium consumption is higher than other modules. Notably, the trend is removed from the data as a part of preprocessing.

Subsequently, to identify drift, the measure relative unconstrained least-square importance fitting (RuLSIF) (Liu et al., 2013) is used. The RuLSIF method computes dissimilarity score by direct computation of density ratio defined as:

$$\text{Density Ratio, } s(x) = \frac{\wp_{test}(x)}{(1-\alpha)\wp_{ref}(x)+\alpha\wp_{ref}(x)} \quad (1)$$

where $\wp_{test}(x)$ and $\wp_{ref}(x)$ are probability distribution densities for $x_{ref}(m)$ and $x_{ref}(m-n)$ with window size k;

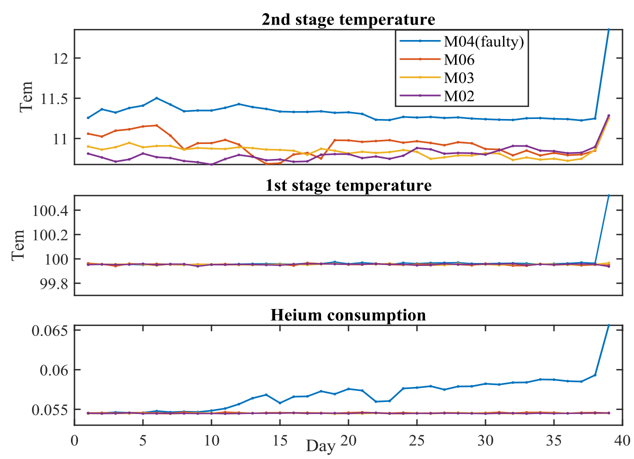

Figure 4. Average RMS values of Tool-7

here α is an adjustable parameter. In this research, $k = 12$, $n = 24$ and $\alpha = .1$ is used, meaning that change score is calculated every 12 hours and cumulative sum of score is used to identify the threshold. Figure 5 shows the cumulative change score for last 10 days before the fault occurred. Module 4 (M04) shows a higher change score in helium consumption compared to other modules. Consequently, it can be concluded that if change score of the helium consumption exceeds 40, a potential fault is likely and a warning for regeneration should be set. All analyses are performed on limited datasets with a small number of SVIDs. After obtaining the full dataset, the aim is to generalize this phenomenon for fab-wide predictive maintenance. However, in future, this study will focus on multidimensional extension of the proposed drift detection method and accounting for significant correlations among modules within the same tool.

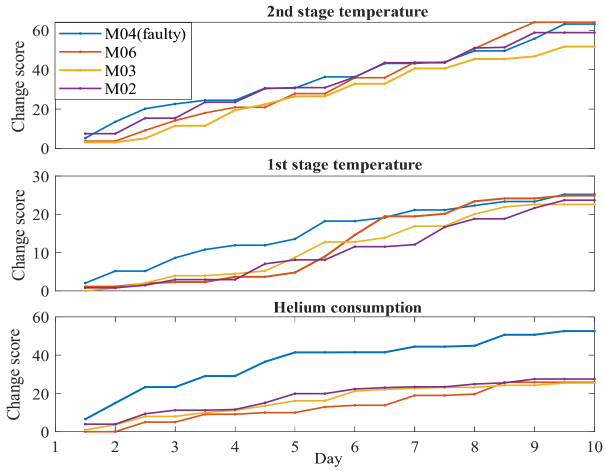

Figure 5. Cumulative change score of Tool-7

4.3. Feature Extraction from Regeneration

To analyse the time-varying behaviour of regeneration, rate of change in whole regeneration cycle is accomplished by estimating first derivative (Figure 6). The first derivative of point x_m with adaptive time difference Δm is obtained using backward difference as,

$$\Delta x_m = \frac{x_m - x_{m-1}}{\Delta m} \quad (2)$$

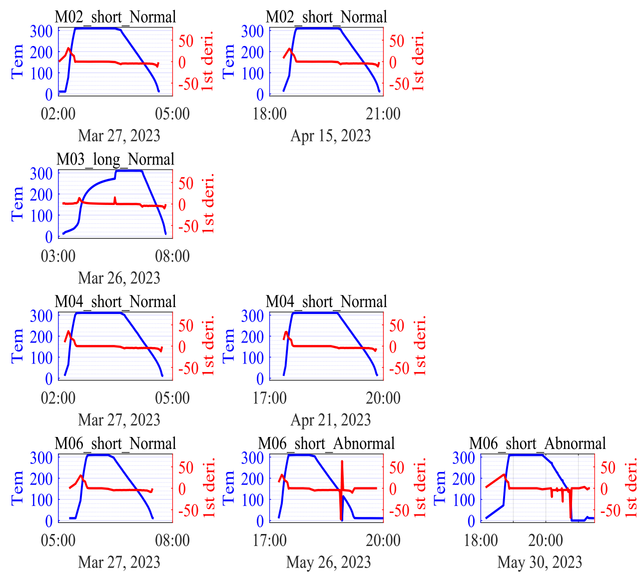

Figure 6. Rate of change for 2nd stage temperature from short and long regenerations of Tool-8; where y-axis(left) and y-axis(right) represents 2nd stage temperature and 1st derivative of temperature, respectively.

4.4. Abnormal Regeneration Identification

Based-on the proposed method 2, fifteen statistical features are calculated from each regeneration including: time, average change rate, standard deviation, maximum change rate, minimum change rate in warm-up, steady and cooldown stages. After that, these features from normal regenerations are used to train a one-class support vector machine (OCSVM). During testing, normal and abnormal regenerations are used. However, only fifty-four fast regeneration instances (33 for training and 21 for testing) from four different tools are used in this analysis. Moreover, this study only considered short regeneration. Again, proposed method's performance in terms of accuracy, specificity and precision are 90.91%, 94.74% and 66.67% respectively.

Table 1. Confusion matrix

		Predicted		Total
		Normal (0)	Abnormal (1)	
Actual	Normal (0)	18	1	19
	Abnormal (1)	1	2	3
	Total	19	3	21

5. CONCLUSION

The current market size of cryogenic pumps for semiconductor industry is around four billion dollar and is expected to increase annually by 11.8%. Thus, improved efficiency of these pumps has direct impact on the market. The first method monitors health of pumps and provides maintenance guidance. Additionally, the purpose of the second method is detection of abnormality during regenerations. Initial analysis shows promising results by utilizing sample data. In future, more heterogenous features from the frequency domain such as dominant frequency and mean peak frequency will be extracted and tested with the proposed method. Additionally, feature ranking and continual learning will be studied. Generally, distribution of equipment characteristics changes continuously due to aging under normal operating conditions and continual learning can adapt to this scenario. Moreover, in future, the proposed methods will be tested on larger datasets after completing data collection and labelling. Finally, improved methods will be implemented in a real semiconductor manufacturing plant.

ACKNOWLEDGEMENT

The funding support for this work from the UKRI Strength in Places Fund Project (81801): Smart Nano-Manufacturing Corridor is gratefully acknowledged.

REFERENCES

Abdallah, M., Joung, B.-G., Lee, W. J., Mousoulis, C., Raghunathan, N., Shakouri, A., Sutherland, J. W., & Bagchi, S. (2023). Anomaly Detection and Inter-Sensor Transfer

Learning on Smart Manufacturing Datasets. Sensors, 23(1), 486.

Alippi, C., Boracchi, G., & Roveri, M. (2017). Hierarchical Change-Detection Tests. IEEE Transactions on Neural Networks and Learning Systems, 28(2), 246–258.

Blázquez-García, A., Conde, A., Mori, U., & Lozano, J. A. (2022). A Review on Outlier/Anomaly Detection in Time Series Data. ACM Computing Surveys, 54(3), 1–33.

Collart, E., Longley, A., Gordon, D., Nordquist, J., & Matthews, P. (2022a). Predictive Maintenance Practices for Cryogenic Pumps in Semiconductor Manufacturing. 2022 33rd Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC), 1–6.

Liu, S., Yamada, M., Collier, N., & Sugiyama, M. (2013). Change-point detection in time-series data by relative density-ratio estimation. Neural Networks, 43, 72–83.

Ma, M., Zhang, S., Pei, D., Huang, X., & Dai, H. (2018). Robust and Rapid Adaption for Concept Drift in Software System Anomaly Detection. 2018 IEEE 29th International Symposium on Software Reliability Engineering (ISSRE), 13–24.

Rostami, H., Blue, J., Chen, A., & Yugma, C. (2021). Equipment deterioration modeling and cause diagnosis in semiconductor manufacturing. International Journal of Intelligent Systems, 36(6), 2618–2638.

Yamada, M., & Sugiyama, M. (2009). Direct Importance Estimation with Gaussian Mixture Models. IEICE Transactions on Information and Systems, E92-D(10), 2159–2162.