Acoustic Data-based Characterization of Incipient Boiling for Space Applications: A Progress Report

Krishnanshu Gupta
Data Science Department
Cal Poly, SLO
San Luis Obispo, CA
kgupta15@calpoly.edu

James Lamkin
Data Science Department
Cal Poly, SLO
San Luis Obispo, CA
lamkin@calpoly.edu

Andrew Martinez
Data Science Department
Cal Poly, SLO
San Luis Obispo, CA
amart531@calpoly.edu

Zachary Weinfeld
Data Science Department
Cal Poly, SLO
San Luis Obispo, CA
zweinfel@calpoly.edu

***Abstract*: In collaboration with NASA, this mid-project report details a machine learning framework designed to analyze and interpret acoustic data from cryogenic fuel tanks. Our framework detects and classifies early-stage boiling behaviors, vital for maintaining the safety and stability of fuel storage in microgravity environments. We utilize signal processing techniques, including peak detection and spectral analysis, to capture the distinct acoustic signatures associated with different boiling regimes. By applying unsupervised clustering models, we distinguish these regimes as quasi-homogeneous and heterogeneous nucleation.**

Preliminary results suggest that our approach can significantly improve the responsiveness of incipient boiling detection systems, in comparison to traditional thermal sensor systems used in spacecraft. This report details our methodologies and initial findings. Looking ahead, we plan on refining these methods to enhance early detection capabilities, ensuring more reliable cryogenic fuel management for space missions.

Keywords—Boiling Regimes, Acoustic signals, Unsupervised learning,

I. Introduction

In space exploration, the safe and reliable storage of cryogenic fuels – liquefied gases stored at very low temperatures – is crucial for powering spacecraft propulsion systems. Maintaining stability in these fuel tanks is essential, as any disruption can compromise mission safety. One of the key challenges in microgravity environments is the unpredictable behavior of boiling in cryogenic fuel tanks. Under these conditions, even minor heat leaks can create localized hot spots where the liquid fuel becomes superheated, exceeding its boiling point without immediately vaporizing. This high wall superheat can eventually trigger incipient boiling, leading to rapid vapor formation resulting in dangerous pressure spikes. Such unpredictable pressure fluctuations pose a significant risk of tank rupture and catastrophic mission failure.

NASA currently relies on thermal sensors to monitor boiling activity within cryogenic fuel tanks. However, these instruments provide limited spatial precision and cannot reliably detect the subtle, early signals of boiling. This limitation highlights the urgent need for a faster and more sensitive detection method.

Consequently, NASA researchers proposed an innovative solution to use acoustic-based detection methods that capture the distinct signals produced by bubble formation, growth, and collapse. By conducting experiments under simulated microgravity conditions, the researchers detected amplitude signals over specific time intervals following the onset of boiling.

Our project analyzes these acoustic signatures and uses machine learning models to characterize and distinguish among various boiling regimes. These regimes are broadly categorized into quasi-homogeneous and heterogeneous nucleation, each of which encompass more specific behaviors.

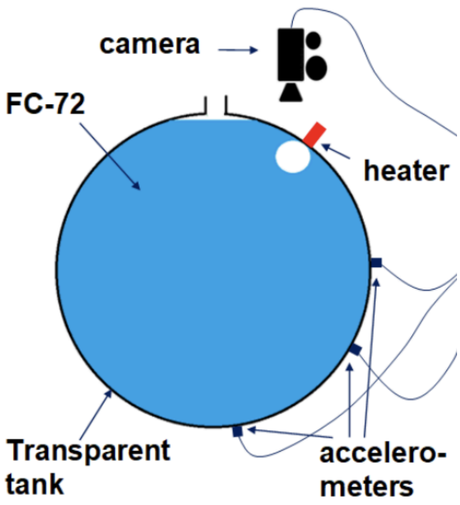
Ultimately, we aim to build a robust, real-time system that accurately classifies boiling patterns and provides early warning of potentially hazardous conditions. This approach could significantly enhance the safety and reliability of cryogenic fuel management systems, supporting NASA's long-term mission to enable safer, more sustainable space exploration.

II. Dataset Description

We were given data from 441 separate boiling experiments conducted by NASA, each lasting from 1 to 30 seconds.

A. Data Collection and Experimental Setup

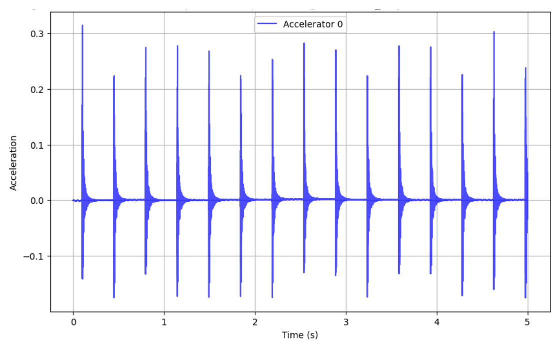
The experiments used a liquid-filled tank with acoustic sensors placed on its sides. These sensors captured sounds generated by bubble formation and growth, reflecting boiling patterns that ranged from random bubble formation at lower heat levels to more consistent boiling at higher levels. The figure below shows a diagram of this experimental setup.



B. Data Representation

The dataset comprises 441 individual comma-separated value (CSV) files, each representing a single experimental run. Each file contains accelerometer data sampled every 1/10,000th of a second from two sensors. Because the accelerometers produce similar results, we focused on data from a single sensor.

We visualize these data as plots in the time domain, with time (seconds) on the horizontal axis and accelerometer readings on the vertical. The figure below shows one example of this visualization, using data from one of the runs (Thursday, November 7, 2024, 12:54 pm Run 7).



C. Physical Interpretation

Both the time-domain and spectral representations emphasize the acoustic signals produced during bubble nucleation and collapse in a boiling liquid, especially under conditions simulating space environments. As superheated liquid reaches its boiling point near a heat source, bubbles form and grow rapidly before collapsing or detaching. This sudden formation and collapse of vapor bubbles releases energy in the form of energy waves, which manifest as sharp-spikes in the time domain signal. These spikes are linked to the behavior of boiling and bubble activity, with different boiling regimes, such as quasi-homogeneous or heterogeneous nucleation, producing distinct acoustic signals. In microgravity conditions, like those in space, convection is suppressed, leading to more intense superheating and potentially explosive boiling, which would be particularly pronounced in acoustic signals as distinct spikes.

III. Feature Engineering

Feature engineering plays a substantial role in our project, helping us uncover meaningful patterns from the raw accelerometer data.

The data can be analyzed from two different perspectives: time domain and spectral domain. These two separate domains provide distinct key insights, allowing us to extract features from each. By leveraging both, we enhance our ability to characterize boiling behavior and distinguish different experimental conditions.

A. Time Domain

The first set of extracted features relates to the time domain. When viewing the data in this domain, the most apparent attributes are the peaks. Since peaks correspond to nucleation events, extracting information about them proved to be one of the most effective feature extraction methods.

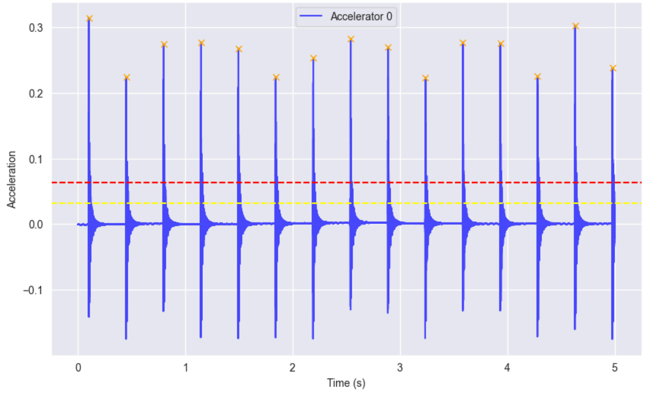
1) Peak Extracting Algorithm

We used SciPy's peak extraction algorithm to identify peaks in each run. The algorithm required two parameters: a vertical threshold and a horizontal threshold. The vertical threshold set a minimum height for detected peaks, while the horizontal threshold enforced a minimum separation between adjacent peaks. A fixed threshold across all runs proved ineffective due to significant variations in the scale of accelerometer data between experiments. To address this, we implemented dynamic thresholding.

a) Vertical Threshold: To determine the vertical threshold for each run, we computed both the 99.5th percentile of the data and 10% of the maximum value. We then took the maximum of these two values and clamped it between 0.015 and 0.1, ensuring the threshold didn't exceed these values and remained within a reasonable range.

b) Horizontal Threshold: Set the horizontal threshold to $350 + 5(vertical\ threshold)^{-1}$. Some runs only have one or two prominent peaks, causing the vertical threshold to be unusually low, creating false positive peaks. This problem is solved when lower vertical thresholds yield larger horizontal thresholds.

The figure below illustrates our peak finding algorithm applied to Thursday November 7, 2024, 12:54 pm Run 7. The red line represents the 99.5th percentile of the data, while the yellow line represents 10% of the maximum value in the data. The orange x's indicate the detected peaks.



2) Peak Based Features

Using our peak extraction algorithm, we derived several features from the x (time) and y (magnitude) values of each detected peak:

- Standard deviation of time between peaks
- Mean time difference between peaks
- Median time difference between peaks
- Maximum peak magnitude
- Median peak magnitude
- Standard deviation of peak magnitudes
- Average number of peaks per second
- Sum of peak magnitudes per second

These features effectively capture key aspects of nucleation behavior and boiling dynamics.

3) Other Time Domain Features

In addition to peak-based features, we extracted several other time-domain features directly from the raw data, independent of peak detections, including:

- Percent of time above threshold
- Standard deviation of signal

B. Frequency & Spectral Domain

Our second set of features comes from the spectral domain, which reveals patterns in frequency regularity. Intuitively, periodic nucleation events would produce distinct peaks in the power spectral density plot, while irregular or chaotic boiling regimes would exhibit a scattered frequency distribution without prominent peaks. Thus, transforming the accelerometer data to the spectral domain could prove useful to identify boiling regime patterns.

1) Frequency & Spectral Transformation

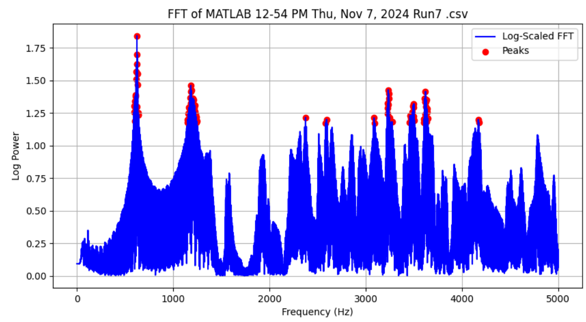
To convert the accelerometer data into the spectral domain, we applied a Fast Fourier Transform (FFT). Although we considered alternatives such as Wavelet Transform (WT) and Short-Time Fourier Transform (STFT), the largely stationary nature of the boiling regimes within each experiment made FFT more appropriate. We implemented FFT using SciPy for efficient spectral analysis.

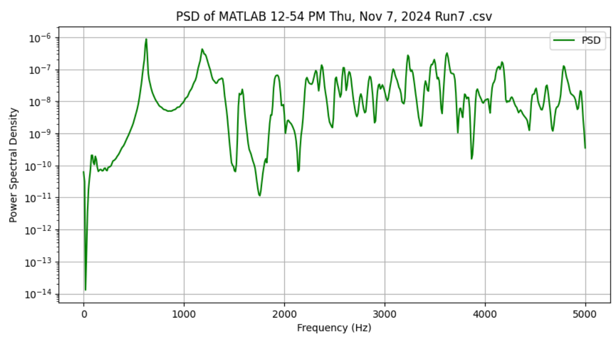
2) Frequency & Spectral Transformation Algorithm

To extract meaningful insights into the regularity of a boiling experiment, we implemented a series of methodologies to derive useful features. First, a high-pass Butterworth filter was applied to remove low-frequency noise. The filtered signal was then transformed into the frequency domain using a Fast Fourier Transform (FFT), decomposing it into its respective frequency components. The magnitude of the resulting FFT values was calculated to generate the power spectrum, highlighting the frequencies that contribute the most energy to the signal. The peaks of the power spectrum were identified as the “dominant” frequencies.

Next, Welch’s method was used to calculate the power spectral density, providing insight into how the signal’s energy is distributed across different frequency bands. The power spectral density also helps determine the signal’s entropy, where higher entropy indicates greater irregularity and lower entropy suggests a more stable signal.

See the figures below for the plotted spectral and frequency domains from **Run 7, recorded on Thursday, November 7, 2024, at 12:54 PM.**





3) Spectral Domain Features

Using our Frequency & Spectral Transformation Algorithm, we extracted a variety of different features.

- Top peak in the frequency domain
- Number of peaks in the frequency domain
- Weighted average frequency
- Mean power of power spectrum
- SD of power of power spectrum
- Spectral Entropy

These features aim to capture entropy within our signal, along with pointing out dominant frequencies to assess rhythmic trends.

IV. Progress To Date

A. Pipeline

To analyze the acoustic data, we implemented a multi-stage pipeline. This pipeline encompasses feature extraction, feature scaling, dimensionality reduction via Principal Component Analysis (PCA), and clustering using KMeans. The following subsections detail each step of this process.

1) Feature Extraction & Scaling

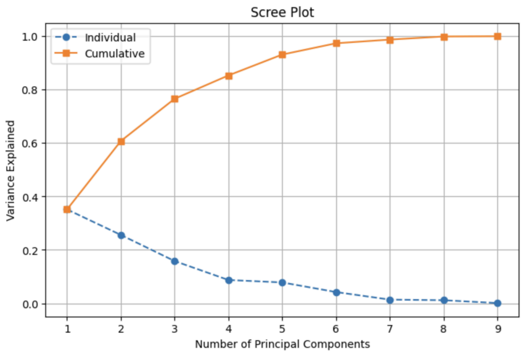
Using the raw acoustic data to identify peaks in each sample, we calculated and extracted the peak-based features mentioned in Section 3.2. After performing a Fast Fourier Transformation to convert our data from the time domain to the spectral domain, we extracted spectral domain features and compiled them into a CSV file alongside the other peak-based features. Most features were manually calculated using their mathematical definition, while others were derived using functions from SciPy's library.

To prepare the extracted features for dimensionality reduction, we employed sklearn's StandardScaler. Standardization was crucial because our dataset contained features with varying scales, which could significantly distort the results of PCA. StandardScaler standardized each feature by removing the mean and scaling it to unit variance, ensuring that all features contributed equally to the PCA. This preprocessing step was essential for accurately capturing the underlying variance in our data and preventing bias towards features with larger magnitudes.

2) Principal Component Analysis

To reduce the feature space for easier interpretability and insight into potential boiling regimes, we applied Principal Component Analysis. PCA is a statistical technique that transforms a set of potentially correlated variables into a set of linearly uncorrelated components, known as principal components. These components are ranked based on the amount of variance they explain, with the first component capturing the most variance.

After extracting features from the acoustic data, we performed PCA and analyzed a **Scree plot**, which details the individual and cumulative variance explained by each principal component. These six components collectively accounted for 98.75% of the total variance. Based on this, we proceeded with KMeans clustering using only these six principal components, effectively retaining essential information while significantly reducing data dimensionality.



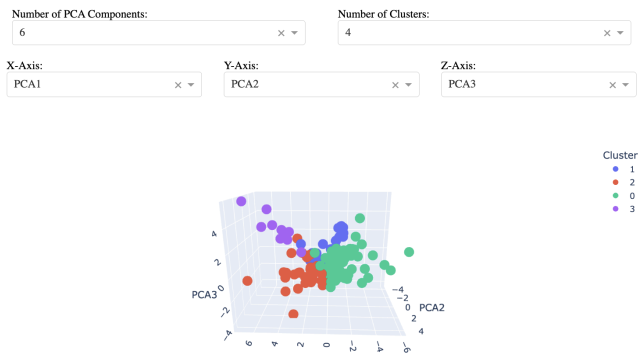
3) Unsupervised Learning - KMeans Clustering

Following dimensionality reduction with PCA, we applied KMeans clustering to the resulting six principal components. This unsupervised learning approach aimed to uncover distinct patterns within our acoustic data, effectively discerning different boiling regimes based on the key features highlighted by PCA. By clustering the data in this reduced space, we sought to uncover inherent groupings corresponding to various boiling behaviors, without relying on pre-existing labels.

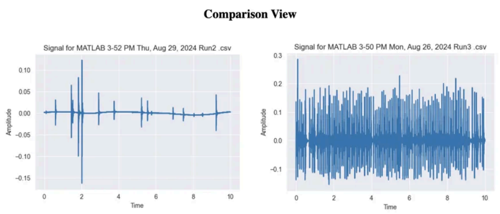
To facilitate further analysis and visualization, we performed KMeans clustering with varying numbers of clusters (ranging from 2 to 7) and evaluated projections across 2 to 6 principal components These clustering results were then integrated into our web application, providing an interactive platform for users to explore the data, visualize the clusters, and ultimately determine the optimal number of boiling regimes based on the observed patterns.

B. Web Application

To enhance our understanding of the clustering results derived from our pipeline, we developed an interactive web application designed to visualize the Principal Component Analysis (PCA) projections.



The platform features visualization controls, as displayed in the figure above, allowing NASA researchers to adjust the PCA dimensions and number of clusters displayed. Through interactive 2D and 3D scatterplots, researchers can clearly observe and analyze clustering patterns in the acoustic data, with each cluster representing a different boiling regime within quasi-homogeneous or heterogeneous boiling.



A key feature of the application is the comparison view, which enables our client to select specific data points directly from the scatterplots. Upon selection, the application displays detailed time-domain plots showing amplitude versus time for each data point. As shown in the figure above, this side-by-side comparison reveals distinct acoustic signatures between different clusters: one showing sparse, isolated peaks versus another with continuous high-frequency oscillations. This visual comparison helps NASA researchers identify and understand the characteristic acoustic patterns associated with different boiling regimes, making complex acoustic datasets more accessible and interpretable.

Overall, the application significantly simplifies the exploration of complex clustering and PCA component results, allowing NASA to derive more meaningful insights into the dynamics of different boiling regimes in microgravity conditions.

V. Future Plans

A. Exploring Periodicity For Multiple Boiling Regimes

A key limitation in our current analysis is the inability to capture multiple boiling regimes within a single experiment. While our existing feature set effectively characterizes experiment runs with a single nucleation site, it struggles to effectively differentiate between overlapping signals.

To address this, we plan to implement a Poisson mixture model to identify multiple boiling regimes within a single experiment. By exploring periodicity and the presence of multiple regimes, we aim to improve clustering accuracy and refine regime classification. For example, within a single boiling regime, boiling may follow either a rhythmic or irregular boiling pattern. However, in cases where multiple regimes overlap, the mixture could include a combination of rhythmic and irregular regimes, multiple irregular regimes, or multiple rhythmic ones. Rather than simply increasing the number of clusters, these new features will enable more precise and distinct boiling classifications, enhancing our model's ability to characterize complex boiling behaviors

B. Consolidating and Expanding the Pipeline

To enhance the efficiency and usability of our data processing pipeline, we are planning several key improvements. First, we aim to integrate the feature extraction stage directly into the pipeline, eliminating the current manual transfer of data from a separate features csv file. This integration will streamline the workflow and allow for quicker feature engineering iteration.

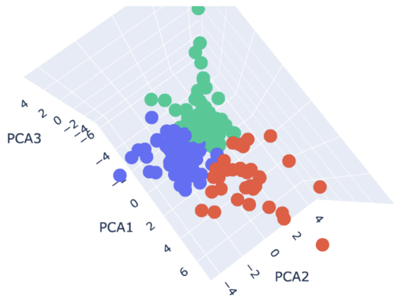
Additionally, we plan to automate the upload of KMeans clustering results to our web application. Currently, this process involves a manual transfer, which we intend to replace with an automated upload workflow. We will also implement a verbosity feature, enabling users to control the level of detail printed during pipeline execution through command-line arguments. This will allow for further analysis and understanding of the pipeline's behavior at various stages, ultimately improving its usability and interpretability.

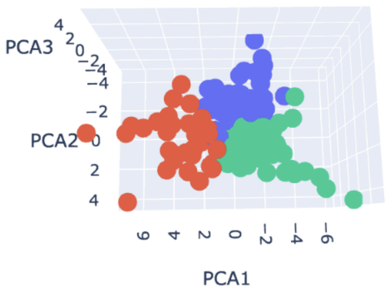
These enhancements will increase efficiency, reduce manual intervention, and improve the overall usability of our pipeline, ultimately making it a more powerful tool for NASA researchers analyzing acoustic boiling data.

Appendix A: Sample Clustering Result

A. Overview

This appendix presents a sample clustering outcome from our web application, using three principal components (PC1, PC2, PC3) and grouping the data into three clusters. The figure below visualizes each experimental run in 3D space, with colors indicating cluster assignments.





B. Principal Components

Our analysis identifies three principal components. PC1 represents timing vs. peak magnitude contrast, where positive values indicate longer intervals with smaller peaks, while negative values reflect shorter intervals with more intense peaks. PC2 captures overall signal intensity, with high values signifying a more energetic dataset—larger peaks, greater variability, and frequent bubble events—while low values correspond to a quieter boiling regime. PC3 represents peak frequency vs. individual peak intensity, where high values indicate many moderate peaks per second, and low values suggest fewer but stronger individual peaks.

C. Interpreting Scatterplot

The X-axis (PC1) represents the contrast between timing and peak magnitude, moving from shorter, more intense peaks on the negative side to longer intervals with milder peaks on the positive side. The Y-axis (PC2) reflects overall signal activity, with lower values indicating a quieter regime and higher values corresponding to a more energetic dataset. The Z-axis (PC3) captures peak frequency versus intensity, shifting from fewer, stronger peaks at the bottom to a higher frequency of moderate peaks at the top.

Cluster 0, shown in green, appears in regions with moderately negative PC1 and positive PC3. This suggests shorter intervals between peaks combined with frequent, moderate peaks, likely representing an energetic but relatively uniform boiling regime. Cluster 1, in blue, is centered around mid-range PC1 and PC3 values but skews toward lower PC2. This indicates a generally weak signal with larger peak heights and greater variability, though not necessarily the highest frequency of peaks. Cluster 2, in red, leans toward positive PC1 and lower or middle PC3, reflecting longer time gaps between peaks and fewer moderate peaks overall. This cluster likely corresponds to a more subdued boiling regime, where events are less frequent but of moderate intensity.

References

[1] C. Dunlap, H. Pandey, E. Weems, H. Hu. Nonintrusive heat flux quantification using acoustic emissions during pool boiling. Applied Thermal Engineering 228 (2023) 120558.

[2] S. Barathula, K. Srinivasan. Review on research progress in boiling acoustics. International Communications in Heat and Mass Transfer 139 (2022) 106465.

[3] J. Tang, G. Xie, J. Bao, Z. Mo, H. Liu, M. Du. Experimental study of sound emission in subcooled pool boiling on a small heating surface. Chemical Engineering Science 188 (2018) 179-191.

[4] B. M. Dorofeev and V. I. Volkova. An acoustic method of investigation of the process of boiling. High Temp. 43 (4), 620–627, 2005.

[5] B. M. Dorofeev and V. I. Volkova. The effect of evaporation and condensation in vapor bubbles on the hydrodynamic sound generation in a subcooled boiling liquid. Acoust. Phys. 52 (2), 173–179, 2006.

[6] Y. Ueki and K. Ara. Proof of concept of acoustic detection of boiling inception and state transition using deep neural network. International Communications in Heat and Mass Transfer 129 (2021) 105675.

[7] V. K. Dhir. Boiling under microgravity conditions. In: Proceedings of the 12th Int. Heat Transfer Conf., Grenoble, France (2002).

[8] C. B. Muratov, V. V. Osipov, V. N. Smelyanskiy. Issues of Long-Term Cryogenic Propellant Storage in Microgravity, NASA Report, 2011,TM-2011-215988, ARC-E-DAA-TN4295

[9] M. Hasan and R. Balasubramaniam. Analysis of the Pressure Rise in a Partially Filled Liquid Tank in Microgravity with Low Wall Heat Flux and Simultaneous Boiling and Condensation, 50th AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, 2012

[10] H. Merte and H. S. Lee. Quasi-Homogeneous Nucleation in Micro-gravity at Low Heat Flux: Experiments and Theory, ASME Journal of Heat Transfer, Vol. 119, pp. 305–312, 1997.

[11] M. Kassemi, O. Kartuzova and S. Hylton. Validation of two-phase CFD models for propellant tank self-pressurization: Crossing fluid types, scales, and gravity levels. Cryogenics, Vol. 89, No. 2018, pp. 1–15, 2018.

[12] https://www1.grc.nasa.gov/space/iss-research/msg/zbot/zbot-1/#implication-of-zbot-findings-for-propellant-tank-design

[13] W. L. Johnson, R. Balasubramaniam and R. Grotenrath. Analysis of heat transfer from a local heat source at cryogenic temperatures. IOP Conf. Ser.: Mater. Sci. Eng. 1240 012013, 2022.

[14] Z. Benkő, T. Bábel, and Z. Somogyvári. Model-free detection of unique events in time series. Scientific Reports, 12(1):227, 2022.

[15] H.C. Hansen, W.L. Johnson, M. Meyer, A. Werkheiser, J.R. Stephens, Cryogenic fluid management technologies enabling for the Artemis program and beyond, in: ASCEND 2020, p. 4000, 2020.