Hybrid Energy-Based Models and Flow Matching for Sensor Anomaly Detection

Literature Review & Project Proposal

(FIT5126)

Shosuke Asano
Student
sasa0005@student.monash.edu

Dr Loo Junn Yong
Supervisor
loo.junnyong@monash.edu

Dr Ting Chee Ming
Co-Supervisor
ting.cheeming@monash.edu

5880 words

Submission Date: 4th May 2025

Table of Contents

1. Introduction

Sensor signals are indispensable components in countless modern systems, providing the data essential for monitoring, control, and automation across industries ranging from manufacturing to healthcare (Zamanzadeh Darban et al., 2024). However, the reliability of these systems hinges on the integrity of sensor data, which can be compromised by various faults or anomalies such as bias, drift, spikes, or stuck values (Li et al., 2020; Yi et al., 2017). These anomalies, arising from sensor degradation, environmental factors, or system malfunctions, can lead to significant operational inefficiencies, safety hazards, and economic losses. Detecting these anomalies is complicated by the nature of contemporary sensor data, often characterized by high dimensionality, noise, complex temporal dependencies, and non-stationarity (Shiva et al., 2024). Consequently, robust anomaly detection methods are critical. While traditional statistical process control techniques have been used (Montgomery, 2024), they often fall short when faced with the complexities and dynamics of modern sensor signals, sometimes requiring unrealistic assumptions about data distributions. The rise of machine learning and deep learning has provided powerful alternatives, particularly unsupervised generative models, which excel at learning complex patterns from unlabeled data (Hoh et al., 2022). These models learn the distribution of normal operational data and identify anomalies as deviations, circumventing the common challenge of scarce labeled anomaly data in real-world settings (Geiger et al., 2020). Key generative approaches include Energy-Based Models (EBMs), Diffusion Models, and Flow Matching.

Given this context, the primary aim of the literature review within this report is to comprehensively survey and synthesize foundational concepts alongside the state-of-the-art in generative modeling techniques as applied to sensor anomaly detection. This involves examining the underlying principles, strengths, and limitations of dominant and emerging methods, assessing their suitability for handling typical sensor data characteristics and fault types, and identifying critical knowledge gaps in current research. To fulfill this aim, the Literature Review (Section 2) first establishes essential background on common sensor faults and signal transformation techniques like STFT (Oppenheim & Lim, 1981) and DST (Dewhurst et al., 2020) crucial for feature extraction. It then critically examines prominent generative models:

- **Energy-Based Models (EBMs):** Valued for their potential interpretability via energy scores but often hampered by intractable partition functions and challenging MCMC-based training (Du & Mordatch, 2019; Nijkamp et al., 2019; Yoon et al., 2023).
- **Diffusion Models:** Known for strong generative performance but typically rely on computationally intensive, iterative reconstruction processes for anomaly detection, often lacking transparency (Liu et al., 2025; Pintilie et al., 2023).
- **Flow Matching:** A more recent framework offering efficient and stable training for continuous normalizing flows, potentially faster than diffusion methods (Lipman et al., 2023; Patel et al., 2024). The review also explores nascent Hybrid Approaches that propose combining EBM principles with Flow Matching training methodologies to potentially achieve both interpretability and efficiency (Balcerak et al., 2025; Loo et al., 2025).

Synthesizing these findings (Section 3), the literature review reveals a significant research gap: despite the theoretical appeal of hybrid EBM/Flow Matching models, there is a lack of empirical studies validating their performance and practical applicability specifically for detecting diverse anomaly types within sensor signal data. Addressing this gap is the core motivation for the research project proposed herein. This project focuses on the development, implementation, and rigorous evaluation of a hybrid EBM/Flow Matching model, employing techniques like Variational Potential

Flow Bayes (VPFB), explicitly tailored for unsupervised sensor anomaly detection. The importance of this work lies in its potential to yield a solution that synergizes the interpretability advantages of EBMs with the computational efficiency and stability of Flow Matching. Successfully demonstrating such a model could lead to more robust, scalable, and trustworthy anomaly detection systems, particularly valuable in domains where understanding the reasoning behind an anomaly flag is crucial (Rudin, 2019). Providing empirical evidence for this approach on benchmark sensor data (MFP dataset, Ruiz-Cárcel et al., 2015) with realistic fault scenarios moves beyond theoretical promise to practical application.

The remainder of this report details this endeavour. Section 2 presents the full literature review, Section 3 summarizes the state-of-the-art and the identified gap, Section 4 outlines the proposed Research Project Plan, and Section 5 offers concluding remarks.

2. Literature Review

This section provides a comprehensive review of approaches to sensor anomaly detection. Through this review, we aim to bridge foundational understanding with cutting-edge approaches, highlighting the evolution from basic principles to novel methodologies.

2.1. Background

This section provides foundational context for sensor anomaly detection by outlining common types of sensor faults, signal transformation techniques, and general approaches to anomaly detection. Together, these components establish a basis for understanding the challenges and motivations behind recent advances in the field.

2.1.1. Common Sensor Fault Types in Anomaly Detection

Anomalies in sensor data are categorized into five common types of sensor faults—bias, drift, erratic, spike, and stuck—as identified in prior studies (Li et al., 2020; Yi et al., 2017). As illustrated in Figure 1, each fault type exhibits distinct characteristics. Bias faults introduce a constant offset to the sensor readings, while drift faults involve a gradual deviation over time. Erratic faults result in unpredictable, noisy fluctuations. Spike faults manifest as sudden, short-lived deviations, and stuck faults occur when the sensor output remains fixed at a constant value. These faults persist throughout the entire input sequence, with each component of the input equally likely to be affected.

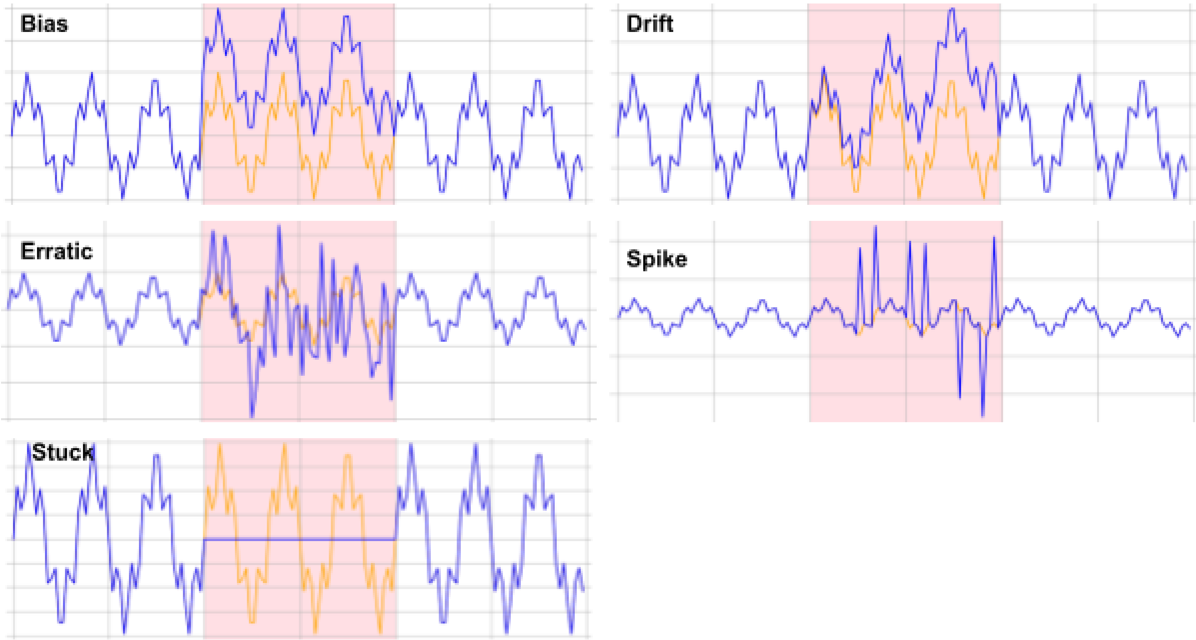

Figure 1: Examples of sensor faults (blue line) compared to the correct signal (orange line).

2.1.2. Signal Transformation Techniques

Sensor signals often exhibit complex, non-stationary temporal dynamics that are not easily captured in their raw time-domain form. To better extract meaningful features, signal transformation techniques are commonly applied prior to modeling (Zhao et al., 2018).

One widely used method is the Short-Time Fourier Transform (STFT), which converts the signal into a time-frequency representation (Oppenheim & Lim, 1981) as illustrated in Figure 2. STFT divides the signal into overlapping segments and applies the Fourier transform to each window, enabling the extraction of localized frequency components over time. This representation allows models to detect transient events, periodic structures, and evolving spectral patterns, all of which may indicate anomalies. Recent studies have shown the effectiveness of STFT-based features in generative models for time series prediction, especially in contexts with strong temporal dependencies and non-stationary behavior (Naiman et al., 2024). By expanding the feature space, STFT enhances a model's ability to learn complex patterns and improves its capacity to detect subtle anomalies.

Complementing frequency-based approaches, the Discrete Shocklet Transform (DST) provides a shape-based method for signal transformation (Dewhurst et al., 2020). Unlike STFT, DST does not rely on frequency decomposition but instead uses cross-correlation with predefined "shock-like" kernels to identify localized patterns in the time domain. This approach is timescale-independent and focuses on detecting abrupt changes and transient behaviors that do not exhibit clear periodicity. DST is particularly valuable in applications where anomalies manifest as sudden, irregular shapes rather than frequency shifts, offering a mechanism-driven, interpretable view of the data.

Together, STFT and DST provide complementary perspectives on signal dynamics—STFT capturing oscillatory and spectral structures, and DST highlighting shape-based, non-oscillatory patterns—both enriching the input representation for downstream anomaly detection models.

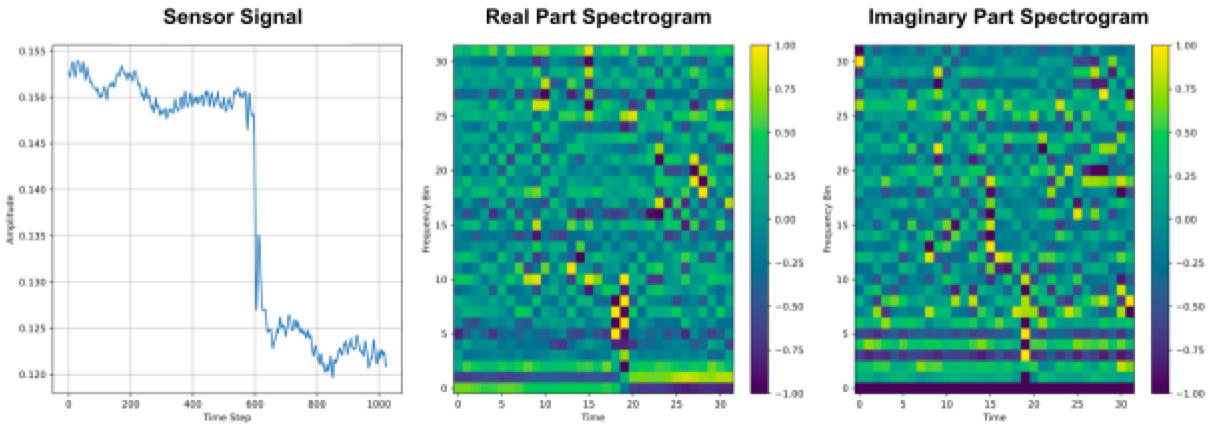

Figure 2: Real and imaginary components of the spectrograms generated by applying the Short-Time Fourier Transform (STFT) to the corresponding signal.

2.1.3. General Approaches to Anomaly Detection

Anomaly detection in sensor data is a well-established field with a rich history spanning several decades. It plays a critical role across various domains, including fraud detection, cybersecurity, network monitoring, industrial systems, and healthcare (Zamanzadeh Darban et al., 2024). Traditional approaches to outlier detection have often relied on statistical methods, which typically assume specific underlying data distributions and methods that depend on predefined thresholds (Montgomery, 2024). However, contemporary sensor data presents significant challenges to these traditional techniques due to its complex characteristics, including high dimensionality, inherent noise, temporal dependencies, and the phenomenon of concept drift, where the statistical properties of the data evolve over time (Shiva et al., 2024).

To overcome the limitations of traditional methods, recent research has increasingly focused on machine learning and, more notably, deep learning-based approaches for anomaly detection (Zamanzadeh Darban et al., 2024). These methods excel at capturing complex, nonlinear patterns in data and do not require strong assumptions about the underlying distribution. Deep learning models, in particular, are well-suited for handling high-dimensional data such as multivariate time series (Zamanzadeh Darban et al., 2024), making them highly effective in modern sensor-based applications. However, a common drawback of deep neural networks is their lack of interpretability, which can be a major limitation in domains such as finance and healthcare, where transparency and explainability are critical (Montgomery, 2024; Rudin, 2019).

2.1.4. Generative Models for Anomaly Detection

Among modern techniques, generative models have emerged as a particularly promising direction for anomaly detection in time series data (Hoh et al., 2022). These models learn the underlying distribution of normal patterns and detect anomalies by identifying data points with low likelihood or high reconstruction error under the learned distribution. This paradigm enables more robust detection of subtle and complex anomalies compared to traditional rule-based systems. In addition to their effectiveness, generative models offer several practical advantages, especially in real-world, unsupervised settings (Geiger et al., 2020):

- **No need for labeled anomalies:** Generative models operate in an unsupervised manner, learning directly from raw time series data without requiring prior knowledge of known anomalies.

- **Independence from simulation-based "normal baselines":** They do not rely on physics-based models or predefined normal signals, making them suitable for systems where generating or defining a normal baseline is difficult.
- **Adaptability to real-world noise and external variability:** Some generative models, such as flow matching (Lipman et al., 2023), exhibit robustness to data containing unexpected shifts or patterns caused by external phenomena (e.g., environmental changes or control regime shifts).
- **Applicability to unsegmented and variable-length data:** Generative approaches are well-suited for signals that cannot be easily segmented or have variable durations, which is common in many industrial and sensor-based applications.

By combining strong modeling capacity with adaptability and minimal supervision requirements, generative models stand out as a scalable solution for anomaly detection in complex, real-world environments.

2.2. Related Works

Recent generative modeling advances have opened new avenues for unsupervised sensor anomaly detection, particularly in settings with scarce labeled data. While originally developed for tasks like image synthesis or density estimation, models such as Energy-Based Models (EBMs), Diffusion Models, and Flow Matching have shown strong potential for capturing complex data distributions and identifying anomalies. This section reviews these approaches with a focus on their applicability to anomaly detection and highlights emerging hybrid models that aim to combine the interpretability of EBMs with the efficiency of Flow Matching.

2.2.1. Energy-Based Models (EBMs) for Anomaly Detection

Energy-Based Models (EBMs) offer a powerful and flexible framework for modeling complex, high-dimensional data distributions. At their core, EBMs define an energy function E(x), typically parameterized by a neural network, which assigns low energy values to in-distribution (normal) data points and higher values to out-of-distribution (anomalous) ones (Nijkamp et al., 2019; Yoon et al., 2023). Since anomalies often lie in low-density regions of the data manifold, they are effectively characterized by high energy scores (Du & Mordatch, 2019).

In the context of anomaly detection, the key inference task involves computing the energy E(x) for a given test input. A sample is flagged as anomalous if its energy exceeds a predefined threshold (Yoon et al., 2023). While finding configurations that minimize the energy function is a related inference task within EBMs (used in training or generation), direct energy evaluation is the standard approach for anomaly scoring. Probabilistically, EBMs define a Boltzmann distribution over data x as:

$$P(x) = \frac{e^{-E(x)}}{Z}$$

Here, E(x) represents the energy assigned to data point x, and $Z = \int e^{-E(x)} dx$ is the partition function. It is an intractable normalization constant for high-dimensional data. Obtaining samples directly from this distribution is also generally intractable. This intractability poses the primary challenge for maximum likelihood training of EBMs (Xiao et al., 2021). To address this, sampling methods, particularly Markov Chain Monte Carlo (MCMC) techniques, are frequently employed to approximate the required expectations. Consequently, several approximate training methods have been developed to circumvent this challenge:

- Contrastive Divergence (CD): Approximates the log-likelihood gradient using a small number of sampling steps (e.g., Langevin dynamics or Gibbs sampling) starting from observed data (Du et al., 2020).
- Score Matching (SM): Avoids the partition function altogether by minimizing the difference between the score functions (gradients of log-probability) of the model and the empirical data distribution (Li et al., 2023).
- Noise Contrastive Estimation (NCE): Reframes density estimation as a binary classification task, distinguishing real data from noise samples drawn from a known distribution (Ma & Collins, 2018).

As discussed in Section 2.1.3, model interpretability is a critical consideration, especially in high-stakes domains like finance and healthcare. In this regard, EBMs offer a degree of transparency by directly modeling energy landscapes that are semantically meaningful (Carbone, 2024). Furthermore, their unsupervised training framework aligns naturally with real-world anomaly detection scenarios, where labeled anomalies are rare or unavailable.

While the training of EBMs remains a significant challenge due to the intractability of the partition function and the complexity of sampling-based methods, which often involve computationally expensive MCMC procedures with potential convergence issues, ongoing advances in approximate learning techniques continue to enhance their viability for practical use. Given their expressive power, interpretability, and compatibility with unsupervised learning, EBMs represent one of the compelling approaches for robust, scalable anomaly detection.

2.2.2. Diffusion Models for Anomaly Detection

Diffusion Models are a powerful class of generative models that have achieved remarkable performance across various domains, including image synthesis, speech generation, and time-series modeling (LingYang et al., 2023). Their core mechanism involves a two-step process: a forward diffusion process, where noise is gradually added to the training data over several time steps, and a reverse diffusion process, where the model learns to denoise this noisy data to recover the original input distribution. The reverse process is typically parameterized by a neural network (often a U-Net or a Transformer), which is trained to predict and remove the noise introduced during the forward process at each time step.

In the field of anomaly detection, diffusion models are primarily used in reconstruction-based frameworks (Pintilie et al., 2023). These approaches involve training the diffusion model exclusively on normal data. During inference, the model attempts to reconstruct potentially anomalous test inputs by running the learned reverse diffusion process. If the input is anomalous, the model typically struggles to reconstruct it accurately. The reconstruction error, measured in terms of the difference between the input and the generated output, serves as the anomaly score (Pintilie et al., 2023). Larger errors indicate a higher likelihood of anomaly. Beyond reconstruction error, recent studies have also investigated alternative anomaly scoring methods. Some utilize the learned latent representations or directly leverage the score function produced during training as an indicator of deviation from the learned distribution (Sakai & Hasegawa, 2025). These methods aim to exploit deeper internal signals from the generative process for more sensitive or interpretable detection.

Despite their strong modeling capacity, diffusion models face notable practical challenges, particularly in terms of computational cost. Because the reverse denoising process involves hundreds to thousands of iterative steps, computing an anomaly score at inference time can be time-consuming, especially when applied to long sequences or high-dimensional data (Liu et al., 2025; Yang et al., 2023). This makes diffusion-based anomaly detection less suitable for real-time or resource-constrained scenarios.

Compared to Energy-Based Models (EBMs) discussed in Section 2.2.1, diffusion models tend to offer greater training stability and end-to-end differentiability, avoiding the need for MCMC-based inference techniques that are often sensitive to hyperparameter tuning and can be slow to converge (Carbone, 2024). However, like many deep generative models, diffusion models also lack interpretability (Katsuoka et al., 2024; Liu et al., 2025), making it difficult to provide clear explanations for why certain inputs are classified as anomalous. This lack of transparency can hinder their adoption in high-stakes, real-world settings where trust and explainability are critical.

2.2.3. Flow Matching for Anomaly Detection

Flow Matching is a recent and efficient generative modeling framework designed to train Continuous Normalizing Flows (CNFs) without relying on simulation-based solvers (Lipman et al., 2023). CNFs transform a simple base distribution into complex data distributions through a continuous-time flow defined by a time-dependent vector field. Traditional CNF training requires solving differential equations, which can be slow and unstable. Flow Matching simplifies this by directly regressing a neural network onto a target vector field, which is defined by interpolated probability paths between noise and data samples.

A major strength of Flow Matching lies in its training efficiency and stability. Unlike diffusion models, which rely on second-order partial differential equations (PDEs) and require hundreds of denoising steps, Flow Matching trains using first-order PDEs, resulting in faster convergence and more robust learning (Patel et al., 2024). In terms of inference speed, Flow Matching often produces straighter and more direct sampling paths, which reduce the number of steps needed to evaluate or generate data samples compared to diffusion-based models (Lipman et al., 2023).

Furthermore, its flexibility allows it to integrate with techniques from diffusion models and path-based generative methods, offering a unified and generalizable framework for generative modeling (Lipman et al., 2023). It can also be used to train EBMs, offering better training stability and efficiency compared to MCMC-based methods (Chao et al., 2023). This makes Flow Matching a promising and adaptable tool in the growing landscape of unsupervised anomaly detection, especially when balancing modeling power, interpretability, and computational efficiency.

2.2.4. Hybrid Approaches (EBMs and Flow Matching)

As previously discussed in Section 2.2.1, EBMs offer a highly interpretable framework for unsupervised anomaly detection by assigning scalar energy values to data points—low energy to in-distribution data and high energy to anomalies. Despite their theoretical appeal and compatibility with label-scarce scenarios, EBMs suffer from practical limitations, most notably the intractability of the partition function and the reliance on slow, often unstable MCMC methods for sampling and training (Du & Mordatch, 2019).

In contrast, Flow Matching (Section 2.2.3) has emerged as a training-efficient, simulation-free alternative for modeling data distributions through CNFs. By learning a deterministic vector field between base and target distributions via regression, rather than solving differential equations or relying on iterative denoising, it supports stable, first-order PDE-based training and faster inference (Patel et al., 2024). Moreover, Flow Matching provides a versatile generative framework that aligns well with both diffusion models and path-based generative techniques, offering a bridge between model expressiveness and computational feasibility (Lipman et al., 2023).

Recent hybrid approaches propose leveraging the strengths of both EBMs and Flow Matching to address the shortcomings of each (Loo et al., 2025). Specifically, they aim to retain the interpretability and outlier sensitivity of EBMs while improving training and inference scalability through Flow Matching. A central idea in these methods is to train a Flow Matching model such that its induced distribution converges to a Boltzmann distribution. This allows the EBM's energy function to be implicitly learned within the parameters of the flow model, bypassing the need for explicit energy function parameterization and traditional MCMC-based training (Loo et al., 2025).

One such approach, rooted in the Jordan–Kinderlehrer–Otto (JKO) scheme for Wasserstein gradient flows, proposes learning a scalar potential V whose corresponding Boltzmann distribution models the data (Balcerak et al., 2025). This is typically achieved via a two-phase training strategy: an initial warm-up phase using a flow-like objective at zero temperature ($\varepsilon=0$) to efficiently transport noise samples towards the data manifold and generate high-quality negative samples, followed by a main phase that jointly optimizes the flow objective with a contrastive divergence loss at increasing temperatures (ε) to refine the potential and ensure the learned Boltzmann distribution accurately reflects the data's likelihood (Balcerak et al., 2025).

For anomaly detection, the implicitly learned energy landscape can be directly used to assign energy scores to new data points. As with traditional EBMs, those with high energy (low likelihood under the model) are flagged as anomalies (Yoon et al., 2023). Crucially, this approach eliminates the slow MCMC-based sampling loops typically required for inference in EBMs, making it scalable and robust even in high-dimensional settings (Loo et al., 2025).

3. Summary of the State of the Art

Sensor anomaly detection is a critical task challenged by the complex, high-dimensional, and dynamic nature of modern sensor data (Shiva et al., 2024; Zamanzadeh Darban et al., 2024). While traditional statistical methods struggle with these complexities (Montgomery, 2024), machine learning, particularly deep generative models, has emerged as a powerful alternative capable of learning intricate patterns in an unsupervised manner (Hoh et al., 2022). These models learn the distribution of normal data and identify anomalies as deviations, overcoming the need for labeled anomaly data, which is often scarce (Geiger et al., 2020). Among generative approaches, several prominent models have been explored. EBMs provide an interpretable framework by assigning an energy score to data points, with higher scores indicating anomalies (Nijkamp et al., 2019; Yoon et al., 2023). Their main drawback lies in challenging training dynamics, often requiring computationally expensive and potentially unstable MCMC methods due to the intractable partition function (Du & Mordatch, 2019; Xiao et al., 2021). Diffusion Models represent another potent class, achieving state-of-the-art generation quality (LingYang et al., 2023). However, their application to anomaly detection typically relies on reconstruction error, which can be computationally intensive due to the iterative denoising process, and they suffer from a lack of interpretability (Katsuoka et al., 2024; Liu et al., 2025).

More recently, Flow Matching has gained attention as an efficient and stable method for training CNFs without simulation (Lipman et al., 2023). It offers faster training and potentially faster inference compared to diffusion models, using direct regression onto target vector fields (Lipman et al., 2023; Patel et al., 2024). Flow Matching also presents flexibility, bridging concepts from diffusion and path-based generation, and has shown potential for stabilising EBM training (Chao et al., 2023; Lipman et al., 2023). This has led to emerging hybrid approaches that aim to combine the interpretability of EBMs with the training efficiency and stability of Flow Matching (Loo et al., 2025). These methods propose learning the EBM's energy function implicitly within a Flow Matching framework, potentially leveraging techniques like the JKO scheme to bypass traditional EBM training hurdles (Balcerak et al., 2025; Loo et al., 2025). The literature describes how the learned energy landscape can be used for anomaly detection by scoring data points (Yoon et al., 2023).

However, a clear gap exists in the empirical validation and specific application of these novel hybrid EBM and Flow Matching models directly for the task of sensor signal anomaly detection. While the theoretical foundations, improved performance and potential benefits of combining these frameworks have been proposed (Balcerak et al., 2025; Loo et al., 2025), their practical effectiveness, scalability, and performance compared to established methods within the specific context of the anomaly detection for complex signal data (characterized by phenomena like bias, drift, and spikes) remain unexplored in the current researches. The proposed research project aims to directly fill this gap. It will involve developing, implementing, and rigorously evaluating a hybrid model leveraging Flow Matching to train an EBM specifically tailored for sensor signal anomaly detection. This project will investigate whether this hybrid approach can effectively retain the interpretability benefits of EBMs while harnessing the computational efficiency of Flow Matching, thereby providing a robust, scalable, and interpretable solution for identifying diverse fault types in real-world sensor signals. The evaluation will focus on performance metrics and the model's ability to handle various sensor fault types compared to existing generative approaches.

4. Research Project Plan

This section outlines the plan to address the identified gap in Section 3: the lack of empirical validation and specific application of hybrid EBMs and Flow Matching models for the task of sensor signal anomaly detection. The plan details the objectives, methodology, evaluation strategy, and ethical considerations for developing and assessing such a hybrid model.

The core aim is to empirically investigate the practical possibility of combining EBMs and Flow Matching for robust, interpretable, and efficient sensor anomaly detection. This translates into the following specific objectives:

1. **Develop a Hybrid EBM/Flow Matching Model for Sensor Signal Anomaly Detection:** Implement a functional hybrid model inspired by recent theoretical work (e.g., Balcerak et al., 2025; Loo et al., 2025).
2. **Train and Test the Model on Benchmark Sensor Data:** Employ the dataset, training the model in an unsupervised fashion. Generate realistic fault types for robust anomaly detection evaluation.
3. **Evaluate Performance:** Assess the model’s performance using standard anomaly detection metrics to determine its effectiveness.

4.1. Methodology

In this section, we describe the full pipeline for anomaly detection, starting with signal preprocessing—normalization, segmentation, and transformation—to extract meaningful temporal and spectral features. We then implement and train a hybrid Energy-Based Model with Flow Matching on clean sensor data to learn its normal distribution, enabling unsupervised detection of anomalies via learned energy scores.

4.1.1. Data Preparation

We utilize the Multiphase Flow Process (MFP) dataset introduced by Ruiz-Cárcel et al. (2015), which captures time-series data from a wide range of industrial-grade sensors monitoring a controlled flow of water, oil, and air through a complex pipeline system. The dataset captures rich, nonlinear sensor dynamics characteristic of real-world industrial processes, making it a valuable resource for developing and evaluating anomaly detection models under realistic operating conditions.

The data preprocessing pipeline consists of three key phases: normalization, segmentation, and signal transformation. Each phase plays an important role in preparing the raw sensor signals for effective learning and anomaly detection.

- **Normalization:** Sensor signals can vary significantly in scale depending on the type of sensor or measurement unit used. For example, one sensor may produce values ranging from 0 to 100, while another may range from 0 to 1. To ensure consistent input across different sensors and to stabilize training, each signal is normalized to a common scale. This helps the model focus on the signal's shape and pattern rather than absolute magnitudes.
- **Segmentation:** Sensor data is typically continuous and high-frequency. To process this data with models that expect fixed-size input, the signals are segmented into overlapping windows of a fixed length. This converts the time series into a sequence of smaller sub-sequences, where each window captures a local pattern of signal behavior. The window size and overlap ratio are treated as hyperparameters that can be tuned based on preliminary experiments. This windowing approach also provides a temporal context for anomaly detection.
- **Signal Transformation:** To enhance the representation of each segmented window, signal transformation techniques are applied. These transformations extract meaningful features that reveal both the temporal and spectral properties of the signal.
    - The Short-Time Fourier Transform (STFT) (Oppenheim & Lim, 1981) is used to convert each time-domain window into a time-frequency representation, as illustrated in Figure 2. STFT helps the model capture non-stationary behaviors such as drifts, spikes, or frequency shifts, which are common in sensor faults.
    - In addition, the Discrete Shocklet Transform (DST) (Dewhurst et al., 2020) may be explored to extract shape-based features from the signals. DST is particularly effective at capturing sharp transitions and localized signal changes, complementing the smooth, frequency-focused representations from STFT. The inclusion of DST will be based on exploratory analysis and its contribution to model performance.

For evaluation purposes, we generate labeled test samples by injecting five common sensor fault types—bias, drift, erratic, spike, and stuck—into segments of the normal MFP dataset, following augmentation methodologies proposed in prior work (Ding et al., 2022). To simulate both in-distribution and out-of-distribution fault scenarios, we introduce positive bias and drift, as well as normally distributed erratic noise. Additionally, we inject negative bias and drift, uniformly distributed erratic noise, and two distinct fault types—spike and stuck—that represent more challenging, unseen conditions. These synthetic faults are applied exclusively during testing to rigorously evaluate model robustness, while the model itself is trained in an unsupervised manner on clean, fault-free data. This setup enables a comprehensive assessment of anomaly detection performance across both familiar and novel fault distributions.

4.1.2. Model Implementation

The core of our generative model is built upon the principles of Variational Potential Flow Bayes (VPFB), drawing inspiration from recent advancements (Loo et al., 2025). Instead of directly training a vector field to match predefined targets, this approach focuses on learning a time-dependent potential energy function $\Phi(x, t)$, parameterized by a neural network. The negative gradient of this potential, $-\nabla_x \Phi(x, t)$, implicitly defines the velocity field of a continuous-time flow. The model is trained such that the probability density transported by this potential flow matches a target density homotopy (a smooth path of distributions from a simple prior to the data distribution) over time $t \in [0, t_{end}]$. A key theoretical result of VPFB is that, as the density homotopy approaches a stationary equilibrium (for $t \geq t_{max}$), the learned potential $\Phi(x, t)$ relates directly to the Boltzmann energy $E(x)$ of a standard EBMs, such that $P(x) \propto e^{-E(x)}$. Our goal is to learn this implicit EBM representing the distribution of normal sensor data without relying on traditional, often unstable, MCMC-based EBM training. Figure 3 illustrates the complete pipeline for our VPFB-based sensor signal anomaly detection system, divided into training (upper panel) and anomaly detection (lower panel).

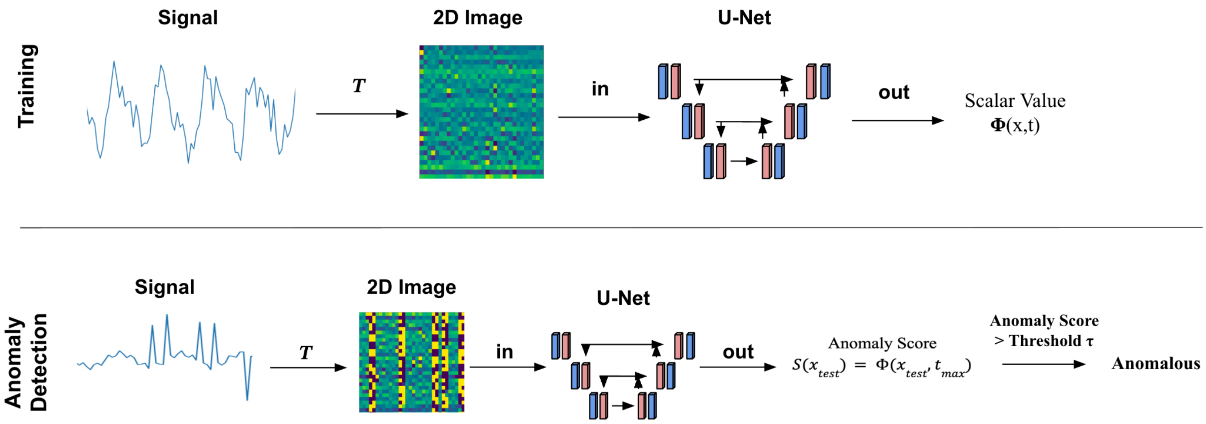

Figure 3: Anomaly Detection Pipeline using Hybrid EBM/Flow Matching. Illustrates training and detection workflows including signal processing, U-Net, and energy scoring.

- **Neural Network Architecture:** Based on VPFB (Loo et al., 2025), the time-dependent potential energy function $\Phi(x, t)$ is parameterized by a neural network based on the U-Net architecture (Ronneberger et al., 2015). Hyperparameters for the U-Net were adopted from Lipman et al. (2023). The network accepts the data input x and time t, outputting the scalar potential energy value $\Phi(x, t)$.

- **Training Protocol:** The neural network parameterizing Φ is trained by minimizing a variational loss function derived from the principle of matching the flow-driven density homotopy to the target marginal density homotopy. This loss, specific to the VPFB framework (Loo et al., 2025), avoids explicit density evaluation and MCMC sampling. The overall training objective is to minimize the expected value of a time-dependent loss $L(\Phi, t)$ over the training duration, effectively averaging over sampled times:

$$min_{\theta} E_{t \sim U(0, t_{end})}[L(\Phi_{\theta}, t)]$$

where Φ_θ denotes the potential function parameterized by network weights θ, and $L(\Phi, t)$ is calculated for each sampled time t and batch of data. The time-dependent loss function is (Loo et al., 2025, Equation 25):

$$L(\Phi, t) = Cov_{p(x|\bar{x},t)P_{data}(\bar{x})}[\Phi(x, t), w(t)\gamma(x, \bar{x}, t)]$$

$$- \frac{\nabla_x \Phi(x,t) \cdot v(x|\bar{x},t)}{||\nabla_x \Phi(x,t)|| \, ||v(x|\bar{x},t)||} + E_{p(x|\bar{x},t)P_{data}(\bar{x})}[||\nabla_{(x,t)} \Phi(x, t)||^2 + \eta||\Phi(x, t)||^2]$$

In practice, the expectations and covariances are computed using mini-batches of samples. For each training step:

1. Sample a batch of normal sensor data segments $\bar{x} \sim P_{data}$.
2. Sample a batch of time points $t \sim U(0, t_{end})$
3. For each $(\bar{x}, t)$ pair, sample a perturbed point x from the conditional distribution $p(x|\bar{x},t)$.
4. Compute the corresponding innovation term $\gamma(x, \bar{x},t)$ and the target conditional vector field $v(x|\bar{x},t)$
5. Compute $\Phi(x, t)$, its spatial gradient $\nabla_x \Phi(x, t)$, and its time derivative $\vartheta_t \Phi(x, t)$ using the neural network and automatic differentiation.
6. Calculate the loss $L(\Phi, t)$ for each sample in the batch and the average.
7. Backpropagate and update network weights using optimizers like Adam.

This training protocol ensures that the learned potential function guides the flow to match the desired density evolution, thereby learning the energy landscape of the normal data distribution. Elements from advanced training strategies, such as multi-stage training based on insights from Balcerak et al. (2025), may be explored to further stabilize training or enhance the discriminative power of the learned energy function. Hyperparameters will be tuned using a validation set of normal data.

- **Anomaly Scoring Mechanism:** Upon successful training of the potential function $\Phi(x, t)$, the implicitly learned energy function for a test sample x_{test} is related to $\Phi(x_{test}, t)$ at or near the stationary equilibrium time t_{max}. Following common practice in energy-based anomaly detection (Yoon et al., 2023), we use the learned potential energy evaluated at this equilibrium time as the anomaly score for a test sample x_{test}:

$$S(x_{test}) = \Phi(x_{test}, t_{max})$$

A higher energy value $S(x_{test})$ indicates that the sample x_{test} falls in a region of low probability density under the learned model of normal data, thus signifying a higher likelihood of being anomalous. Anomalies are identified by setting a threshold on this score; test samples with $S(x_{test})$ exceeding the threshold are flagged as anomalies.

4.2. Evaluation Strategy

To evaluate the effectiveness of our anomaly detection model on time-series sensor data, we adopt the Area Under the Receiver Operating Characteristic Curve (AUROC) as the primary performance metric, following the approach introduced by Koizumi et al. (2020). AUROC is widely used in anomaly detection tasks as it provides a threshold-independent measure of a model's ability to separate normal from anomalous signal patterns, which is particularly important when ground-truth labels are sparse or uncertain. In addition, we report F1-score, precision, and recall, which are standard metrics for evaluating classification performance in imbalanced settings. These metrics allow us to quantify the trade-off between false positives and false negatives, especially relevant in sensor-based monitoring where anomalies are rare but critical to detect. This combination of metrics provides a robust and interpretable evaluation framework for assessing both detection accuracy and reliability.

4.3. Ethical Considerations

The dataset used in this study (Ruiz-Cárcel et al., 2015) contains only industrial sensor data with no human subjects or sensitive information. Its use is permitted under a permissive open-source license, and we have obtained the necessary rights for research use. Thus, there is no need for ethical considerations in this research.

5. Conclusion

This report has provided a comprehensive review of sensor anomaly detection, highlighting the complexities of modern sensor data and the evolution from traditional statistical methods to advanced deep learning approaches. We examined key generative models, including Energy-Based Models (EBMs), valued for their interpretability but challenged by training complexities involving MCMC methods; Diffusion Models, recognized for state-of-the-art performance but often hampered by slow inference and limited explainability; and Flow Matching, an emerging efficient and stable framework for training continuous normalizing flows. Recent theoretical developments propose hybrid models combining the interpretability strengths of EBMs with the computational efficiency and stability of Flow Matching, by learning an implicit energy potential via flow-based objectives like the Variational Potential Flow Bayes (VPFB) framework. However, a notable gap exists in the literature regarding the specific application and empirical validation of these promising hybrid approaches for the distinct challenges of sensor anomaly detection. This research project aims to fill this crucial gap. We propose the development, implementation, and rigorous evaluation of a hybrid EBM/Flow Matching model tailored for sensor signals, utilizing the VPFB approach on the benchmark MFP dataset with synthetically generated faults. The primary objective is to empirically assess if this hybrid strategy can deliver a robust, scalable, and interpretable solution for identifying diverse sensor anomalies, thereby advancing the state-of-the-art in unsupervised anomaly detection for critical monitoring systems.

6. References

Balcerak, M., Amiranashvili, T., Shit, S., Terpin, A., Kaltenbach, S., Koumoutsakos, P., & Menze, B. (2025). *Energy Matching: Unifying Flow Matching and Energy-Based Models for Generative Modeling*. https://doi.org/10.48550/arXiv.2504.10612

Carbone, D. (2024). *Hitchhiker's guide on Energy-Based Models: a comprehensive review on the relation with other generative models, sampling and statistical physics*. https://doi.org/10.48550/arXiv.2406.13661

Chao, C.-H., Sun, W.-F., Hsu, Y.-C., Kira, Z., & Lee, C.-Y. (2023). *Training Energy-Based Normalizing Flow with Score-Matching Objectives*. https://doi.org/10.48550/arxiv.2305.15267

Dewhurst, D. R., Alshaabi, T., Kiley, D., Arnold, M. V., Minot, J. R., Danforth, C. M., & Dodds, P. S. (2020). The shocklet transform: a decomposition method for the identification of local, mechanism-driven dynamics in sociotechnical time series. *EPJ Data Science*, *9*(1). https://doi.org/10.1140/epjds/s13688-020-0220-x

Ding, Z. Y., Loo, J. Y., Surya Girinatha Nurzaman, Tan, C. P., & Baskaran, V. M. (2022). A Zero-Shot Soft Sensor Modeling Approach Using Adversarial Learning for Robustness Against Sensor Fault. *IEEE Transactions on Industrial Informatics*, *19*(4), 5891–5901. https://doi.org/10.1109/tii.2022.3187708

Du, Y., Li, S., Tenenbaum, J., & Igor Mordatch. (2020). Improved Contrastive Divergence Training of Energy Based Models. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2012.01316

Du, Y., & Mordatch, I. (2019). Implicit Generation and Modeling with Energy Based Models. *Neural Information Processing Systems*, *32*, 3603–3613.

Geiger, A., Liu, D., Alnegheimish, S., Cuesta-Infante, A., & Veeramachaneni, K. (2020, December 1). *TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks*. IEEE Xplore. https://doi.org/10.1109/BigData50022.2020.9378139

Hoh, M., Schöttl, A., Schaub, H., & Wenninger, F. (2022). A Generative Model for Anomaly Detection in Time Series Data. *Procedia Computer Science*, *200*, 629–637. https://doi.org/10.1016/j.procs.2022.01.261

Katsuoka, T., Shiraishi, T., Miwa, D., Nguyen Le Duy, V., & Takeuchi, I. (2024). *Statistical Test on Diffusion Model-based Anomaly Detection by Selective Inference*. https://doi.org/10.48550/arXiv.2402.11789

Koizumi, Y., Kawaguchi, Y., Imoto, K., Nakamura, T., Nikaido, Y., Tanabe, R., Purohit, H., Suefusa, K., Endo, T., Yasuda, M., & Harada, N. (2020). *Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring*. https://doi.org/10.48550/arXiv.2006.05822

Li, D., Wang, Y., Wang, J., Wang, C., & Duan, Y. (2020). Recent advances in sensor fault diagnosis: A review. *Sensors and Actuators A: Physical*, *309*, 111990.

https://doi.org/10.1016/j.sna.2020.111990

Li, Z., Chen, Y., & Sommer, F. T. (2023). Learning Energy-Based Models in High-Dimensional Spaces with Multiscale Denoising-Score Matching. *Entropy*, *25*(10), 1367–1367. https://doi.org/10.3390/e25101367

LingYang, L., Zhang, Z., Song, Y., Shenda, H., Xu, R., Zhao, Y., Zhang, W., Cui, B., & Yang, M.-H. (2023). Diffusion Models: A Comprehensive Survey of Methods and Applications. *ACM Computing Surveys*, *56*(4), 1–39. https://doi.org/10.1145/3626235

Lipman, Y., Chen, R. T. Q., Ben-Hamu, H., Nickel, M., & Le, M. (2023). *Flow Matching for Generative Modeling*. https://doi.org/10.48550/arXiv.2210.02747

Liu, J., Ma, Z., Wang, Z., Liu, Y., Wang, Z., Sun, P., Song, L., Hu, B., Boukerche, A., & C.M. Leung, V. (2025). A Survey on Diffusion Models for Anomaly Detection. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2501.11430

Loo, J. Y. L. P., Vishnu Monn Baskaran, Chee-Ming Ting, Raphaël C.-W. Phan, Adeline, M., Lau, J. K., Leong, F. Y., Tew, H. H., Pal, A., Baskaran, V. M., Ting, C.-M., & Phan, R. C.-W. . (2025). *Learning Energy-Based Generative Models via Potential Flow: A Variational Principle Approach to Probability Density Homotopy Matching*. https://doi.org/10.48550/arXiv.2504.16262

Ma, Z., & Collins, M. (2018). Noise Contrastive Estimation and Negative Sampling for Conditional Models: Consistency and Statistical Efficiency. In E. Riloff, D. Chiang, J. Hockenmaier, & J. Tsujii (Eds.), *In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing* (pp. 3698–3707). Association for Computational Linguistics. https://doi.org/10.48550/arXiv.1809.01812

Montgomery, R. M. (2024). *Techniques for Outlier Detection: A Comprehensive View*. https://doi.org/10.20944/preprints202410.1603.v1

Naiman, I., Berman, N., Pemper, I., Arbiv, I., Fadlon, G., & Azencot, O. (2024). Utilizing Image Transforms and Diffusion Models for Generative Modeling of Short and Long Time Series. *Neural Information Processing Systems 2024*. https://doi.org/10.48550/arXiv.2410.19538

Nijkamp, E., Hill, M., Zhu, S.-C., & Wu, Y. N. (2019). Learning Non-Convergent Non-Persistent Short-Run MCMC Toward Energy-Based Model. *Neural Information Processing Systems*, *32*, 5232–5242.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE*, *69*(5), 529–541. https://doi.org/10.1109/proc.1981.12022

Patel, Z., DeLoye, J., & Mathias, L. (2024). *Exploring Diffusion and Flow Matching Under Generator Matching*. https://doi.org/10.48550/arXiv.2412.11024

Pintilie, I., Manolache, A., & Brad, F. (2023). Time Series Anomaly Detection using Diffusion-based Models. *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, 570–578. https://doi.org/10.1109/icdmw60847.2023.00080

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1505.04597

Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Ruiz-Cárcel, C., Cao, Y., Mba, D., Lao, L., & Samuel, R. T. (2015). Statistical process monitoring of a multiphase flow facility. *Control Engineering Practice*, *42*, 74–88. https://doi.org/10.1016/j.conengprac.2015.04.012

Sakai, S., & Hasegawa, T. (2025). *Reconstruction-Free Anomaly Detection with Diffusion Models via Direct Latent Likelihood Evaluation*. https://doi.org/10.48550/arXiv.2504.05662

Shiva, K., Etikani, P., Bhaskar, V. V. S. R., Mittal, A., Dave, A., Thakkar, D., Kanchetti, D., & Munirathnam, R. (2024). Anomaly Detection in Sensor Data with Machine Learning: Predictive Maintenance for Industrial Systems. *Journal of Electrical Systems*, *20*(10s), 454–462. https://journal.esrgroups.org/jes/article/view/5137

Xiao, Z., Yan, Q., & Amit, Y. (2021). *EBMs Trained with Maximum Likelihood are Generator Models Trained with a Self-adverserial Loss*. https://doi.org/10.48550/arXiv.2102.11757

Yang, S., Chen, Y., Wang, L., Liu, S., & Chen, Y. (2023). *Denoising Diffusion Step-aware Models*. https://doi.org/10.48550/arxiv.2310.03337

Yi, T.-H., Huang, H., & Li, H.-N. (2017). Development of sensor validation methodologies for structural health monitoring: A comprehensive review. *Measurement*, *109*, 200–214. https://doi.org/10.1016/j.measurement.2017.05.064

Yoon, S., Jin, Y.-U., Noh, Y.-K., & Park, F. C. (2023). Energy-based models for anomaly detection: a manifold diffusion recovery approach. *Proceedings of the 37th International Conference on Neural Information Processing Systems*, 49445–49466.

Zamanzadeh Darban, Z., Webb, G. I., Pan, S., Aggarwal, C., & Salehi, M. (2024). Deep learning for time series anomaly detection: A survey. *ACM Computing Surveys*, *57*(1), 1–42. https://doi.org/10.1145/3691338

Zhai, S., Cheng, Y., Lu, W., & Zhang, Z. (2016). Deep Structured Energy Based Models for Anomaly Detection. *In Proceedings of the 33rd International Conference on International Conference on Machine Learning*, *48*, 1100–1109. https://doi.org/10.48550/arxiv.1605.07717

Zhao, H., Zuo, S., Hou, M., Liu, W., Yu, L., Yang, X., & Deng, W. (2018). A Novel Adaptive Signal Processing Method Based on Enhanced Empirical Wavelet Transform Technology. *Sensors*, *18*(10), 3323. https://doi.org/10.3390/s18103323