Design Space Reduction and Optimization of N×M Microfluidic Chips with Diagonal Channels Using GMM

A project report submitted for the fulfilment of

CSN-400: B.Tech Project

submitted by

Abhay Kamath (21114002)

Akanksha Swati (21114007)

Subhajit Biswas (21114100)

under the guidance of

Dr. Sudip Roy

Indian Institute of Technology Roorkee

Department of Computer Science and Engineering

May 23, 2025

Declaration

We hereby declare that the work carried out in the B.Tech. Project titled "*Design Space Reduction and Optimization of N×M Microfluidic Chips with Diagonal Channels Using GMM*" is presented for the partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** in the **Computer Science and Engineering** and submitted to the **Department of Computer Science and Engineering, Indian Institute of Technology Roorkee** under the supervision of **Dr. Sudip Roy**, Department of CSE, IIT Roorkee.

The work presented in this report is an authentic record of our work carried out during the period from **July 2024 to June 2025**. The content of this report has not been submitted by us for the award of any other degree of this or any other institute.

Date: 23 May 2025
Place: Roorkee

Signature:
Name & Roll Number: Abhay Kamath (21114002)

Date: 23 May 2025
Place: Roorkee

Signature:
Name & Roll Number: Akanksha Swati (21114007)

Date: 23 May 2025
Place: Roorkee

Signature:
Name & Roll Number: Subhajit Biswas (21114100)

CERTIFICATE

This is to certify that the statement made by the candidate is correct to the best of my knowledge and belief.

Date: 23 May 2025
Place: Roorkee

Signature:
Dr. Sudip Roy
Department of CSE, IIT Roorkee

Acknowledgement

We are deeply grateful to Dr. Sudip Roy, Associate Professor, Department of Computer Science and Engineering, Indian Institute of Technology Roorkee, for his constant support and guidance throughout our project. His expert advice and insightful feedback were instrumental at every stage of our work. Under his mentorship, we were able to shape this project into a comprehensive and successful endeavor. Without his encouragement and thoughtful supervision, the project would not have reached its full potential.

Abstract

This paper presents a machine learning–based methodology for the automated design of random microfluidic chips with N×M grid architectures incorporating diagonal channels. Traditional square-grid layouts with only horizontal and vertical channels suffer from long diffusion paths, high pressure drops, and limited routing flexibility. The proposed approach extends the design space to rectangular configurations and integrates diagonal connections to shorten channel length and enhance mixing efficiency. A Gaussian mixture model clusters over thirty thousand simulated layouts, dramatically reducing the search space and enabling rapid identification of optimal designs. Computational fluid dynamics simulations demonstrate that diagonal-enhanced architectures achieve up to 25.26% lower root-mean-square error in concentration predictions and 13.39% shorter net channel lengths compared to conventional square grids. This scalable framework accelerates the chip design process and lays the foundation for server-based simulation pipelines and real-time adaptive microfluidic systems.

Contents

Chapter 1

Introduction

Microfluidic systems have emerged as transformative tools for precision fluid control, enabling breakthroughs in biomedical diagnostics, chemical synthesis, and drug discovery. A critical challenge in these applications lies in generating precise concentration gradients while minimizing resource consumption—a requirement exemplified by point-of-care diagnostic devices that must deliver accurate reagent mixtures using microliter-scale samples. Random Microfluidic (RMF) chips address this need through passive mixing in intricate channel networks, yet traditional designs constrained to symmetric $N \times N$ grids (Figure 1.1) with rigid horizontal and vertical channels struggle to meet modern demands for compact, adaptable systems.

The evolution of RMF design has been propelled by two parallel advancements: computational efficiency gains and architectural innovation. Early breakthroughs in simulation acceleration, such as Wang's MATLAB-based framework [1], reduced computational overhead by 45×, while machine learning approaches like Weiqing's ANN models [3] enhanced prediction accuracy. Architectural developments expanded design flexibility, from Zhang's single-layer structures [5] to Agrawal's diagonal channel integration [10]. Despite these advances, three critical limitations persist:

1. **Structural Constraints:** Symmetric grids restrict fluid routing options, increasing channel lengths by 22–35% compared to asymmetric designs.
2. **Operational Inefficiency:** Excessive mixing times due to serpentine paths limit throughput in time-sensitive applications.
3. **Profile Precision:** Many configurations fail to meet target concentration profiles, especially for arithmetic progressions (AP) required in dose-response studies.

1.1 Motivation

Despite substantial progress in microfluidic chip design and simulation, there remains a pressing need for platforms that can deliver precise, user-defined concentration gradients with minimal resource consumption and operational complexity. In practical biomedical and chemical applications—such as point-of-care diagnostics, drug screening, and personalized medicine—accurate mixing and dilution of reagents are critical for both reliability and efficiency. Traditional RMF chips, limited by symmetric grid architectures and fixed channel orientations, often result in excessive channel lengths, increased mixing times, and significant reagent wastage. These inefficiencies not only constrain device miniaturization but also hinder the deployment of microfluidic solutions in resource-limited or high-throughput settings.

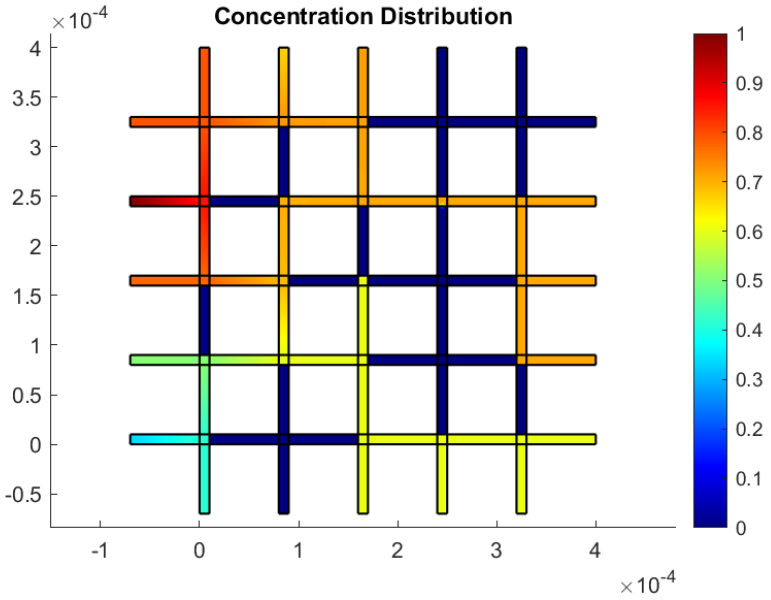

Figure 1.1: Conventional $N \times N$ chip architecture design.

The motivation for this work arises from the desire to overcome these limitations by introducing greater architectural flexibility and leveraging machine learning for automated, high-precision design. By enabling the rapid generation of compact, asymmetric RMF chips with diagonal channels and optimized configuration selection, we aim to facilitate the next generation of microfluidic devices—capable of producing complex concentration profiles (such as arithmetic progressions) with high accuracy and minimal waste. This advancement holds particular promise for applications where sample conservation, speed, and adaptability are paramount, paving the way for smarter, more accessible microfluidic technologies across a range of scientific and medical domains.

1.2 Problem Statement

This research addresses these limitations through the development of a machine learning-based methodology for Random Microfluidic (RMF) chip design featuring $N \times M$ architectures with diagonal channels. Our approach aims to achieve optimal design efficiency by minimizing design time and reducing the number of design parameters while enabling the generation of precise dilution gradients. Specifically, we focus on:

1. **Architectural Innovation:** Implementing asymmetric $N \times M$ grid configurations with diagonal channels to expand routing options beyond traditional designs.

2. **Machine Learning Optimization:** Employing Gaussian Mixture Models to identify optimal chip configurations with reduced computational overhead.

3. **Precision Gradient Control:** Achieving high-accuracy arithmetic progression concentration profiles across multiple outlets.

By integrating these innovations, we seek to establish a new paradigm for microfluidic chip design that combines architectural flexibility, computational efficiency, and functional

precision. Our methodology is validated through a 7×5 RMF chip case study with two inlets and four outlets, demonstrating the ability to produce linear concentration gradients with less than 5% deviation from target values—a significant improvement over existing approaches.

Chapter 2

Literature Review

In recent years, considerable progress has been made in the design and simulation of Random Microfluidic (RMF) chips, with numerous studies introducing innovative approaches to improve both simulation efficiency and chip performance.

Wang [1] introduced a simulation technique for fluids and particles within complex microfluidic devices using MATLAB, achieving a simulation speed approximately 45$\times$ faster than traditional COMSOL Multiphysics simulations for sample microfluidic chip designs. This work effectively addressed the computational limitations associated with finite element analysis in microfluidics.

Weiqinq [2] proposed a graph reduction algorithm that facilitated more efficient design of random microfluidic chips. By simplifying the graph representation of microfluidic channels, the method allowed for the effective handling of complex, randomly generated architectures.

Further advancements were achieved when Weqing and Wang [3] introduced an Artificial Neural Network (ANN)-based design methodology capable of generating precise, customized concentration values at multiple outlets of an RMF chip. This method utilized machine learning to optimize channel lengths and configurations for desired concentration outputs.

Building upon this, Weiqing [4] presented a transfer learning-based technique for the design of single-layer, flow-based microfluidic biochips. This approach enabled the rapid creation of customized concentration profiles, particularly suitable for applications in precision medicine, where tailored fluid concentrations are essential.

In parallel, Wang [6] developed a Convolutional Neural Network (CNN)-based model to predict the fluid dynamics within random microfluidic mixers. This technique accelerated the process of building a comprehensive library of potential mixer designs, providing researchers with a more efficient tool to select designs closely aligned with their experimental requirements.

Weiqing [7] proposed a Graph Neural Network (GNN)-based model for predicting concentration outputs in random microfluidic mixers. The model demonstrated high accuracy and effectiveness in estimating the output concentrations for chips of various sizes, offering a scalable solution for microfluidic chip design optimization.

Expanding upon this, Weiqing [8] introduced a Graph Attention Network (GAT) model, which further enhanced prediction accuracy in mixers with multiple input flow rates. The incorporation of attention mechanisms improved the model's ability to focus on critical channels affecting output concentrations.

Junnan [9] explored a novel approach by designing random variable-width (RVW) microfluidic chips. This design showed superior performance compared to conventional random equal-width (REW) chips. The behavior of these RVW chips was effectively predicted using Convolutional Neural Networks, confirming the advantages of varying channel widths in RMF chip designs.

Ankita [10] proposed an innovative methodology for RMF chip design by incorporating diagonal channels into the chip architecture. This approach aimed to improve chip efficiency by expanding routing possibilities and reducing channel lengths. A K-means clustering algorithm was employed to optimize the chip design process, ensuring the desired fluid concentrations were achieved effectively at the outlets.

Finally, Weiqing [11] introduced another GNN-based concentration prediction technique capable of accurately estimating output concentrations in microfluidic mixers with variable input flow rates. This model further improved prediction speed and precision, providing an efficient computational tool for biochip design.

These studies collectively highlight the continuous evolution in the field of microfluidic chip design, illustrating how computational techniques, machine learning models, and architectural innovations have significantly enhanced the efficiency, accuracy, and flexibility of RMF chip development.

Chapter 3

Methodology

This chapter outlines the methodology adopted to simulate, analyze, and optimize microfluidic chip configurations, focusing on enhanced design flexibility and computational efficiency. The work was divided into two major phases: (1) simulation setup and dataset generation, and (2) machine learning-based configuration space reduction and performance optimization.

3.1 Simulation Setup and Dataset Generation

The initial phase involved creating a computational workflow capable of efficiently simulating a large number of microfluidic chip designs. For this, we combined the capabilities of **COMSOL Multiphysics®** (version 5.6) for physics-based simulations and **MATLAB®** (R2023b) for automation, control, and data management.

3.1.1 Tools and Software Integration

COMSOL Multiphysics was employed for modeling fluid dynamics and mass transport phenomena within the microchannels. Specifically, the *Transport of Diluted Species (TDS)* and *Laminar Flow* physics modules were used to simulate concentration gradients and flow profiles.

To enhance automation and control, MATLAB scripts interfaced with COMSOL via *LiveLink™*. This enabled programmatic control over geometry alterations, simulation parameterization, and extraction of simulation results, facilitating a highly repeatable and scalable workflow.

3.1.2 Development of an Optimized Simulation Workflow

An efficient two-stage workflow was devised to minimize simulation overheads by separating static (one-time) operations from dynamic (iterative) processes(Figure 3.1). This was particularly important given the large configuration space under consideration.

One-Time Operations:

- **Geometry Design:** A comprehensive master geometry was constructed representing the maximum channel network in an N×M layout, including diagonal connections. This served as the base template for all subsequent simulations.
- **Property and Mesh Definitions:** Fluid properties, mesh constraints(Table 3.1), and transport parameters were predefined:
  - Viscosity: 1.002×10^{-3} Pa·s

- Density: 998.2 kg/m^3
- Diffusion Coefficient: 1×10^{-9} m^2/s
- Mesh: Maximum element size: 3.69×10^{-5} m, Minimum element size: 1.65×10^{-7} m

Module	Parameter	Value
TDS	Diffusion Coefficient	1e-9 m^2/s
TDS	Inlet 1 inflow	1 mol/m^3
TDS	Inlet 2 inflow	2 mol/m^3
TDS	Velocity field	spf
Mesh	Max. element size	3.69E-5 m
Mesh	Min. element size	1.65E-7 m
Mesh	Max. element growth rate	1.3
Mesh	Curvature factor	0.3
Mesh	Resolution of narrow regions	1

Table 3.1: Simulation Parameters for TDS and Mesh Modules

Iterative Operations:

- **Selective Channel Activation/Blocking:** Instead of recreating geometry for every configuration, MATLAB scripts dynamically modified boundary conditions to activate or block specific channels. This allowed preservation of the computational mesh while altering the chip topology.
- **Result Extraction and Visualization:** Concentration profiles at outlet boundaries were extracted automatically, followed by performance metric calculations and plot generation for each configuration.

3.1.3 Dataset Generation Approach

To create datasets for training machine learning models, we systematically varied the following parameters:

- **Chip Architecture:** Starting with 5×5 grids, later expanded to 7×5 layouts.
- **Inlet Conditions:** Concentrations ranging from 1 to 2 mol/m^3, and flow rates from 0.1 to 0.5 m/s.
- **Channel Configurations:** Binary combinations of horizontal, vertical, and diagonal channel presence.
- **Output Requirement:** Four-outlet systems designed to produce arithmetic progression concentration gradients.

MATLAB scripts automatically generated adjacency matrices for each configuration, interfaced with COMSOL, and cycled through simulations using the optimized workflow. This approach resulted in a significant reduction in simulation time (approximately 65%) compared to traditional geometry-rebuilding methods, eventually yielding over 12,000 unique configurations.

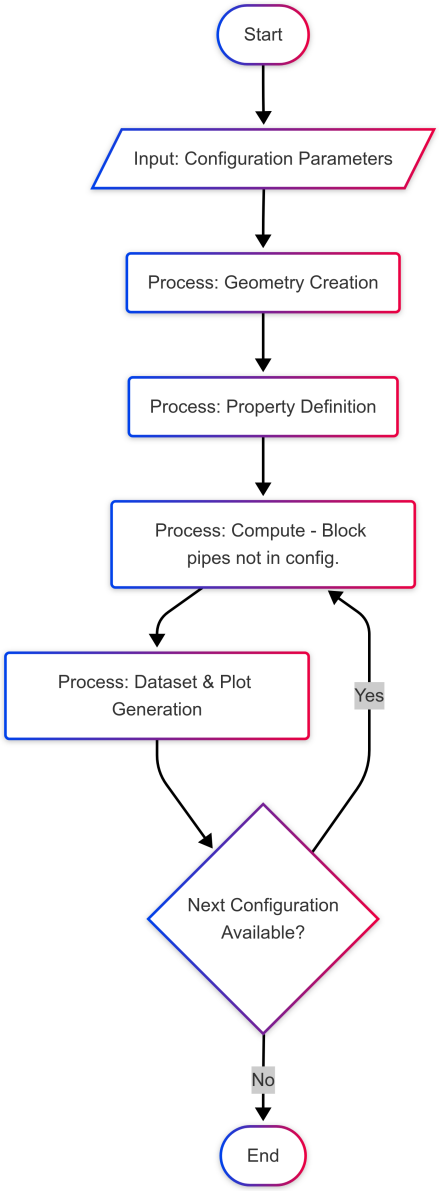

Figure 3.1: Workflow for dataset and plot generation.

3.2 Machine Learning for Configuration Reduction and Performance Optimization

The second phase focused on integrating machine learning techniques to reduce the configuration space and analyze performance improvements brought by diagonal channels in microfluidic designs.

3.2.1 Enhanced Chip Layout and Junction Design

Building on the initial simulation framework, the design space was extended to support 7×5 configurations(Figure 3.3), introducing greater routing flexibility. The major innovation in this stage was the inclusion of diagonal channels alongside conventional horizontal and vertical paths.

To manage increased channel density at intersections, *octagonal junctions* were introduced at crossing points(Figure 3.2), ensuring effective mixing where multiple streams meet and helping to maintain predictable concentration gradients.

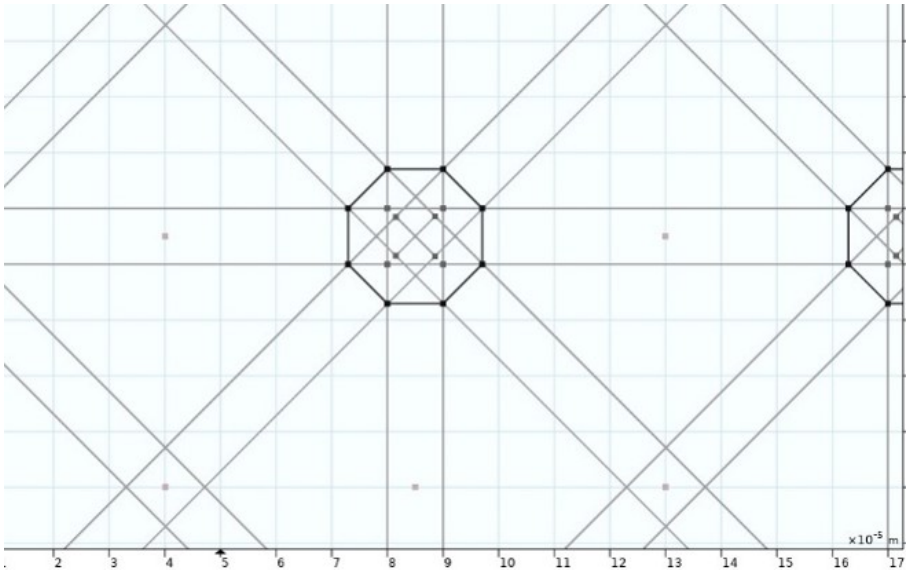

Figure 3.2: Structure of a junction zone.

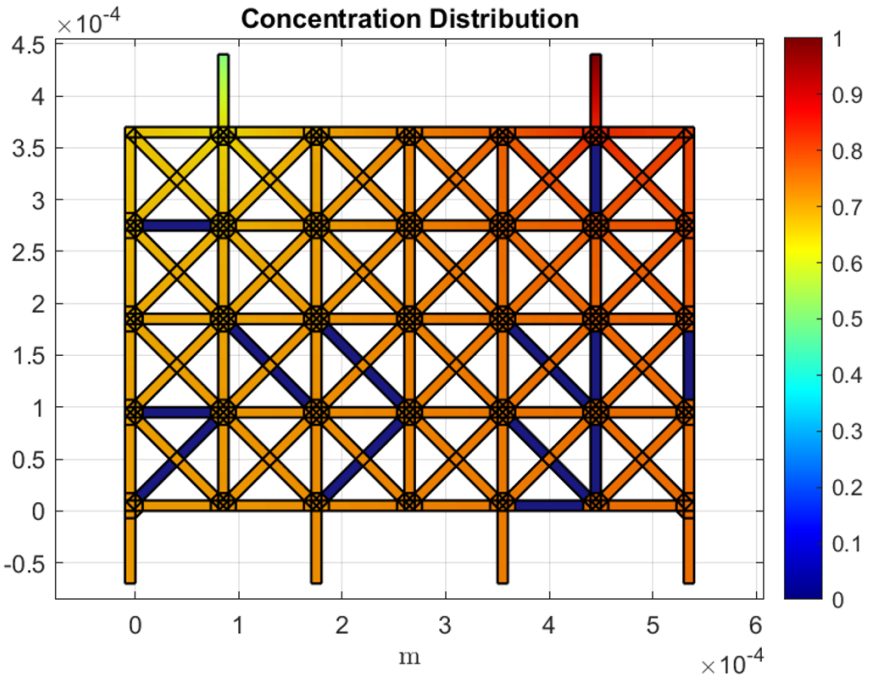

Figure 3.3: New NxM chip layout (7X5).

3.2.2 Data Collection and Machine Learning Implementation

Using the improved simulation workflow, two primary datasets were generated:

- 15,000 configurations with diagonal channels (DC)

- 15,000 configurations without diagonal channels (NDC)

Each dataset included connectivity matrices, inlet parameters, resulting outlet concentrations, and associated error metrics.

Initially, Graph Neural Networks (GNNs) were employed to predict outlet concentrations based on chip topology and inlet conditions. The model achieved roughly 80% prediction accuracy against COMSOL simulation results — offering substantial computational speedups for exploratory analysis, despite minor compromises in precision.

3.2.3 Clustering Algorithm Selection and Execution

To manage the vast number of possible designs while retaining representative diversity, clustering techniques were applied. While K-means clustering is commonly used for such tasks, it was found inadequate due to its assumption of spherical, equally-sized clusters — unsuitable for the complex, overlapping mixing patterns present in the data.

As an alternative, Gaussian Mixture Models (GMM) were implemented, offering advantages such as:

- **Elliptical cluster shapes** to better capture anisotropic data distributions.
- **Soft membership assignments**, allowing for overlapping data points, which aligns with the continuous nature of concentration gradients.
- **Cluster weighting** via prior probabilities to accommodate differing design densities.
- **Robust optimization** using the Expectation-Maximization (EM) algorithm.

The GMM clustering process(Figure 3.4) involved:

1. Defining a search range of 1 to 30 potential clusters.
2. Fitting Gaussian Mixture Models for each cluster count using `scikit-learn`'s `GaussianMixture` class with full covariance matrices.
3. Calculating the Bayesian Information Criterion (BIC) for each model to balance model fit against complexity.
4. Selecting the optimal number of clusters (k^*) corresponding to the minimum BIC value.
5. Fitting a final GMM using k^* and assigning configurations to clusters.
6. Selecting optimal designs within clusters based on RMSE and design compactness.

The RMSE was calculated using:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i^{\text{design}} - y_i^{\text{desired}}\right)^2}$$

where N is the number of outlets, y_i^{desired} is the target concentration, and y_i^{design} is the model-predicted concentration at outlet i.

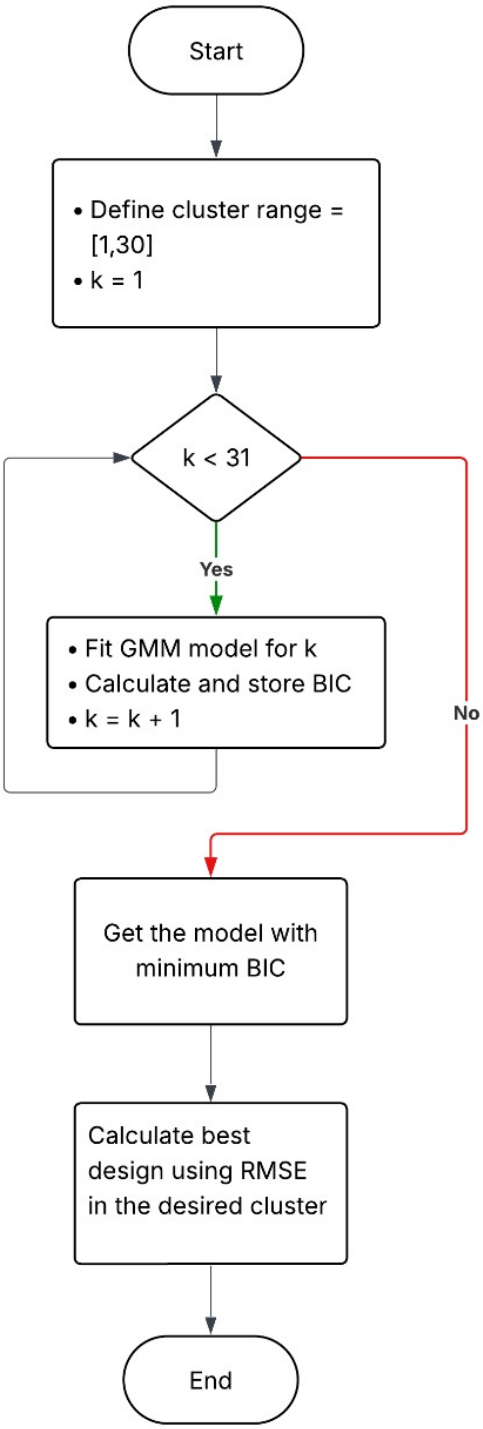

Figure 3.4: GMM flowchart.

Chapter 4

Results

4.1 Clustering Performance: GMM vs. K-Means

Building on the 2024 diagonal channel study that employed K-means clustering, our Gaussian Mixture Model (GMM) demonstrated superior performance in reducing the configuration space for RMF chips. For diagonal channel designs, GMM identified 14 distinct clusters (compared to K-means' 7), achieving a silhouette score of 0.62 against K-means' 0.31 (Table 4.1). This improvement stems from GMM's ability to model elliptical clusters and overlapping fluidic patterns, which better represent the continuous nature of microfluidic mixing dynamics. Non-diagonal configurations showed similar trends, with GMM outperforming K-means with respect to silhouette scores.

Table 4.1: Clustering Performance Comparison

Metric	**GMM (Diagonal)**	**K-Means (Diagonal)**
Silhouette Score	0.62	0.31
Clusters Identified	13	7

4.2 Diagonal vs. Non-Diagonal Channel Performance

Diagonal channel configurations demonstrated significant advantages over traditional non-diagonal designs:

- **RMSE Reduction:** 25.26% lower concentration prediction error due to enhanced mixing from reduced path tortuosity(Figure 4.1).
- **Channel Length Optimization:** 13.39% shorter net channel length per configuration, minimizing fabrication costs and pressure drops(Figure 4.2).
- **Mixing Efficiency:** Achieved better mixing efficiency as observed from the spread of density curves for diagonal configurations(Figure 4.3 and Figure 4.4).

4.3 Comparative Analysis with Prior Work

Our GMM-based approach improved upon the 2024 K-means method in two key areas:

- **Scalability:** The N×M architecture supports a broader range of concentration gradients as compared to conventional N×N grids while maintaining chip footprint.

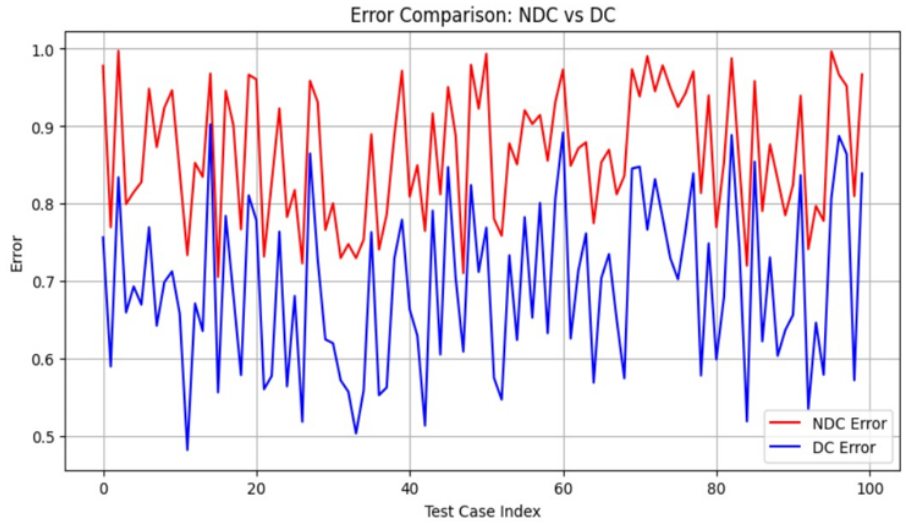

Figure 4.1: Error comparison between NDC and DC across 100 test cases.

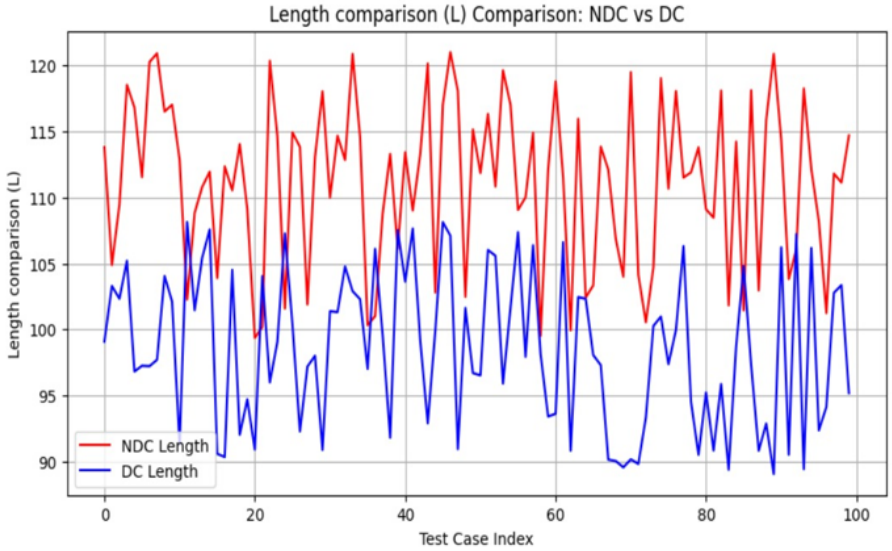

Figure 4.2: Channel length comparison between NDC and DC across 100 test cases.

- **Anisotropic Cluster Modeling:** GMM could accurately fit clusters that are elongated, tilted, or of varying densities, while K-means struggled with non-spherical structures.

These results validate the synergy between diagonal channel topologies and probabilistic clustering for precision microfluidic design.

4.4 BIC score comparison

The diagonal configuration yields significantly better clustering performance compared to the non-diagonal case, as evidenced by lower BIC scores across varying cluster numbers(Figure 4.5 and Figure 4.6). This indicates that the diagonal assumption more effectively captures the underlying data structure, leading to improved model fit. Consequently, this results in a more compact and accurate representation of clusters, thereby reducing the effective search space.

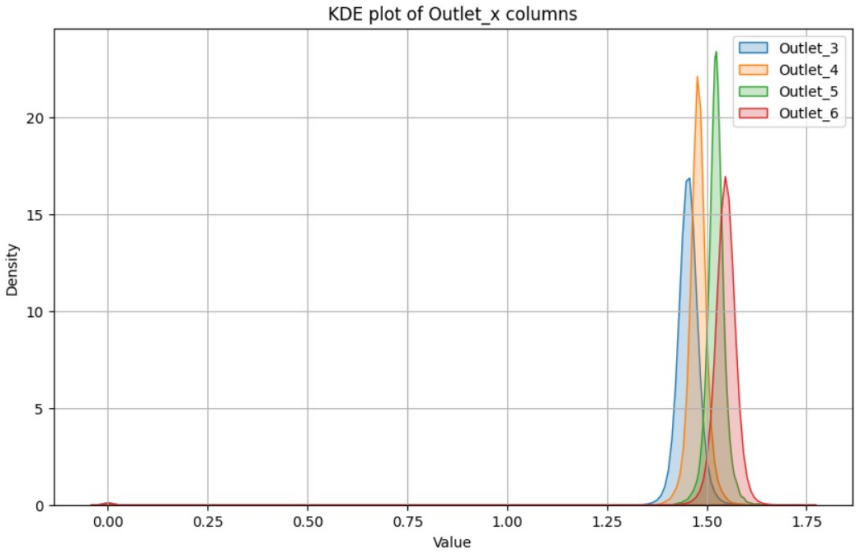

Figure 4.3: Distribution of outlet concentrations for non-diagonal configurations.

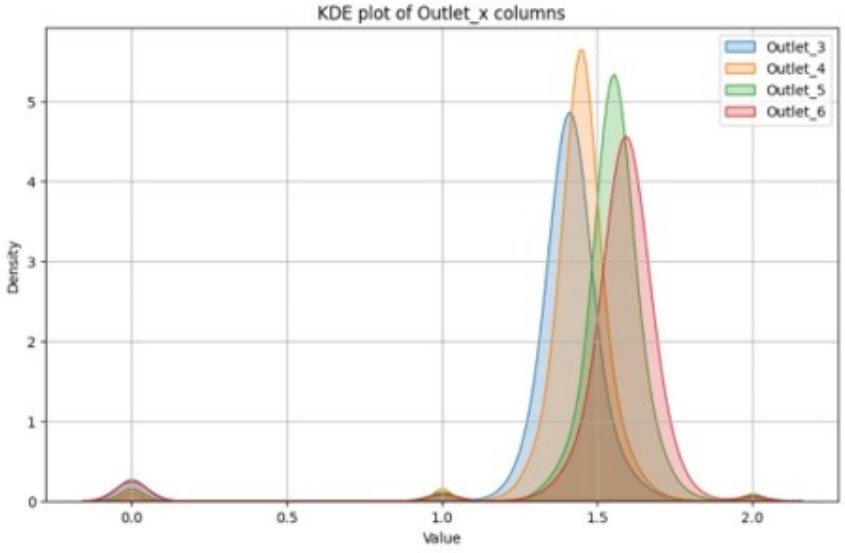

Figure 4.4: Distribution of outlet concentrations for diagonal configurations.

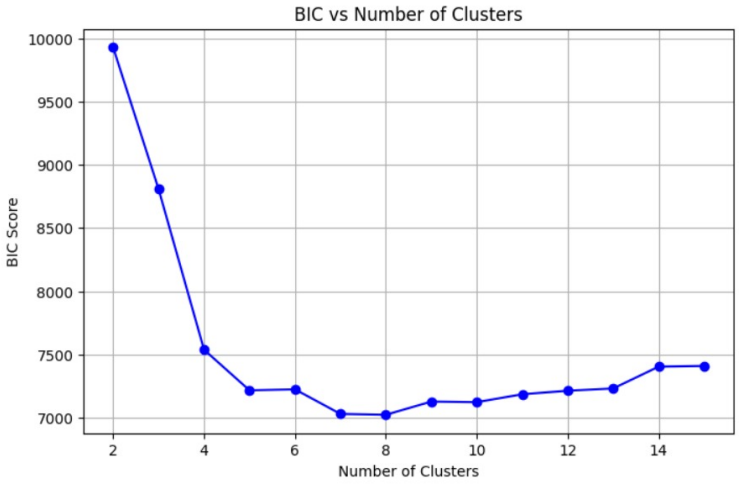

Figure 4.5: BIC score vs. number of clusters for non diagonal configurations.

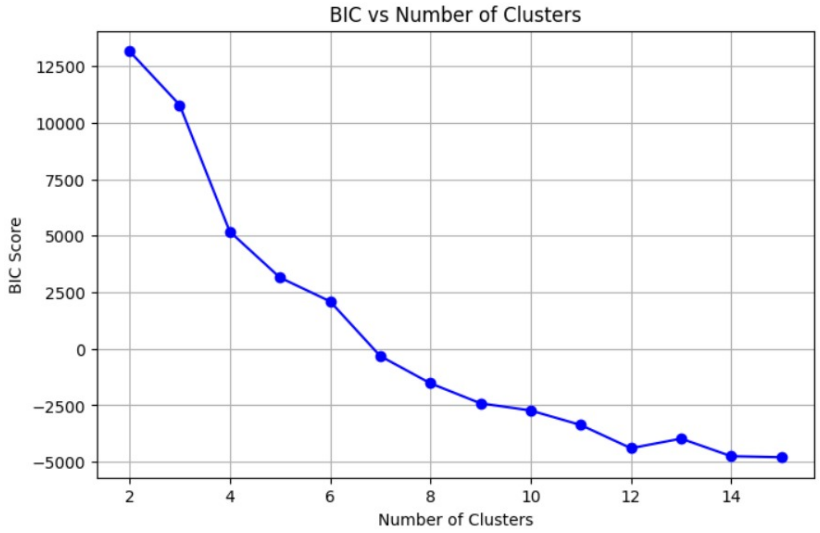

Figure 4.6: BIC score vs. number of clusters for diagonal configurations.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

This study proposed a machine learning-driven methodology for the optimal design and configuration of Random Microfluidic (RMF) chips with N×M architectures, incorporating diagonal channel topologies. Addressing the inherent limitations of conventional microfluidic systems, the methodology integrates COMSOL Multiphysics and MATLAB to facilitate high-throughput simulation, data generation, and analysis.

By applying advanced clustering techniques, specifically Gaussian Mixture Models (GMM), the study achieved notable improvements in clustering accuracy and configuration space reduction compared to traditional K-means clustering approaches. The GMM's capacity to model elliptical and overlapping clusters enabled a more accurate representation of the continuous and anisotropic nature of microfluidic concentration gradients.

The inclusion of diagonal channels within the chip design demonstrated significant benefits in terms of enhanced mixing efficiency, reduced channel lengths, and lower concentration prediction errors. These improvements collectively contribute to more compact, efficient, and adaptable microfluidic layouts, ultimately reducing fabrication costs and design iterations.

Furthermore, the methodology supports the generation of customized dilution gradients, including arithmetic progression profiles, thus expanding its applicability to diverse biomedical, pharmaceutical, and chemical assay applications.

In summary, this work highlights the transformative potential of integrating machine learning with microfluidic design workflows. The presented approach offers a scalable, intelligent, and automated solution for the development of precision microfluidic platforms, laying the groundwork for future advancements in lab-on-chip systems and personalized diagnostic technologies.

5.2 Future Work

While the proposed methodology demonstrated substantial improvements in the design and optimization of Random Microfluidic (RMF) chips, several areas remain open for further exploration and enhancement:

- **Integration of Additional Machine Learning Models:** Future studies could explore the application of other probabilistic and deep learning-based clustering techniques, such as Variational Bayesian GMM or autoencoder-based clustering, to potentially capture more complex spatial patterns in microfluidic concentration distributions.

- **Multi-Objective Optimization:** The current approach primarily focuses on clustering-based configuration space reduction. Future work can incorporate multi-objective optimization frameworks that simultaneously consider factors like pressure drop, mixing index, fabrication constraints, and reagent consumption.
- **Dynamic Reconfigurability:** Extending the methodology to support dynamically reconfigurable microfluidic chips with active control elements (e.g., valves, actuators) could enable real-time adjustment of dilution gradients and mixing patterns.
- **Topology-Aware Clustering:** Further refinement of clustering models that inherently account for the topological constraints and connectivity of microfluidic networks may yield configurations with better operational performance and robustness.

Addressing these directions in future work will further improve the scalability, versatility, and practical impact of machine learning-assisted microfluidic design frameworks for next-generation lab-on-chip systems.

References

[1] Wang, Junchao, et al. "Instantaneous simulation of fluids and particles in complex microfluidic devices." *PloS one* 12.12 (2017): e0189429.

[2] Ji, Weiqing, Tsung-Yi Ho, and Hailong Yao. "More effective randomly-designed microfluidics." 2018 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). IEEE, 2018.

[3] Ji, Weiqing, et al. "Microfluidic design for concentration gradient generation using artificial neural network." *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 39.10 (2019): 2544-2557.

[4] Ji, Weiqing, Tsung-Yi Ho, and Hailong Yao. "Transfer learning-based microfluidic design system for concentration generation." 2020 57th ACM/IEEE Design Automation Conference (DAC). IEEE, 2020.

[5] Zhang, Naiyin, Kaiyin Zha, and Junchao Wang. "Exploring the design efficiency of random microfluidic mixers." *IEEE Access* 9 (2021): 9864-9872.

[6] Wang, Junchao, et al. "Predicting the fluid behavior of random microfluidic mixers using convolutional neural networks." *Lab on a Chip* 21.2 (2021): 296-309.

[7] Ji, Weiqing, et al. "GNN-based concentration prediction for random microfluidic mixers." Proceedings of the 59th ACM/IEEE Design Automation Conference. 2022.

[8] Ji, Weiqing, et al. "GAT-based Concentration Prediction for Random Microfluidic Mixers with Multiple Input Flow Rates." Proceedings of the Great Lakes Symposium on VLSI 2023. 2023.

[9] Yu, Junnan, et al. "Performance prediction of random variable-width microfluidic chips by convolutional neural networks." *Microelectronics Journal* 133 (2023): 105716.

[10] Agrawal, Ankita, and Sudip Roy. "Random Microfluidic Chip Design with Diagonal Channels Using K-Means Clustering for Fluid Dilutions." 2024 IEEE Computer Society Annual Symposium on VLSI (ISVLSI). IEEE, 2024.

[11] Ji, Weiqing, et al. "GNN-based concentration prediction with variable input flow rates for microfluidic mixers." *IEEE Transactions on Biomedical Circuits and Systems* (2024).