A Network-Flow Based Optimal Sample Preparation Algorithm for Digital Microfluidic Biochips

Trung Anh Dinh¹  Shigeru Yamashita¹  Tsung-Yi Ho²

¹Ritsumeikan University, Japan
²National Cheng-Kung University, Taiwan
Agenda

Introduction

Problem Formulation

Proposed Method

Experimental Results

Conclusion
Agenda

Introduction

• Digital Microfluidic Biochips
• Sample Preparation
• Illustrative Example

Problem Formulation

Proposed Method

Experimental Results

Conclusion
Digital Microfluidic Biochips (DMFBs)

- Architecture of a DMFB:
  - 2D microfluidic array: Basic cells for biological reactions
  - Droplets: Biological samples (picoliter unit)
  - I/O ports, peripheral devices (detector, …)

- Applications: Immunoassay, DNA sequencing, protein crystallization, etc.
Sample Preparation (1/4)

- To produce droplets of the **required concentrations**
- A crucial preprocessing step in every application
  - 90% of the cost and 95% of the analysis time

![Diagram showing sample preparation process]

- Expensive cost !!! $$
- Processing time, reliability !!!

Sample preparation

- **Target droplet**
- **Sample droplets** \((c = 100\%)\)
- **Buffer droplets** \((c = 0\%)\)
- **Waste droplets**
Sample Preparation (2/4)

- To produce droplets of the required concentrations
- A crucial preprocessing step in every application
  - 90% of the cost and 95% of the analysis time

Sample preparation process:
- Sample droplets ($c = 100\%$)
- Buffer droplets ($c = 0\%$)
- Waste droplets

Multi-target processing:
- Optimization needed!!!
- Processing time, reliability !!!

Expensive cost !!! $$$
Sample Preparation (3/4)

• Dilution method in DMFBs
  • Samples are diluted using 1:1 mixing/splitting ratio
  • E.g., produce a droplet of concentration 25% ($\frac{1}{4}$)

IEEE/ACM ASP-DAC 2014
Sample Preparation (3/4)

- Dilution method in DMFBs
  - Samples are diluted using 1:1 mixing/splitting ratio
  - E.g., produce a droplet of concentration 25% ($\frac{1}{4}$)

![Diagram showing sample, buffer, and waste droplets with 1 dilution operation]
Sample Preparation (3/4)

• Dilution method in DMFBs
  • Samples are diluted using 1:1 mixing/splitting ratio
  • E.g., produce a droplet of concentration 25% ($\frac{1}{4}$)

1 sample droplet  1 waste droplet
2 buffer droplets  1 dilution operation
Sample Preparation (3/4)

- Dilution method in DMFBs
  - Samples are diluted using 1:1 mixing/splitting ratio
  - E.g., produce a droplet of concentration 25% ($\frac{1}{4}$)

1 sample droplet 1 waste droplet
2 buffer droplets 2 dilution operations
In general:

\[ \frac{C_i + C_j}{2} \]

A concentration value is always expressed by \( \frac{c_i}{2^d} \)

- \( d \): precision level of concentration
- \( d \) is given in each problem
Example: Same targets, but different cost

- \( d = 6 \)
- Targets: \( \frac{21}{64}, \frac{51}{64} \)

<table>
<thead>
<tr>
<th>#sample</th>
<th>#buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>32</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>64</td>
</tr>
<tr>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>42</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>51</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#sample</th>
<th>#buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td>24</td>
<td>0</td>
</tr>
<tr>
<td>38</td>
<td>64</td>
</tr>
<tr>
<td>12</td>
<td>64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#waste</th>
<th>#ops</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

\#sample \#buffer
7 7
32 0
16 64
40 0
42 0
21 51
38 64
12 64

\#waste \#ops
12 12
0 1
0 1
0 1
0 1
0 1
0 1
0 1
0 1
12 12
0 1
0 1
0 1
0 1
0 1
0 1
0 1
0 1
0 1
Problem Formulation (1/2)

• **Given:**
  - Cost of 1 sample droplet: $cost_s$
  - Cost of 1 buffer droplet: $cost_b$
  - Precision level of concentration: $d$
  - A set of $N$ target concentrations: $TC = \{c_1, c_2, \ldots, c_N\}$
  - A set of the required number of droplets for each target concentration: $S = \{s_1, s_2, \ldots, s_N\}$; $S_R = \sum_{i=1}^{N} s_i$

• **Objective:**
  Minimize cost of sample and buffer usage

  $F = u_s \times cost_s + u_b \times cost_b$

  $u_s, u_b$: The numbers of used sample/buffer droplets
Problem Formulation (2/2)

• Example:
  • $cost_s = 2$
  • $cost_b = 1$
  • $d = 3$
  • $TC = \{3, 5\}(\frac{3}{8} & \frac{5}{8})$
  • $S = \{4, 6\}; S_R = 10$

• Output: A valid sample preparation process

• Objective: Minimize cost function

$$F = u_s \times 2 + u_b \times 1$$
Previous Works

• All the previous works are based on heuristics
• All the previous works focus on only one objective optimization

Proposed method: Minimize the cost function

\[ F = u_s \times cost_s + u_b \times cost_b \]

• Optimal solution for multiple-target problem
• Flexible to change objective optimization
  • By varying the values of \( cost_s \) and \( cost_b \)
  • E.g., \( cost_s = 1 \) & \( cost_b = 0 \) \( \implies \) \( F = u_s \) (#sample droplets)
    (more on this later...)

IEEE/ACM ASP-DAC 2014
Agenda

- Introduction
- Problem Formulation
- Proposed Method
- Experimental Results
- Conclusion
Overview

Input

Min-cost max-flow (MCMF) network model construction

Integer equal flows problem transformation

ILP solver

Output
MCMF Network Model – Set of Vertices \( V \)

- **Source**, **Sink**, and **Waste** vertices represent different concentrations.
- Vertices are constructed bottom-up from level 0 to level 3.
- Level 0 contains vertices at level 0 that represent the set of concentration values that can be generated at this level.

Diagram:

- Level 0: Source and Waste vertices with concentration 0.
- Level 1: Source and Waste vertices with concentration 2.
- Level 2: Source and Waste vertices with concentration 4.
- Level 3: Source and Waste vertices with concentration 8.

Levels 4 to 7 are represented by Sink vertices with concentration 8.
Minimize

\[ F = f(\text{buffer}) \times \text{cost}_b + f(\text{sample}) \times \text{cost}_s \]
MCMF Network Model – Set of Arcs $A$

- Source
- Waste
- Sink

Levels:
- Level 0: Source
- Level 1: Waste
- Level 2: Intermediate nodes
- Level 3: Sink

Costs:
- $c_{s} = 2$
- $c_{b} = 1$
- $d = 3$
- $TC = \{3, 5\}$
- $S = \{4, 6\}$

Capacities:
- $cost = 0$
- $capacity = \infty$
MCMF Network Model – Set of Arcs $A$

Integer equal flows constraints: $f(a_i) = f(a_j)$

$cost_s = 2$
$cost_b = 1$
$d = 3$
$TC = \{3, 5\}$
$S = \{4, 6\}$
MCMF Network Model – Set of Arcs $A$

Target concentration constraints: $f(\text{target}, \text{Sink}) = s_i$

- $cost_s = 2$
- $cost_b = 1$
- $d = 3$
- $TC = \{3, 5\}$
- $S = \{4, 6\}$
MCMF Network Model – Set of Arcs $A$

- **Source**
- **Sink**
- **Waste**

Costs:
- $cost_s = 2$
- $cost_b = 1$
- $d = 3$
- $TC = \{3, 5\}$
- $S = \{4, 6\}$

Capacity: $\infty$
ILP Model

• Minimize

\[ F = f(\text{Source, sample}) \times \text{cost}_s + f(\text{Source, buffer}) \times \text{cost}_b \]

• Subject to

  • Capacity constraints:
    \[ f(v_x, v_y) \leq \text{capacity}(v_x, v_y) \quad \forall (v_x, v_y) \in A \]

  • Network-flow conservation:
    \[ \sum_{v_i: (v_i, v_x) \in A} f(v_i, v_x) = \sum_{v_o: (v_x, v_o) \in A} f(v_x, v_o) \]

• Integer equal flow constraints
• Target concentrations constraints
Comparative Studies

• Single-target sample preparation problem
  • [W. Thies et al., Natural Computing’08] [BS]
  • [S. Roy et al., IEEE/ACM DATE’11] [DMRW]
  • [J.-D Huang et al., IEEE/ACM ICCAD’12] [REMIA]

• Multiple-target sample preparation problem
  • [J.-D Huang et al., IEEE/ACM ICCAD’12] [REMIA]
Parameters Settings

• Cost Function

\[ F = u_s \times \text{cost}_s + u_b \times \text{cost}_b \]

• Waste droplets: \[u_s + u_b - S_R\]
  • \( S_R \): The total number of target droplets

<table>
<thead>
<tr>
<th>( \text{cost}_s )</th>
<th>( \text{cost}_b )</th>
<th>( F )</th>
<th>Optimization Objective</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( u_s )</td>
<td>#sample droplets (ours(_S))</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>( u_s + u_b )</td>
<td>#waste droplets (ours(_W))</td>
</tr>
</tbody>
</table>
Single-Target Sample Preparation

• ILP solver: CPLEX

• $d = 10$

• Target concentrations: $\frac{1}{1024} \rightarrow \frac{1023}{1024}$
  • Take average values of all 1023 cases

<table>
<thead>
<tr>
<th></th>
<th>BS</th>
<th>DMRW</th>
<th>REMIA</th>
<th>ours$_S$</th>
<th>ours$_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td># sample droplets</td>
<td>5.00</td>
<td>3.52</td>
<td>2.41</td>
<td>2.22</td>
<td>2.49</td>
</tr>
<tr>
<td># buffer droplets</td>
<td>4.01</td>
<td>3.50</td>
<td>6.09</td>
<td>9.68</td>
<td>2.50</td>
</tr>
<tr>
<td># waste droplets</td>
<td>8.01</td>
<td>6.02</td>
<td>7.50</td>
<td>10.90</td>
<td>3.99</td>
</tr>
<tr>
<td># operations</td>
<td>8.01</td>
<td>12.52</td>
<td>10.13</td>
<td>15.80</td>
<td>9.85</td>
</tr>
</tbody>
</table>
Multiple-Target Sample Preparation

- $d = 8, 9, 10$  \hspace{1em} $N = 10, 20, 50, 100$
- For each pair $(d, N)$, generate 100 random test cases

<table>
<thead>
<tr>
<th>$d = 9$</th>
<th>$N = 10$</th>
<th>$N = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>REMIA</td>
<td>ours$_s$</td>
</tr>
<tr>
<td># sample droplets</td>
<td>19.59</td>
<td>8.07</td>
</tr>
<tr>
<td># buffer droplets</td>
<td>31.19</td>
<td>12.67</td>
</tr>
<tr>
<td># waste droplets</td>
<td>40.78</td>
<td>10.74</td>
</tr>
<tr>
<td># operations</td>
<td>60.90</td>
<td>43.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$d = 10$</th>
<th>$N = 10$</th>
<th>$N = 100$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>REMIA</td>
<td>ours$_s$</td>
</tr>
<tr>
<td># sample droplets</td>
<td>17.33</td>
<td>11.17</td>
</tr>
<tr>
<td># buffer droplets</td>
<td>43.35</td>
<td>19.79</td>
</tr>
<tr>
<td># waste droplets</td>
<td>50.68</td>
<td>20.96</td>
</tr>
<tr>
<td># operations</td>
<td>89.77</td>
<td>73.12</td>
</tr>
</tbody>
</table>
Conclusion

• Sample preparation
  • Pivotal role in every assay, laboratory, and application in biomedical engineering and life science

• The first optimal sample preparation algorithm is proposed
  • Based on a minimum-cost maximum-flow model
  • Reduce the numbers of sample-buffer/waste droplets & dilution operations significantly
    (~60%/70%/85% & 60%, respectively)

Thank you for your attention!!!
MCMF Network Model – Set of Arcs $A$

cost = 0  
cost$_s$ = 2

capacity = $\infty$  
cost$_b$ = 1

cost = 0  
d = 3

capacity = $s_2$ = 6  
$TC = \{3, 5\}$

Source

cost = cost$_b$ = 1

capacity = $\infty$

cost = cost$_s$ = 2

capacity = $\infty$

Sink

cost = 0

capacity = $s_2$ = 6

Waste

cost = 0

capacity = $\infty$
\( d = 3 \)

\[ \text{Waste} \]

\[ \text{Sink} \]

\[ \text{Source} \]

Level 3

Level 2

Level 1

Level 0
Problem Formulation

- Hybrid cost function:
  \[ F = u_s \times \text{cost}_s + u_b \times \text{cost}_b \]
- The number of waste droplets: \( u_s + u_b - S_R \)

<table>
<thead>
<tr>
<th>( \text{cost}_s )</th>
<th>( \text{cost}_b )</th>
<th>( F )</th>
<th>Minimization</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>( u_s )</td>
<td>#sample droplets</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>( u_s + u_b )</td>
<td>#waste droplets</td>
</tr>
<tr>
<td>practical values</td>
<td>( u_s \times \text{cost}_s + u_b \times \text{cost}_b )</td>
<td>practical cost</td>
<td></td>
</tr>
</tbody>
</table>

- The cost function \( F \) is flexible
Experiment Environment

- Implemented by C++
- ILP Solver: CPLEX
- Linux server
  - Intel® Core(TM) i7 CPU 920 2.67GHz
  - 24 GB Memory
- Largest test case \((d = 10, N = 100)\)
  - 352,608 variables
  - 175,374 constraints
  - Computation time: ~30 minutes
Digital Microfluidic Biochips (DMFBs) (2/2)

• Advantages:
  • High portability
  • High throughput
  • Low sample volume consumption
  • Less human intervention errors

• Applications: immunoassay, DNA sequencing, protein crystallization, etc.