NEURAL INFORMATION PROCESSING SYSTEMS

Pairwise Optimal Transports for Training All-to-All Flow-Based Condition Transfer Model

Kotaro Ikeda[*†] · Masanori Koyama[*†] · Jinzhe Zhang[†]

Kohei Hayashi[*†] · Kenji Fukumizu[‡†]

* The University of Tokyo, †Preferred Networks Inc., ‡The Institute of Statistical Mathematics

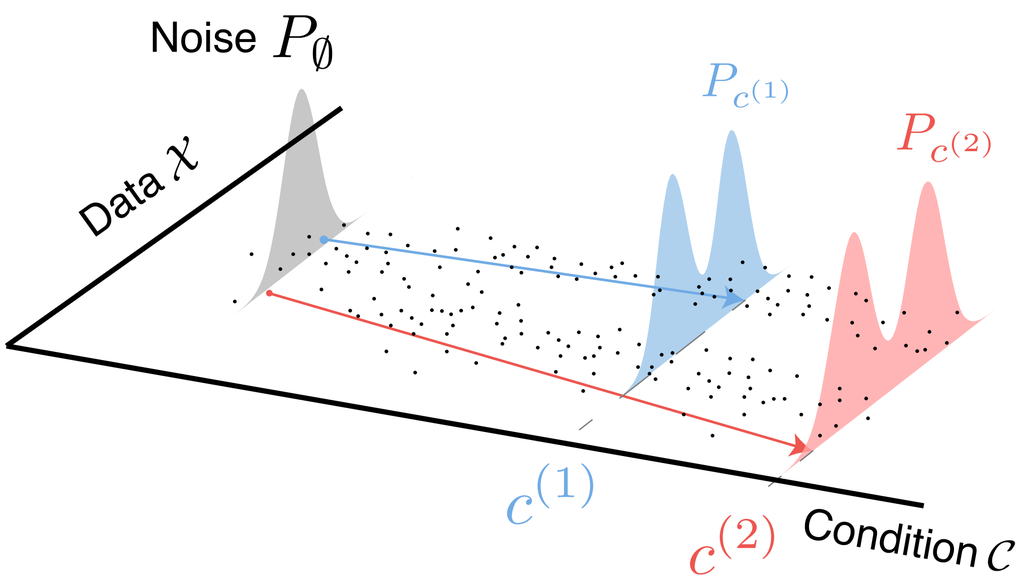
UTokyo

Preferred Networks

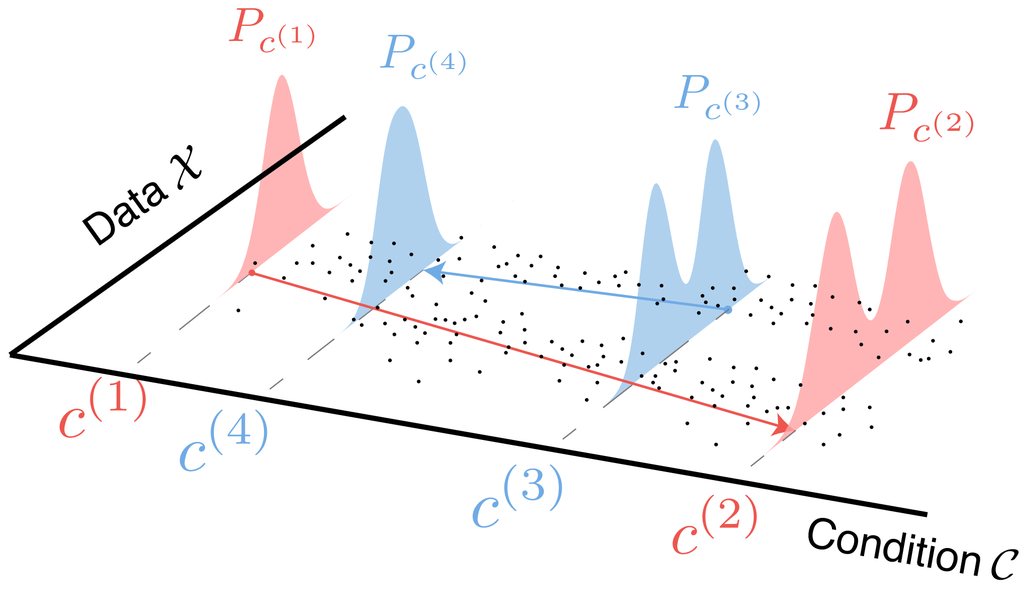
大学共同利用機関法人 情報・システム研究機構 統計数理研究所

Condition Transfer

Conditional Generation



Condition Transfer



Conditional generation: generates samples from a **single** noise distribution

Condition Transfer: transfer samples from **multiple** source conditional distributions

Examples of Condition Transfer

Examples of condition transfer includes…

- Image editing
  - ➡ Reference image + Prompt → Modified image

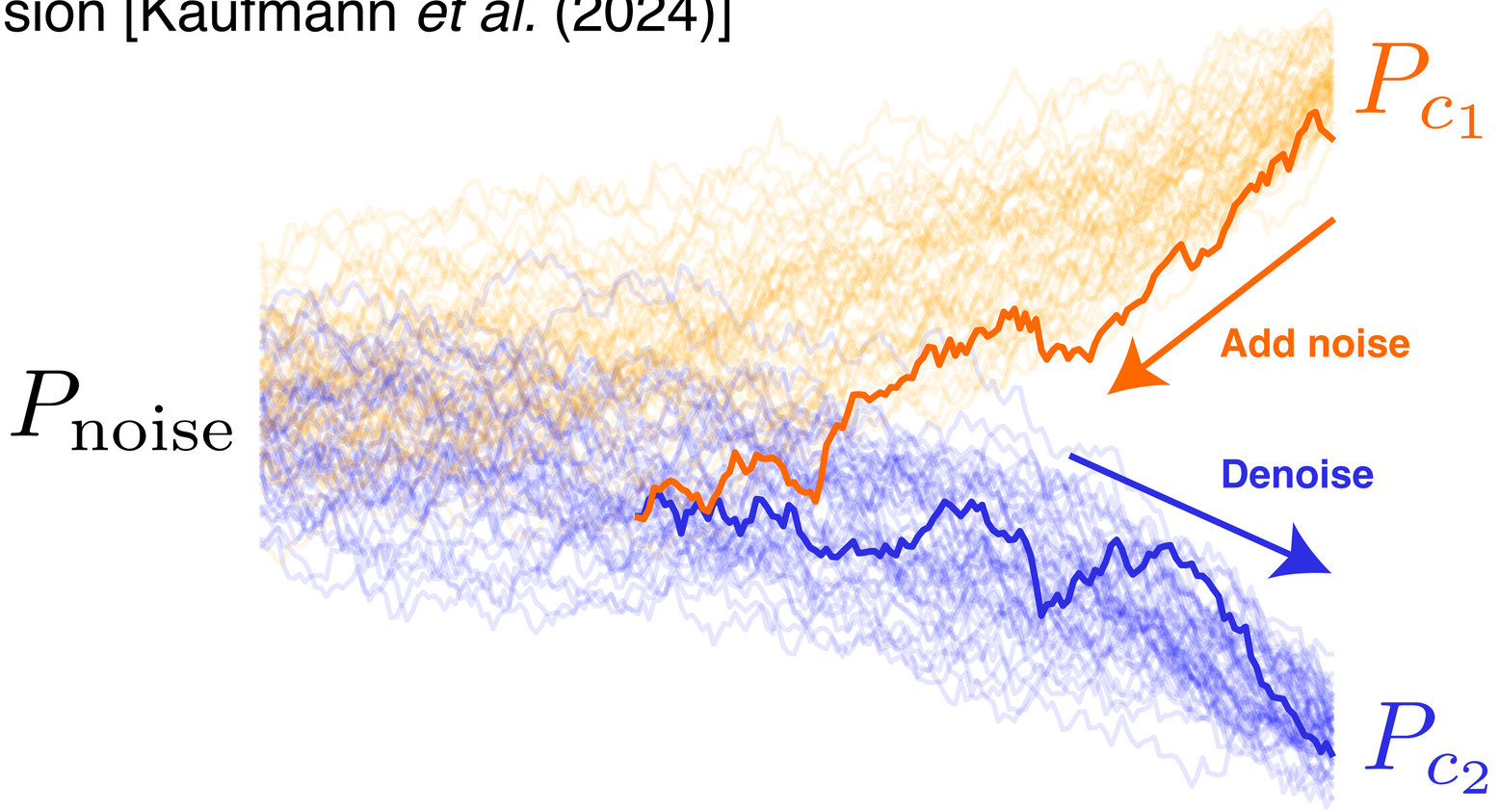
- Molecule optimization
  - ➡ Reference molecule + Physical property → Optimized molecule

- Many other editing tasks……

✅ Condition Transfer is essential when the similarity with reference data is necessary

Prior condition transfer method: Partial diffusion

Partial Diffusion [Kaufmann *et al.* (2024)]

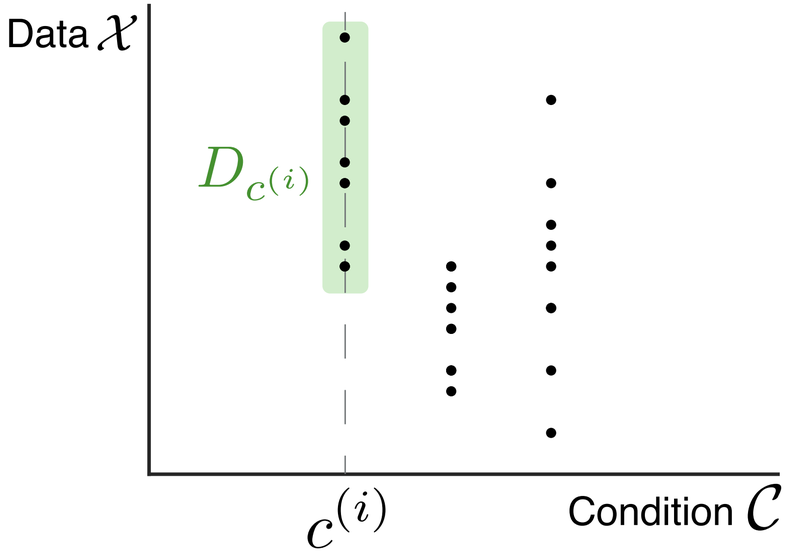


❌ Partial Diffusion does not guarantee similarity between source and target samples

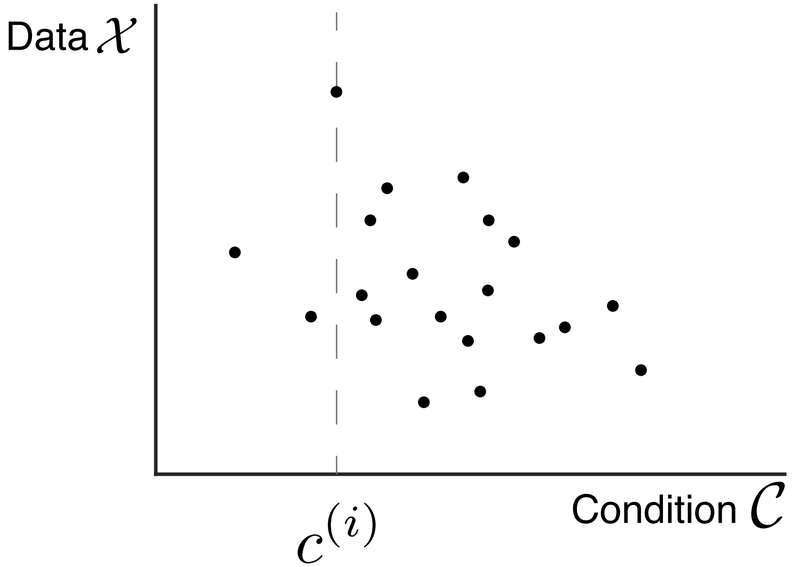
✅ Proposed Method has a theoretical guarantee in similarity through optimal transport (OT)

Grouped and Non-grouped Data

Grouped data
e.g. Image dataset with labels



Non-grouped data
e.g. Molecule dataset with physical properties



Applicability

Multimariginal SI / EFM

A2A-FM (Ours)

❌ Multimarginal SI and EFM leverages OT but limited to grouped settings

✅ Proposed method is applicable to non-grouped settings (even continuous labels!)

All-to-All Flow-based transfer model (A2AFM)

Key Idea: Learn **pairwise optimal transport (OT)** through **Conditional OT flow matching**

Conditional OT flow matching (COT) [Chemseddine et al. (2024), Kerrigan et al. (2024)]

- A conditional generation method that provably converges to conditional OT

Idea of COT: Do OT flow matching in $\mathcal{X} \times \mathcal{C}$ with the following cost function

$$\mathcal{C}(\pi) = \sum_{i=1}^{N} \left\| x_1^{(i)} - x_{\text{noise}}^{\pi(i)} \right\|^2 + \beta \left\| c_1^{(i)} - c_{\text{noise}}^{\pi(i)} \right\|^2$$

$\mathcal{X}$: Data Space

$\mathcal{C}$: Condition Space

π : Permutation in 1, …, N

Theorem (informal)

The OT map of $C(\pi)$ *converges to OT between* $P_{\text{noise}} \to P_c$ *for every* c *as* $\beta \to \infty$ *&* $N \to \infty$

All-to-All Flow-based transfer model (A2AFM)

Key Idea: Learn **pairwise optimal transport (OT)** through **Conditional OT flow matching**

Pairwise OT: Optimal transport between a pair of conditional distributions

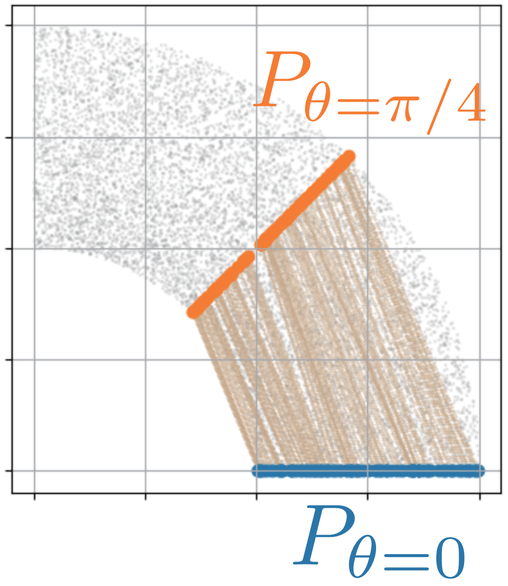
Pairwise OT cost for $P_{c_1} \rightarrow P_{c_2}$

$$\sum_{i=1}^{N} \| x_{c_2}^{\pi(i)} - x_{c_1}^{(i)} \|^2$$

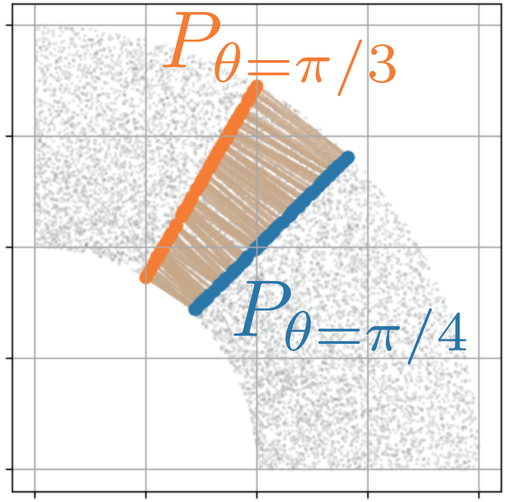
Where, $x_{c_1} \sim P_{c_1}, x_{c_2} \sim P_{c_2}$

π : Permutation in 1, …, N

$\theta = \frac{\pi}{4} \rightarrow 0$



$\theta = \frac{\pi}{3} \rightarrow \frac{\pi}{4}$



All-to-All Flow-based transfer model (A2AFM)

Key Idea: Learn **pairwise optimal transport (OT)** through **Conditional OT flow matching**

Idea of A2AFM: Do OT flow matching in $\mathcal{X} \times \mathcal{C} \times \mathcal{C}$ with the following cost function

$$C(\pi) = \sum_{i=1}^{N} \|x_1^{(i)} - x_2^{\pi(i)}\|^2 + \beta \left(\|c_1^{(i)} - c_1^{\pi(i)}\|^2 + \|c_2^{(i)} - c_2^{\pi(i)}\|^2 \right)$$

Algorithm 1 Training of A2A-FM

Input: (i) Dataset of sample-condition pairs $D := \{(x^{(i)}, c^{(i)})\}$, where each $x^{(i)} \in \mathcal{X}$ is sampled from $P_{c^{(i)}}$. (ii) A parametric model of a vector field $v_\theta : \mathcal{X} \times [0,1] \times \mathcal{C} \times \mathcal{C} \to \mathcal{X}$. (iii) The scalar parameter β. (iv) An algorithm OPTC for optimal coupling.

Return: The parameter θ of v_θ

1: **for** each iteration **do**

`# Step 1: Compute the coupling`

2: Subsample batches $B_1 = \{(x_1^{(i)}, c_1^{(i)})\}_{i=1}^N, B_2 = \{(x_2^{(i)}, c_2^{(i)})\}_{i=1}^N$ from D.

3: Minimize (11) about π_β^* over N indices by OPTC

`# Step 2: Update` v_θ

4: Sample $t_i \sim \mathbf{unif}[0,1], i \in [1:N]$.

5: Update θ by $\nabla_\theta L(\theta)$ with

$L(\theta) = \sum_{i=1}^N \|v_\theta(\psi_i(t_i), t_i | c_1^{(i)}, c_2^{\pi_\beta^*(i)}) - \dot{\psi}_i(t_i)\|^2$ where $\psi_i(t) = (1-t)x_1^{(i)} + t x_2^{\pi_\beta^*(i)}$.

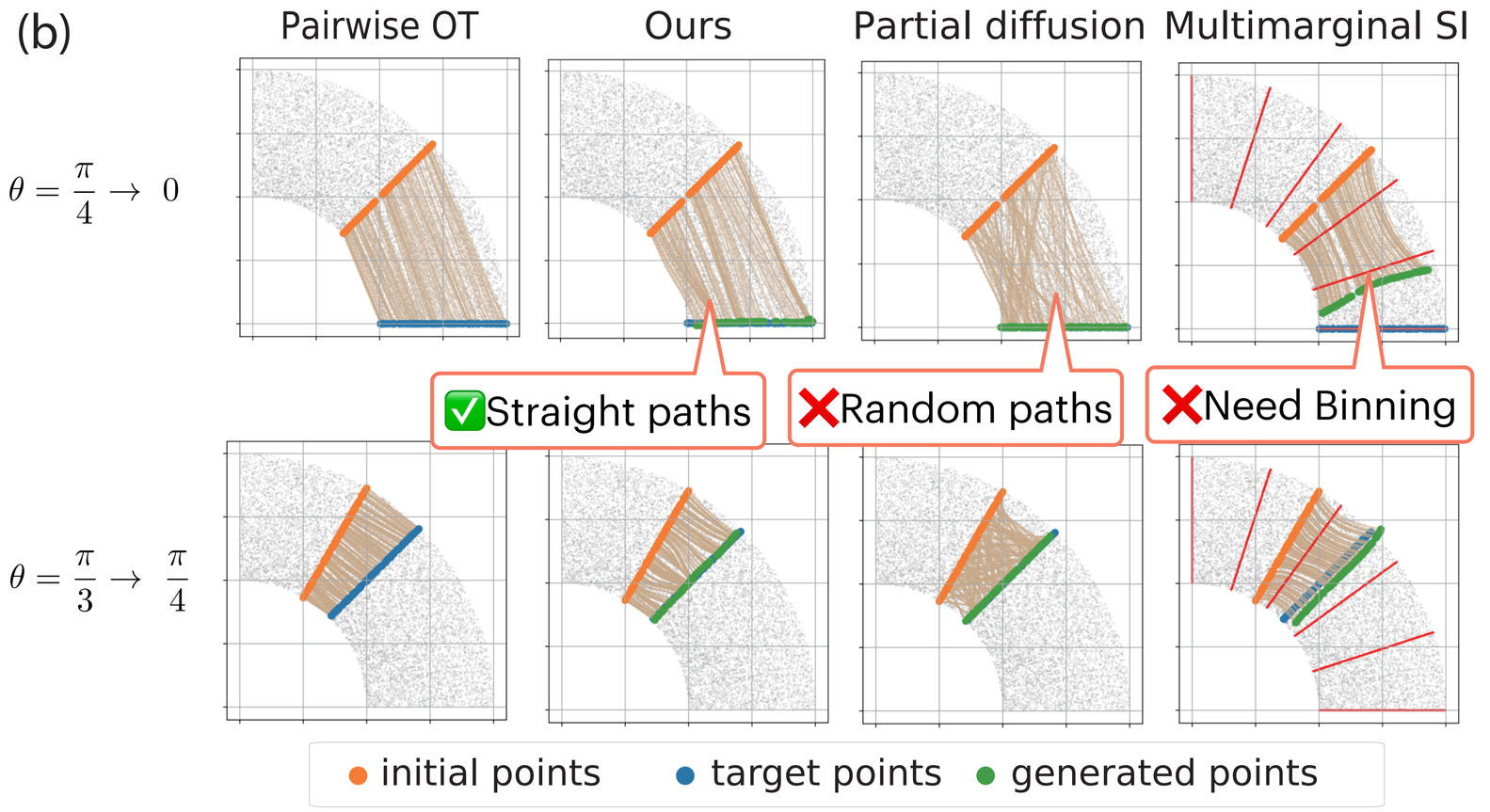
6: **end for**

Theorem (informal)

The OT map of $C(\pi)$ *converges to OT between* $P_{c_1} \to P_{c_2}$ *for every* (c_1, c_2) *as* $\beta \to \infty$ *&* $N \to \infty$

✅ Achieves pairwise OT at $\beta \to \infty$

Synthetic Experiments

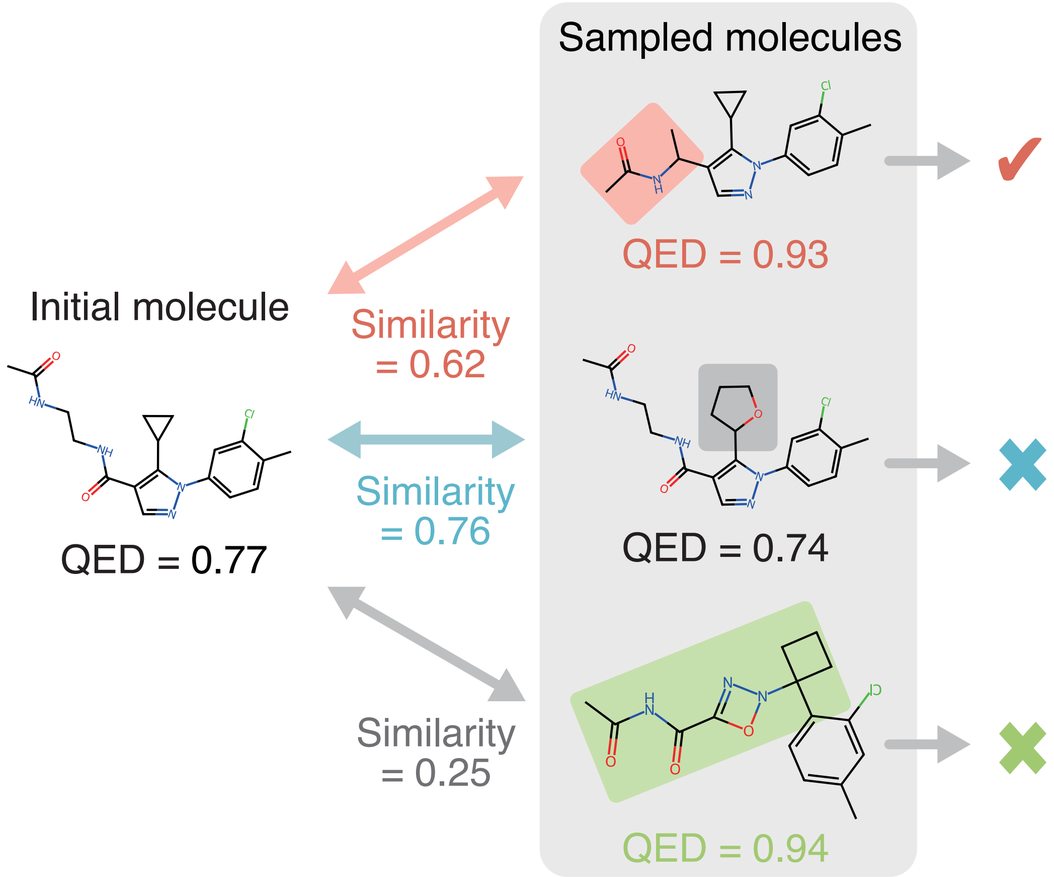


A2AFM succeeds to mimic pairwise OT paths compared to other methods

Application to molecule optimization

QED experiment

- Edit molecules to higher QED

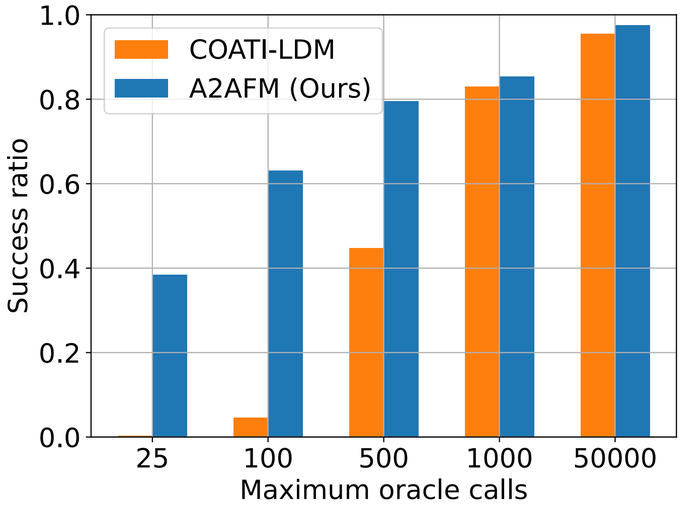


Results

Method	Success (%)
DESMILES [34]	76.9
QMO [19]	92.8
MolMIM [38]	94.6
COATI-LDM [25]	95.6
A2A-FM (Ours)	**97.5**

Table 2: Nearby sampling success rate.

✅ SoTA performance



✅ Efficient sampling

Conclusion

✓ Proposed A2AFM, a method for condition transfer even in **continuous condition** case

✓ A2AFM provably converges to **pairwise OT**

✓ A2AFM achieves **SoTA performance** at QED optimization task

Code: https://github.com/kotatumuri-room/A2A-FM

arXiv: https://arxiv.org/abs/2504.03188

Thanks

Kotaro Ikeda

Department of Mathematical Engineering & Information Physics

The University of Tokyo

kotaro-ikeda@g.ecc.u-tokyo.ac.jp

https://kotatumuri-room.github.io/

I am planning to apply for Ph.D starting at Fall 2026. Feel free to contact!