Exhibit Hall C,D,E
Wed. 3 Dec. 4:30 P.M.

ReDi: Rectified Discrete Flow

Jaehoon Yoo[1] Wonjung Kim[1] Seunghoon Hong[1]

[1]KAIST

Discrete Flow-based Models (DFMs)

Discrete Flow-based models learn $p_\theta(X_1|X_0)$, where X_1 is data and X_0 is initial state.

Masked Generative Models[1]

Uniform Diffusion

Sampling step: 0/8

```
many water during of to water put without thought between necessary million now go
important from general potential this try family over try so building from so
your just general own a someone to always could example state to out been go I
which up after state make how other I oil we possible without shall have while
hot beautiful they faigfywirkx start man could you in thought when take friend
always should had of first me different than is see necessary be example
```

Absorbing Diffusion

Sampling step: 00/30

```
----  -----  ------  --  --  -----  ---  ------  ------  -------  ---------  -------  ---  --
--------  ----  -------  ---------  ----  ---  ------  ----  ---  --  --------  ----  --
----  ----  -------  ---  -  -------  --  ------  -----  -------  -----  --  ---  ----  --  -
-----  --  -----  -----  ----  ---  -----  -  ---  --  --------  -------  -----  ----  -----
---  ---------  ----  -----------  -----  ---  -----  ---  --  ------  ----  ----  ------
------  ------  ---  --  -----  --  --------  ----  --  ---  ---------  --  ------
```

Diffusion Language Models[2]

[1] https://masked-generative-image-transformer.github.io/
[2] https://s-sahoo.com/duo/

What hinders DFM's few-step generation?

DFMs are based on the **factorized modeling** to treat high-dimensional data.

$$p(X_{t+h}|X_t) = \prod_{i=1}^{N} p(X_{t+h}^{(i)}|X_t) + o(h)$$

Joint Modeling

→ D^N logits

Factorized Modeling

→ D x N logits

What hinders DFM's few-step generation?

The necessity of **factorized modeling** introduces **factorization error** that hinders few-step generation of DFMs.

$$p(X_1|X_0) \neq \prod_{i=1}^{N} p(X_1^{(i)}|X_0)$$

Joint Modeling

Factorized Modeling

What hinders DFM’s few-step generation?

We adopt **conditional Total Correlation(TC)** to characterize the factorization error.

$$TC_\pi(X_1|X_0) = \mathbb{E}_{x_0}[D_{KL}(p(X_1|X_0 = x_0)||\prod_{i=1}^{N} p(X_1^{(i)}|X_0 = x_0))]$$

What hinders DFM's few-step generation?

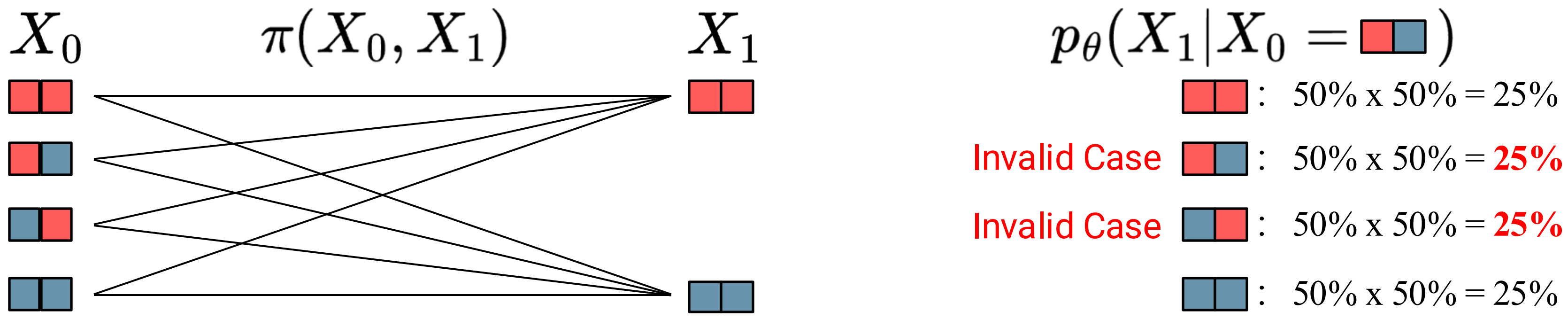
We observe that the conditional TC depends on the coupling $\pi(X_0, X_1)$.

$$TC_\pi(X_1|X_0) = \mathbb{E}_{x_0}[D_{KL}(p(X_1|X_0 = x_0)||\prod^{N} p(X_1^{(i)}|X_0 = x_0))]$$

$$= \frac{\pi(X_0 = x_0, X_1)}{p(X_0 = x_0)}$$

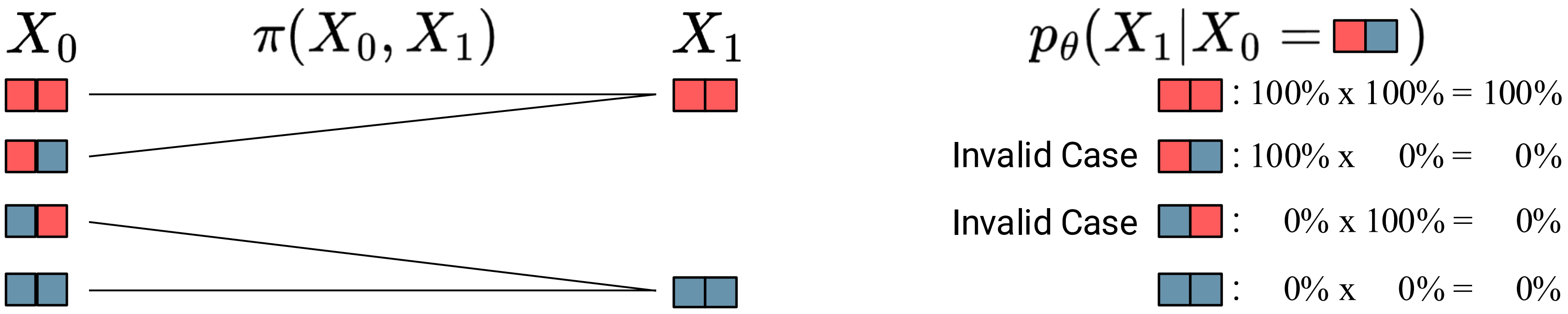
What hinders DFM's few-step generation?

We observe that the conditional TC depends on the coupling $\pi(X_0, X_1)$.



What hinders DFM's few-step generation?

We observe that the conditional TC depends on the coupling $\pi(X_0, X_1)$.



ReDi: Rectified Discrete Flow

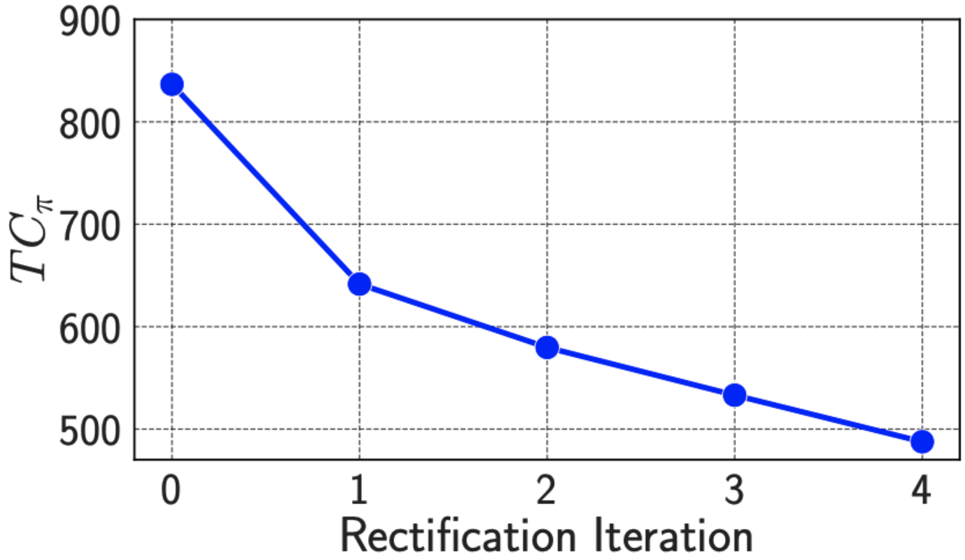
We propose to **rectify the coupling** to reduce the conditional Total Correlation.

$$\pi_{k+1}(X_0, X_1) = p(X_0)\underbrace{p_\theta(X_1|X_0)}_{\text{Data Generation}}$$

ReDi: Rectified Discrete Flow

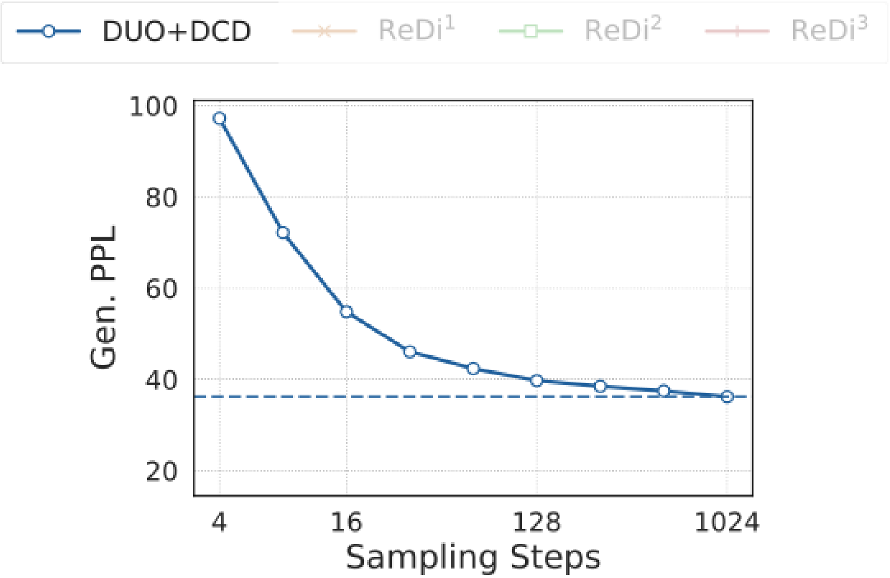
Theoretical Result: each rectification iteration reduces conditional Total Correlation.

$$TC_{k+1}(X_1|X_0) \leq TC_k(X_1|X_0)$$



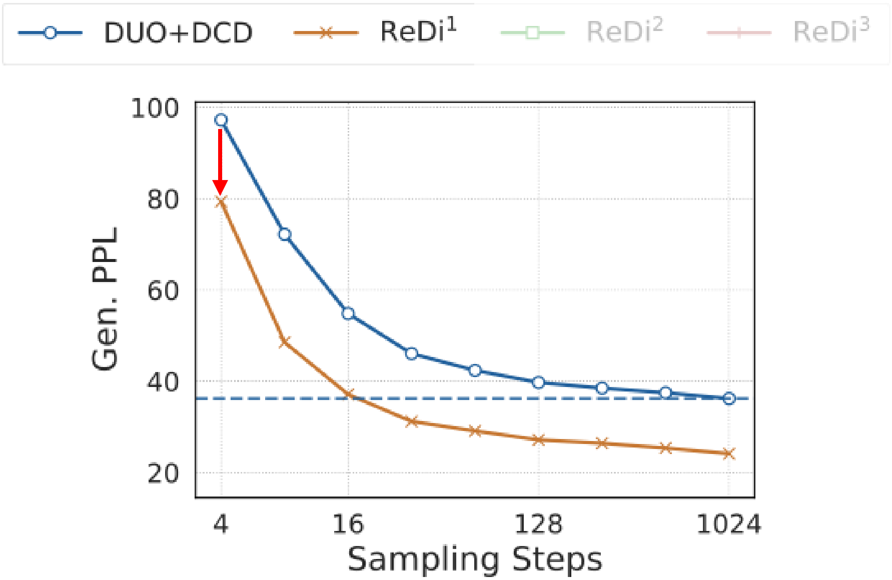
ReDi on Text Generation

ReDi enhances few-step performance of the teacher models on OpenWebText.



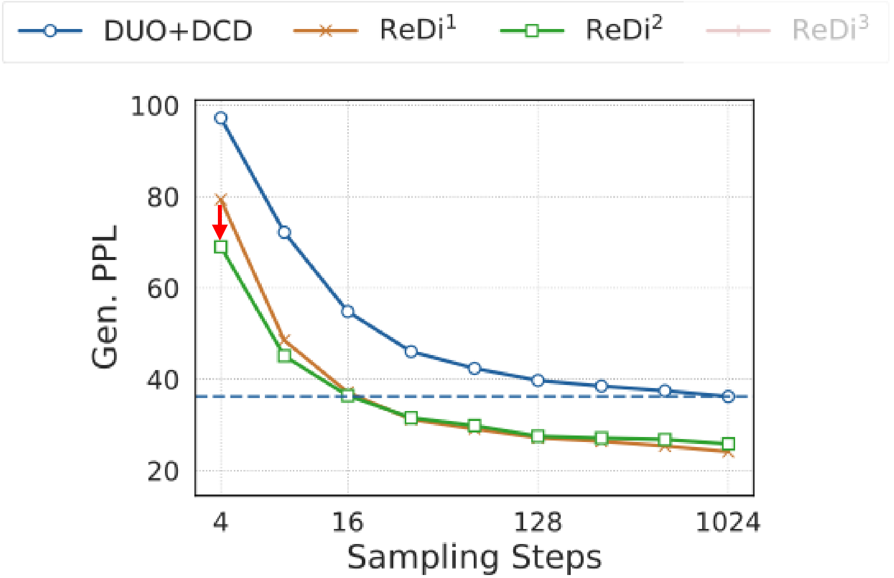
ReDi on Text Generation

ReDi enhances few-step performance of the teacher models on OpenWebText.



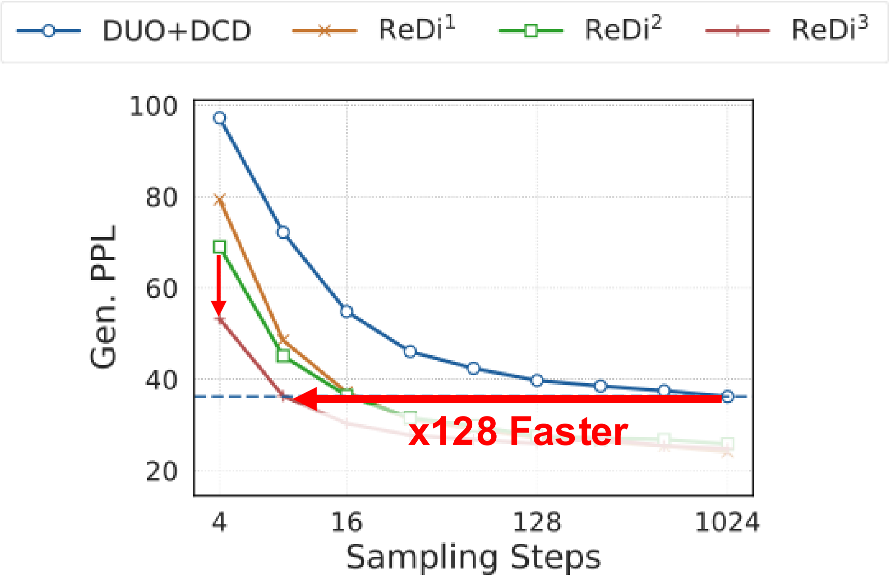
ReDi on Text Generation

ReDi enhances few-step performance of the teacher models on OpenWebText.



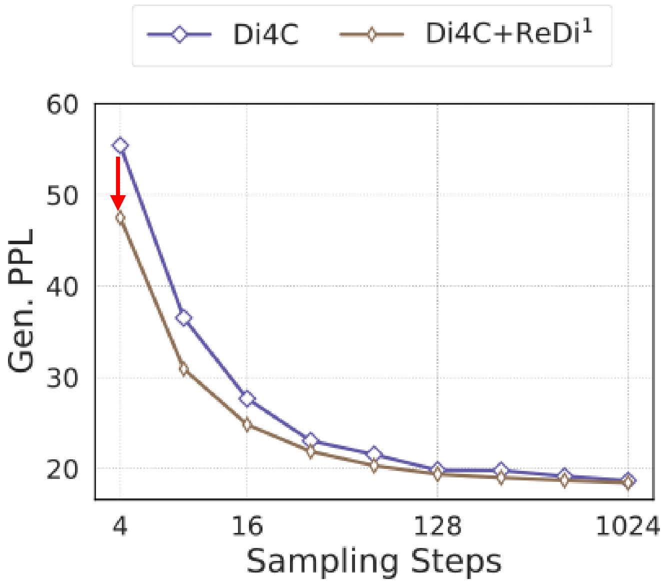
ReDi on Text Generation

ReDi accelerates the teacher model by up to **128x.**



ReDi on Text Generation

ReDi can be **combined with other distillation** method.

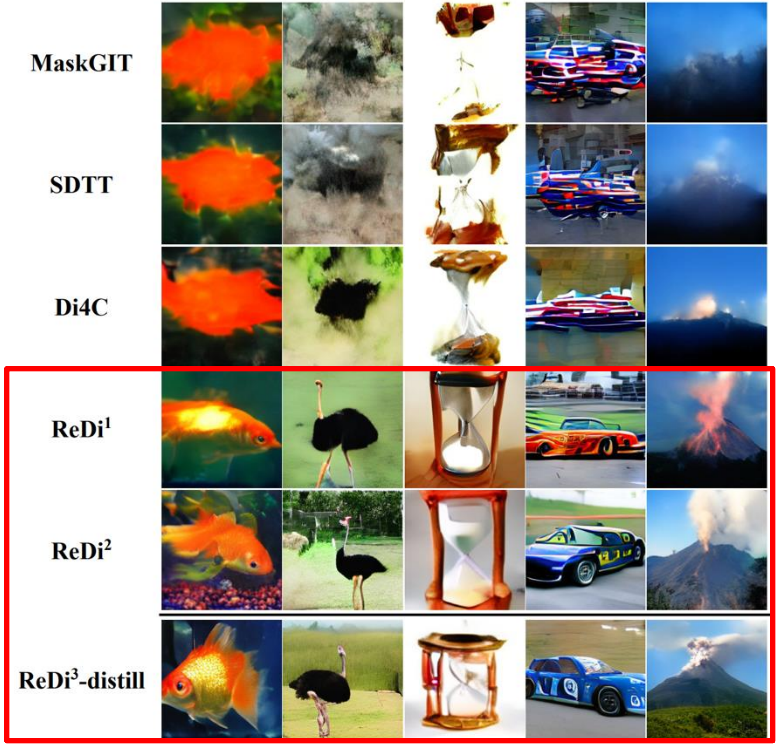


ReDi on Image Generation

ReDi enables **1-step generation** on class-conditional ImageNet.

Step	Model	FID (↓)	IS (↑)
1	MaskGIT [3]	95.16	12
	SDTT† [10]	90.40	14
	Di4C [17]	90.32	13
	ReDi^1	37.43	49
	ReDi^2	21.80	90
	ReDi^3-distill	**11.68**	**182**

1-step Generation Results



Conclusion

- **Factorization error** hinders few-step generation.
- Factorization error can be characterized with **conditional Total Correlation** which **depends on the coupling.**
- We proposed **ReDi, rectifying the coupling** to reduce the factorization error.

For more details, please join our poster session.

Exhibit Hall C,D,E (Wed. 3 Dec. 4:30 P.M.)

arXiv

code