ButterflyFlow: Building Invertible Layers with Butterfly Matrices

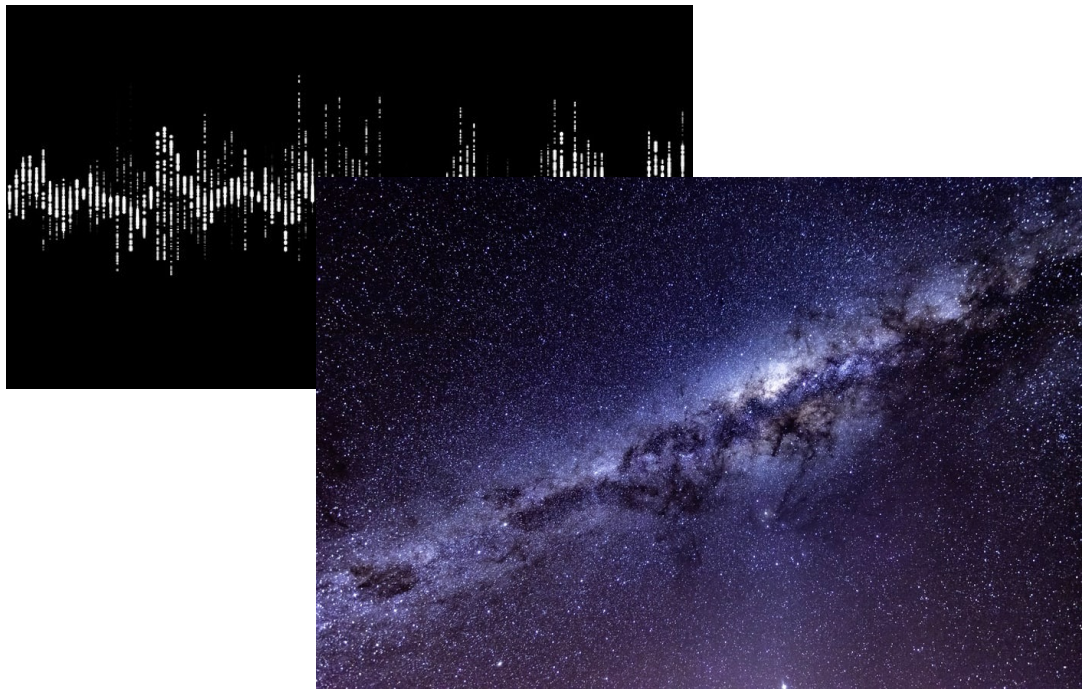
Chenlin Meng*, Linqi Zhou*, Kristy Choi*, Tri Dao, Stefano Ermon

Stanford University

ICML 2022

Motivation and problem setup

Despite the recent successes of generative models, they struggle to capture special structures commonly found in real-world data, such as **permutations and periodicity**.



[Brown et al. 2020, Ho et al. 2020, Ramesh et al. 2022]

Normalizing flows

A normalizing flow models the exact data likelihood via a series of K invertible transformations $\{f_i\}_{i=0}^{K-1}$

$$x = z_K \xleftarrow{f_{K-1}} z_{K-1} \xleftarrow{f_{K-2}} \cdots \xleftarrow{f_0} z_0$$

The data density evolves according to the change of variable formula:

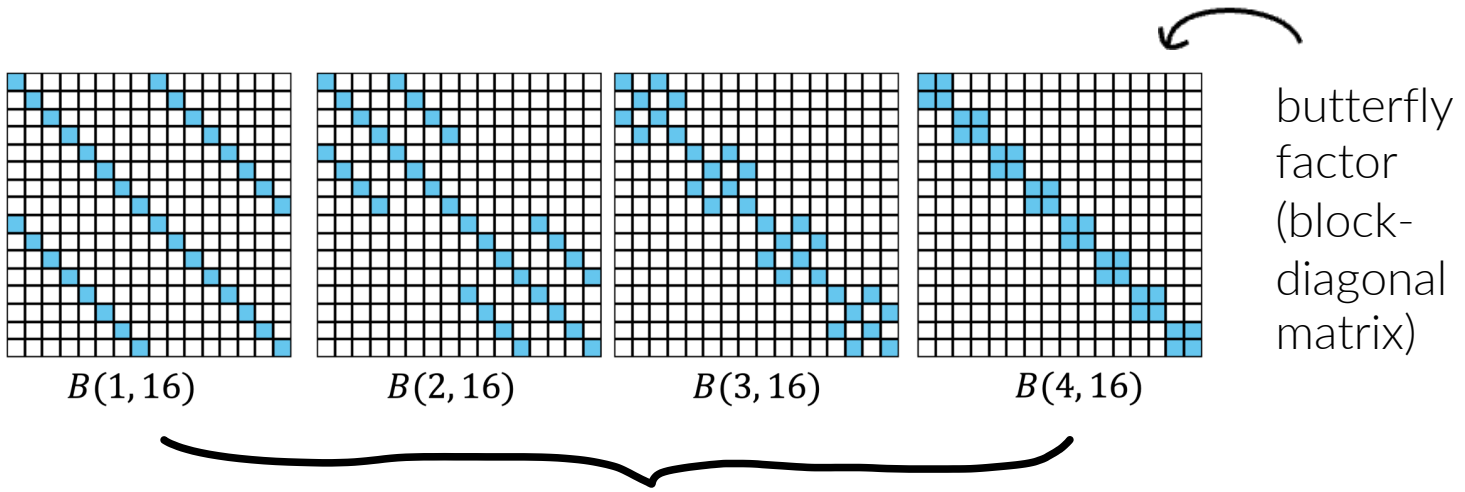
$$p(\mathbf{x}) = p(\mathbf{z}_0) \sum_{i=0}^{K-1} |\det J_{f_i}(\mathbf{z}_i)|$$

Our approach: invertible Butterfly layer

A butterfly layer f_i is a special family of linear layers that can be represented as a **product** of *K* butterfly factors.

Benefit: efficient computation of inverse and Jacobian determinant.

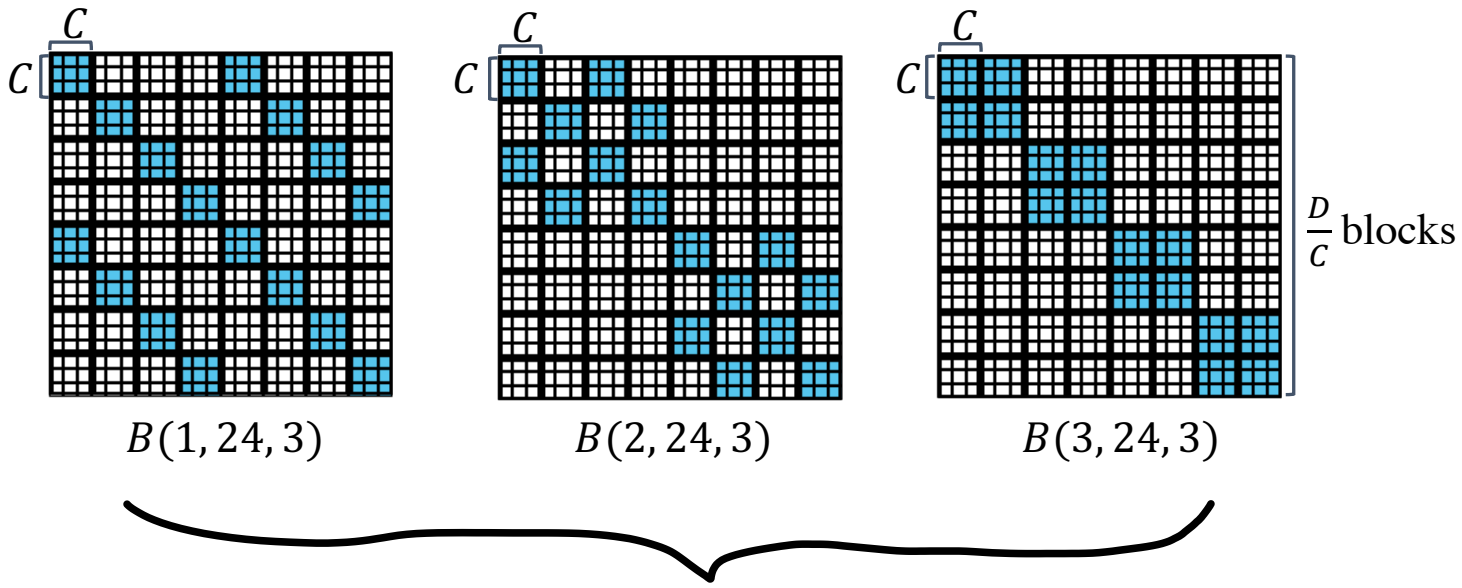
f_i =



[Dao et al. 2019]

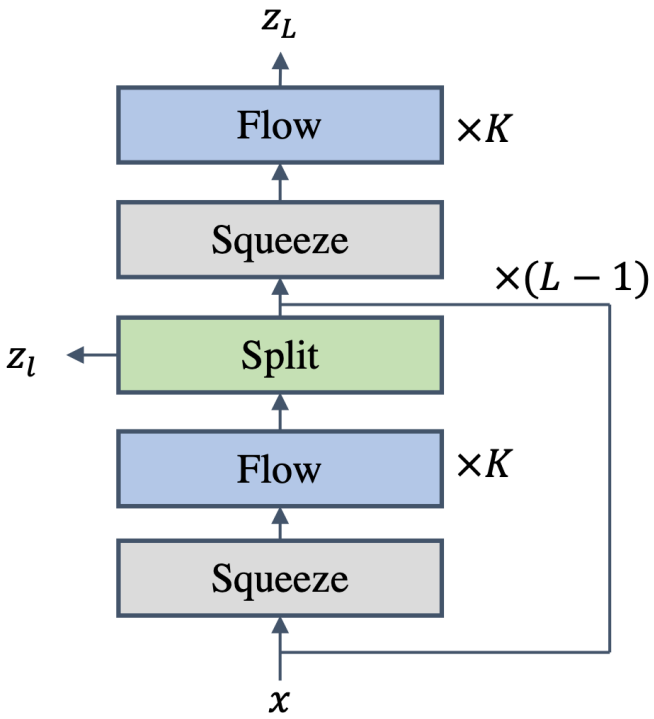
Block-wise Butterfly layers

Each primitive entry is a $C \times C$ block.
Trades off expressivity with computation speed.

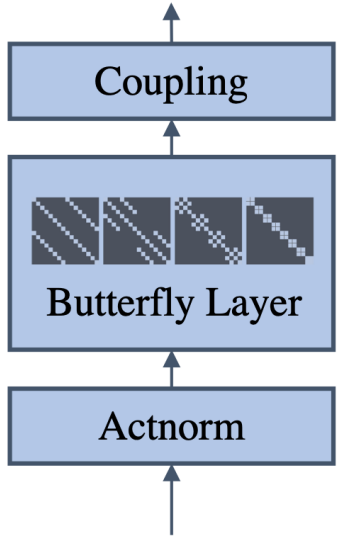


K butterfly factors with data dimension 24 and channel size 3

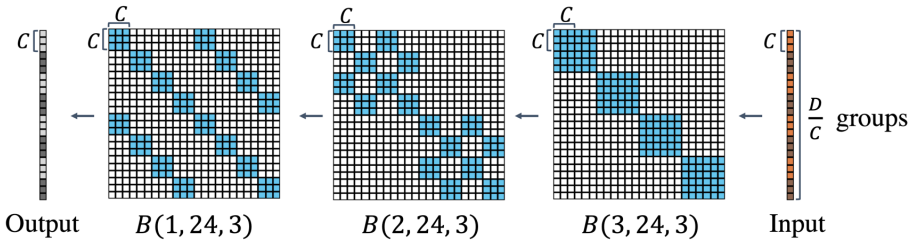
ButterflyFlow model



Model architecture



Flow module



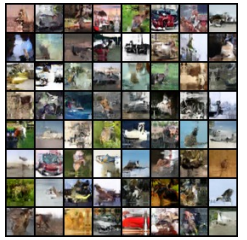
Block-wise butterfly layers

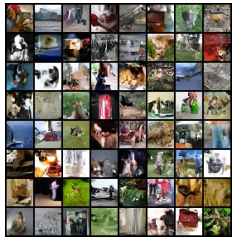
Glow-based backbone

[Dinh et al. 2014, Dinh et al. 2016, Kingma & Dhariwal 2018, Hoogeboom et al. 2019]

Experiments: permuted data


MNIST


CIFAR-10


ImageNet-32


Permuted image (input)


Glow


Emerging


Woodbury

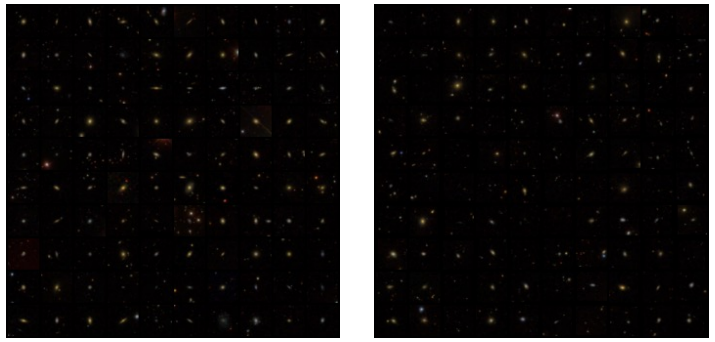

ButterflyFlow (Ours)

	MNIST	CIFAR-10	ImageNet 32×32
Glow (Kingma & Dhariwal, 2018)	1.44	5.48	6.29
Emerging (Hoogeboom et al., 2019)	1.43	5.41	6.25
Woodbury (Lu & Huang, 2020)	1.43	5.41	6.26
ButterflyFlow (Ours)	**1.42**	**5.11**	**6.18**

ButterflyFlow achieves strong density estimation results on **both permuted and original natural image datasets** (log-likelihoods in bits per dimension) relative to baselines.

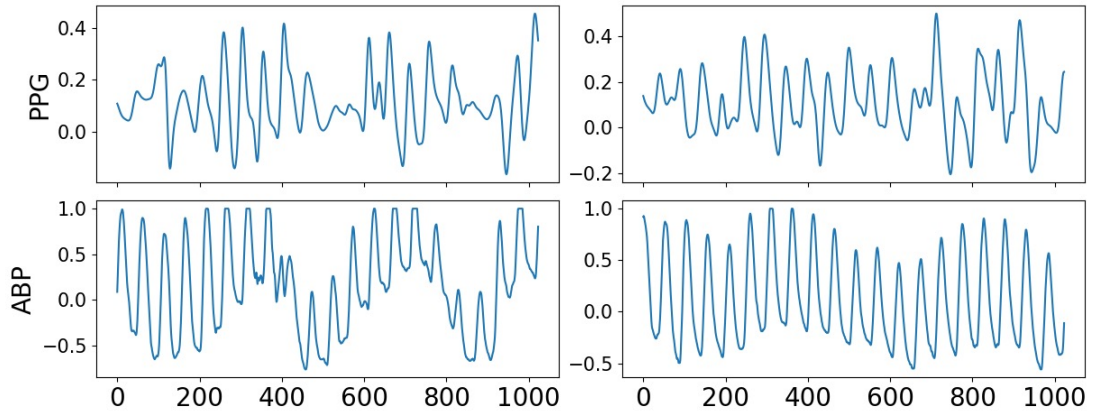
Experiments: periodic data

Galaxy dataset



	Galaxy
1×1 (Glow) (Kingma & Dhariwal, 2018)	2.02
Emerging 3×3 (Hoogeboom et al., 2019)	1.98
Periodic (Hoogeboom et al., 2019)	1.98
Woodbury (Lu & Huang, 2020)	2.01
ButterflyFlow (Ours)	**1.95**

MIMIC-III patient data

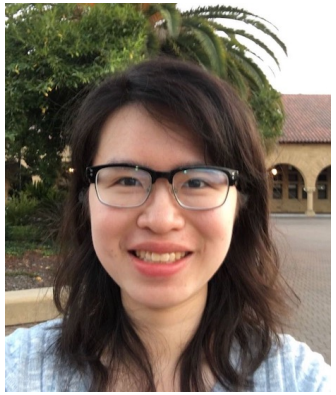


	Patient 1	Patient 2	Patient 3	Avg.
Glow (Kingma & Dhariwal, 2018)	-7.21	-5.59	-6.41	-6.40
Emerging (Hoogeboom et al., 2019)	-6.91	-8.48	-7.25	-7.55
Periodic (Hoogeboom et al., 2019)	-8.47	-9.623	-8.73	-8.94
Woodbury (Lu & Huang, 2020)	-11.68	-11.83	-10.91	-11.47
ButterflyFlow (Ours)	**-29.49**	**-27.07**	**-27.20**	**-27.92**

ButterflyFlow achieves strong performance on datasets with periodic structures relative to baselines.

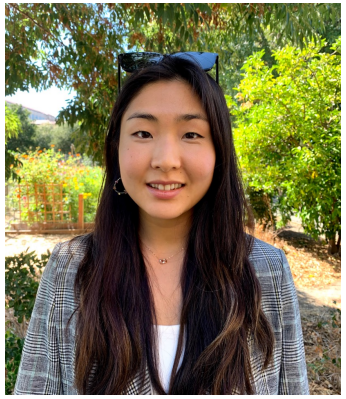
Thank you!

Twitter: @chenlin_meng, @linqi_zhou, @kristyechoi
Email: chelin@stanford.edu



Chenlin Meng*

Linqi Zhou*



Kristy Choi*

Tri Dao

Stefano Ermon