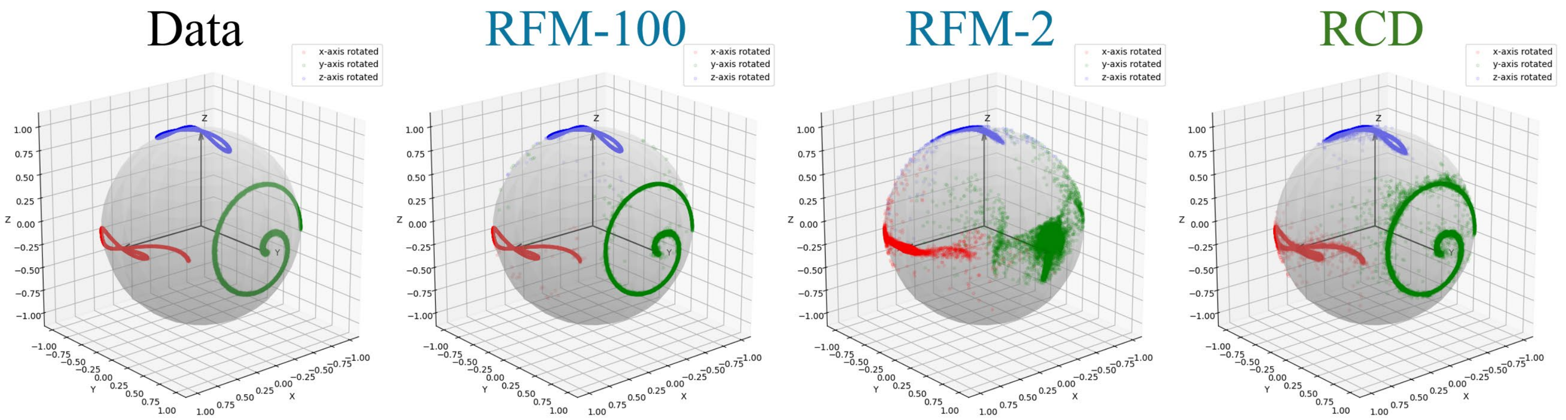
Riemannian Consistency Model

Chaoran Cheng, Yusong Wang, Yuxin Chen, Xiangxin Zhou, Nanning Zheng, Ge Liu

https://github.com/ccr-cheng/riemannian-consistency-model

NeurIPS 2025

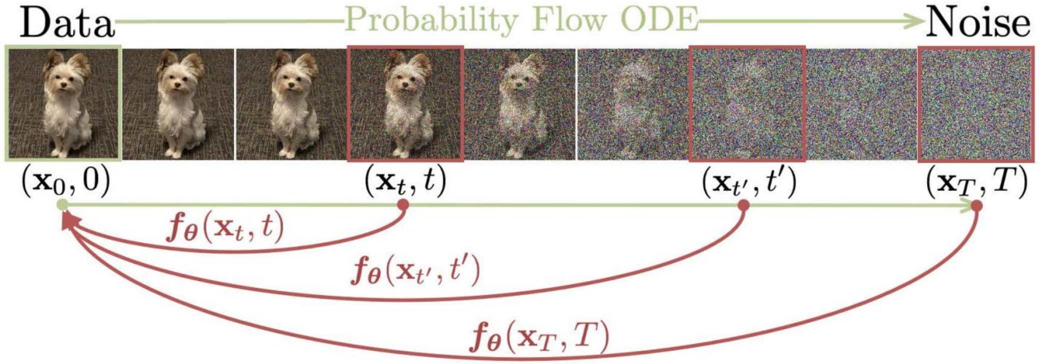
Presenter: Chaoran Cheng



UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

Background: Consistency Model

- Consistency Model (CM) maps any point (*noisy data*) on the ODE trajectory (*probability flow*) directly to its original (*clean data*).



$$\boxed{f(\mathbf{x}_t, t) = \mathbf{x}_0}$$

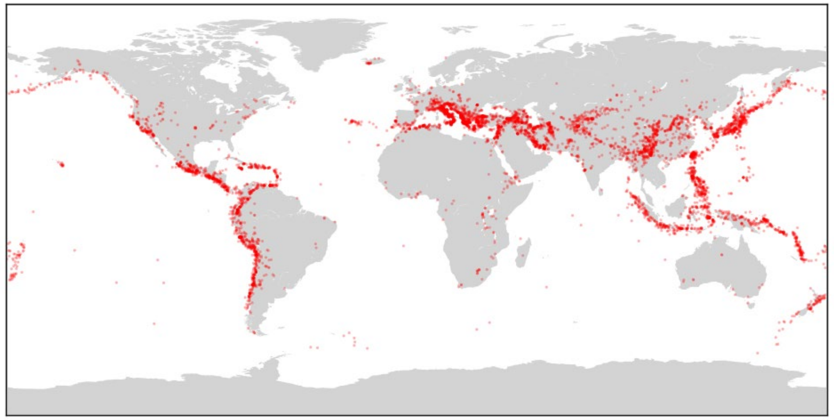
On any level of noisy data, the prediction remains consistent.

- Core ideas:
  - On a generation trajectory, the prediction should be *consistent*!
  - With a consistent prediction, we can *shortcut* the probability flow for **few-step generations**.

Song, Yang, et al. "Consistency models." (2023).

Riemannian Manifolds

- CM Parameterization: $f_\theta(\mathbf{x}_t, t) = c_{\text{skip}}(t)\,\mathbf{x}_t + c_{\text{out}}(t)\,F_\theta(\mathbf{x}_t, t)$
- CM Loss: $\arg\min_\theta \mathbb{E}\left[w(t_i)d(f_\theta(\mathbf{x}_{t_{i+1}}, t_{i+1}), f_{\theta^-}(\tilde{\mathbf{x}}_{t_i}, t_i))\right]$



- Can we extend to Riemannian manifolds?
  - Geology data on the *2-sphere*.
  - Torsion angles in *tori*.
  - Protein orientations in *SO(3)*.
- Challenges
  - Euclidean CM parameterization may break the *manifold constraint*
  - Euclidean CM loss is *ill-defined* for Riemannian manifolds

Riemannian Consistency Model

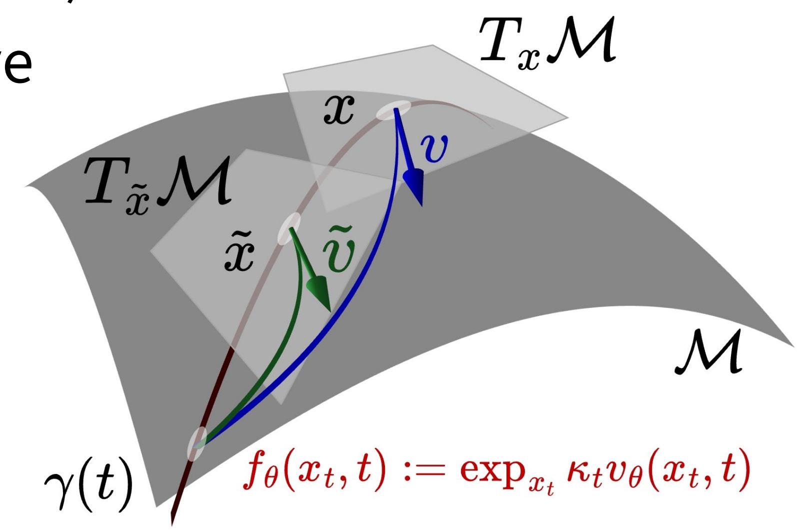
- Our solution: *Riemannian Consistency Model* (**RCM**)
  - Ensure *manifold constraints* by design
  - Support both *distillation* (**RCD**) as well as *training* (**RCT**)
  - Intuitive interpretation from a *kinematics* perspective
- RCM Parameterization:

Exponential map
Ensure manifold constraint

Scheduler

Vector field
$v\colon \mathcal{M} \to T\mathcal{M}$

$$f_\theta(x_t, t) := \boxed{\exp}_{x_t} \boxed{\kappa_t} \boxed{v_\theta}(x_t, t)$$



- RCM Loss:

Riemannian Distance
(geodesic distance)

$$\mathcal{L}_{\mathrm{RCM}}^{N} = N^2 \mathbb{E}_{t, x_t}\left[w_t \boxed{d_g^2}\left(f_\theta(x_t, t), f_{\theta^-}(x_{t+\Delta t}, t + \Delta t)\right)\right]$$

Riemannian Consistency Model

Discrete-Time **RCM** Loss

$$\mathcal{L}_{\mathrm{RCM}}^{N} = N^2 \mathbb{E}_{t,x_t}\left[w_t d_g^2\left(f_\theta\left(x_t, t\right), f_{\theta^-}\left(x_{t+\Delta t}, t+\Delta t\right)\right)\right]$$

Theorem 3.1

Continuous-Time **RCD** Loss

$$\mathcal{L}_{\mathrm{RCM}}^{\infty} := \lim_{N\to\infty} \mathcal{L}_{\mathrm{RCM}}^{N} = \mathbb{E}_{t,x_t}\left[w\| \,\mathrm{d}(\exp_x)_u\left(\dot\kappa v + \kappa \nabla_{\dot x} v\right) + \mathrm{d}(\exp u)_x(\dot x)\|_g^2\right]$$

Marginalization

Theorem 3.2

Stop-gradient is necessary for marginalization

Continuous-Time **RCT** Loss

$$\mathcal{L}_{\mathrm{RCM}}^{\infty} := \mathbb{E}_{t,x_t}\left[w\left\langle f_{\theta^-} - f_\theta + \dot f_{\theta^-}, \dot f_{\theta^-}\right\rangle_g\right], \quad \dot f = \mathrm{d}(\exp_x)_u\left(\dot\kappa v + \kappa \nabla_{\dot x} v\right) + \mathrm{d}(\exp u)_x(\dot x)$$

Approximation

Proposition 3.1

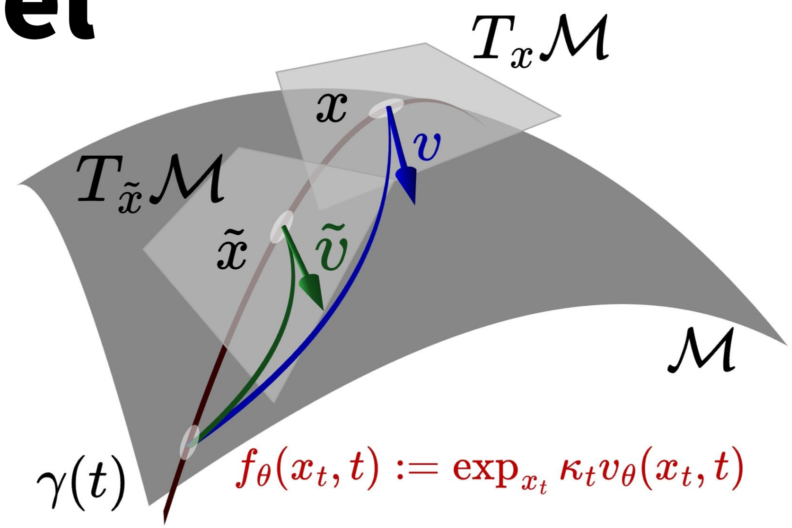
Circumvent the need to compute the differentials of the exponential map

Continuous-Time ***Simplified RCT*** Loss

$$\mathcal{L}_{\mathrm{sRCM}}^{\infty} := \mathbb{E}_{t,x_t}\left[w\langle v_{\theta^-} - v_\theta + \dot u_{\theta^-}, \dot u_{\theta^-}\rangle_g\right], \quad \dot u_\theta := \dot x + \dot\kappa v_\theta + \kappa \nabla_{\dot x} v_\theta$$

Riemannian Consistency Model

- *Riemannian Consistency Model* (**RCM**)
  - **RCD**: distill a pre-trained generative model
  - **RCT**: standalone generative framework



Algorithm 1 Simplified Riemannian Consistency Distillation (sRCD)

Input: Pre-trained RFM s_ϕ.
while not converged **do**
  Sample data x_1, noise x_0, and t.
  Calculate $x_t = \exp_{x_1}(\kappa_t \log_{x_1}(x_0))$.
  Calculate $s_\phi(x_t, t)$.
  Calculate $\dot{u}_\theta = s + \dot{\kappa} v_\theta + \kappa \nabla_s v_\theta$.
  Optimize the loss with $\nabla_\theta w \langle v_{\theta^-} - v_\theta + \dot{u}_{\theta^-}, \dot{u}_{\theta^-} \rangle_g$.
end while

Algorithm 2 Simplified Riemannian Consistency Training (sRCT)

Input: None.
while not converged **do**
  Sample data x_1, noise x_0, and t.
  Calculate $x_t = \exp_{x_1}(\kappa_t \log_{x_1}(x_0))$.
  Calculate $\dot{x}_t = \dot{\kappa}_t \log_{x_t} x_1 / \kappa_t$.
  Calculate $\dot{u}_\theta = \dot{x} + \dot{\kappa} v_\theta + \kappa \nabla_{\dot{x}} v_\theta$.
  Optimize the loss with $\nabla_\theta w \langle v_{\theta^-} - v_\theta + \dot{u}_{\theta^-}, \dot{u}_{\theta^-} \rangle_g$.
end while

Kinematics Perspective

- Uniform circular motion on the unit sphere S^1. Although the velocity is *constant* in some sense, the acceleration is *non-zero*. The centripetal acceleration satisfies:

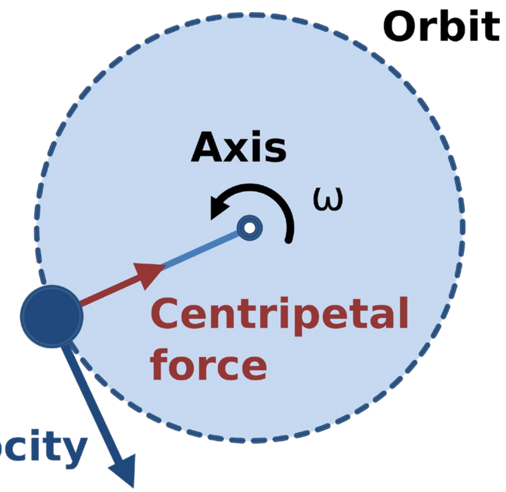
$$\boxed{\dot{v} + x\|v\|^2 = 0} \quad \Leftrightarrow \quad \boxed{\nabla_{\dot{\gamma}}\dot{\gamma} = 0}$$

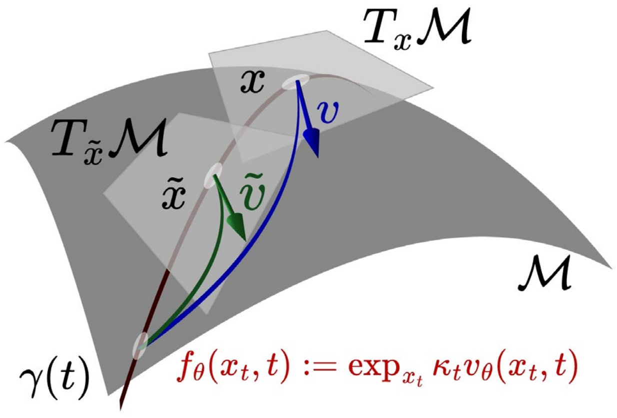
Kinematic Equation

Describe how a point moves on a curved manifold

Geodesic Equation

Describe how probability mass transports on a curved manifold





- For non-flat Riemannian manifolds, there exists an *extrinsic acceleration* due to the manifold constraint!

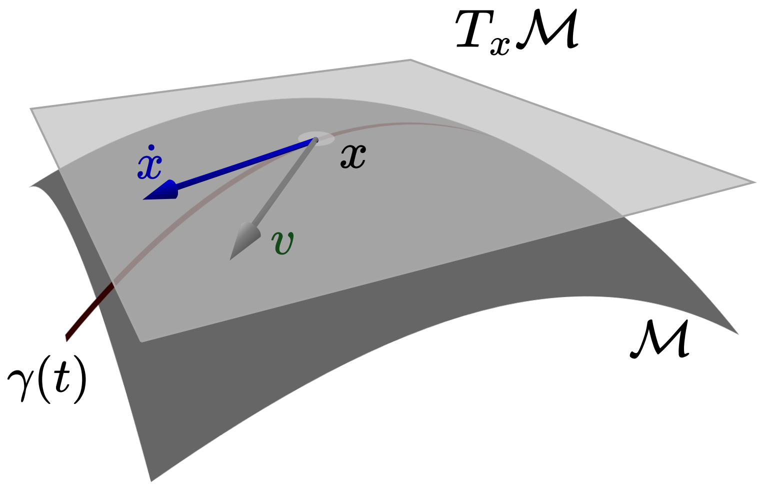
$$\nabla_u v = \dot{v}^k + \boxed{\Gamma_{ij}^k} v^i u^j$$

Christoffel symbols that describe the local geometry of the manifold.

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

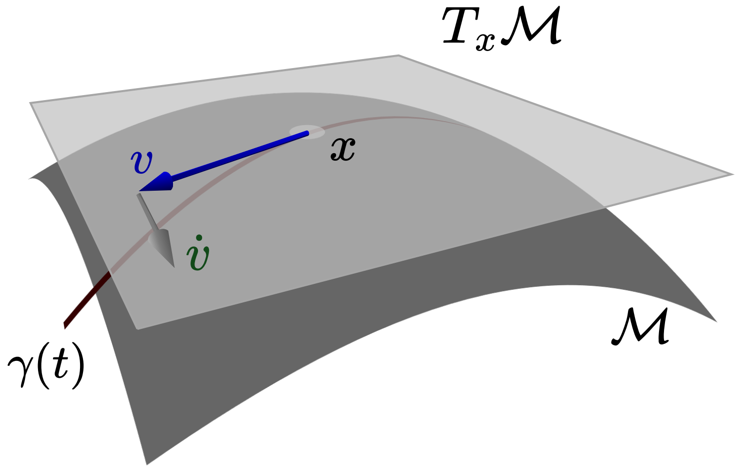
Kinematics Perspective

- The ***RCM Loss*** can be decomposed into three components:



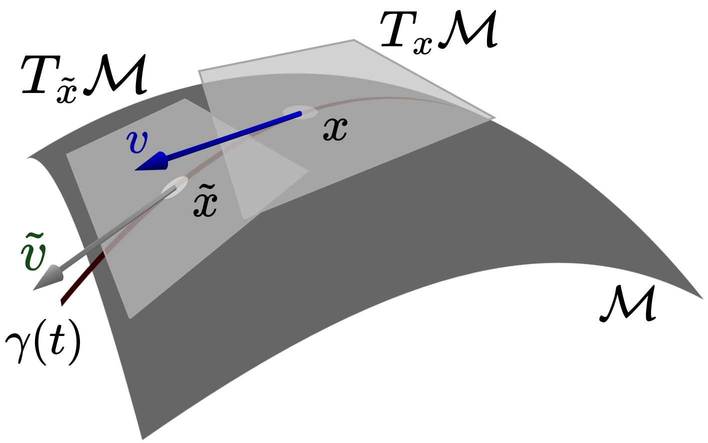
Vector field difference $\dot{x} - v$

The learning error should be small.



Vector field derivative $\dot{v}$

The vector field should remain constant if the manifold is flat.



Covariant derivative $\nabla_{\dot{x}} v$

$$\nabla_u v = \dot{v}^k + \boxed{\Gamma^k_{ij}} v^i u^j$$

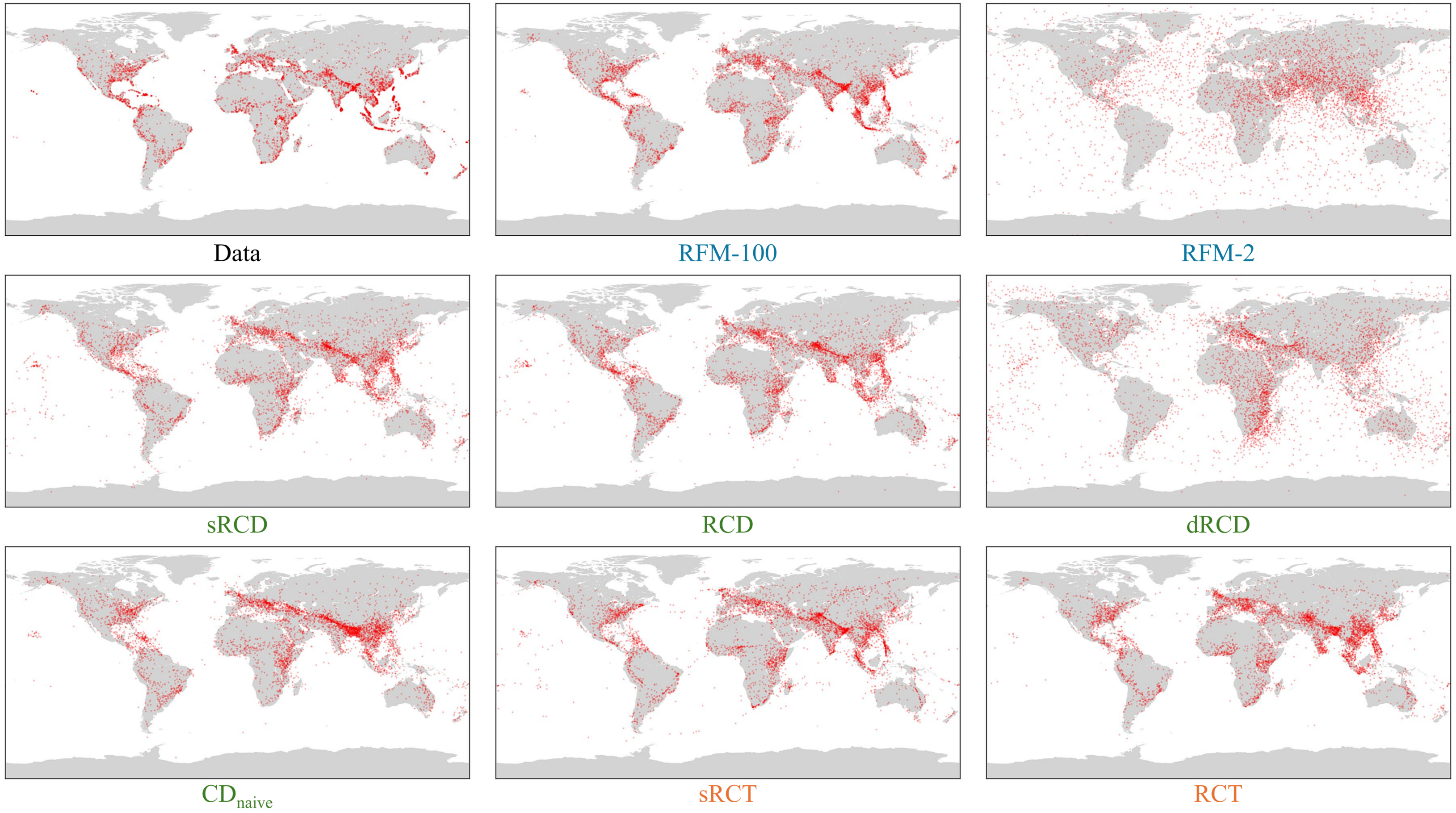
Also appear in Euclidean cases

Exclusive for non-flat Riemannian manifolds!

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Result

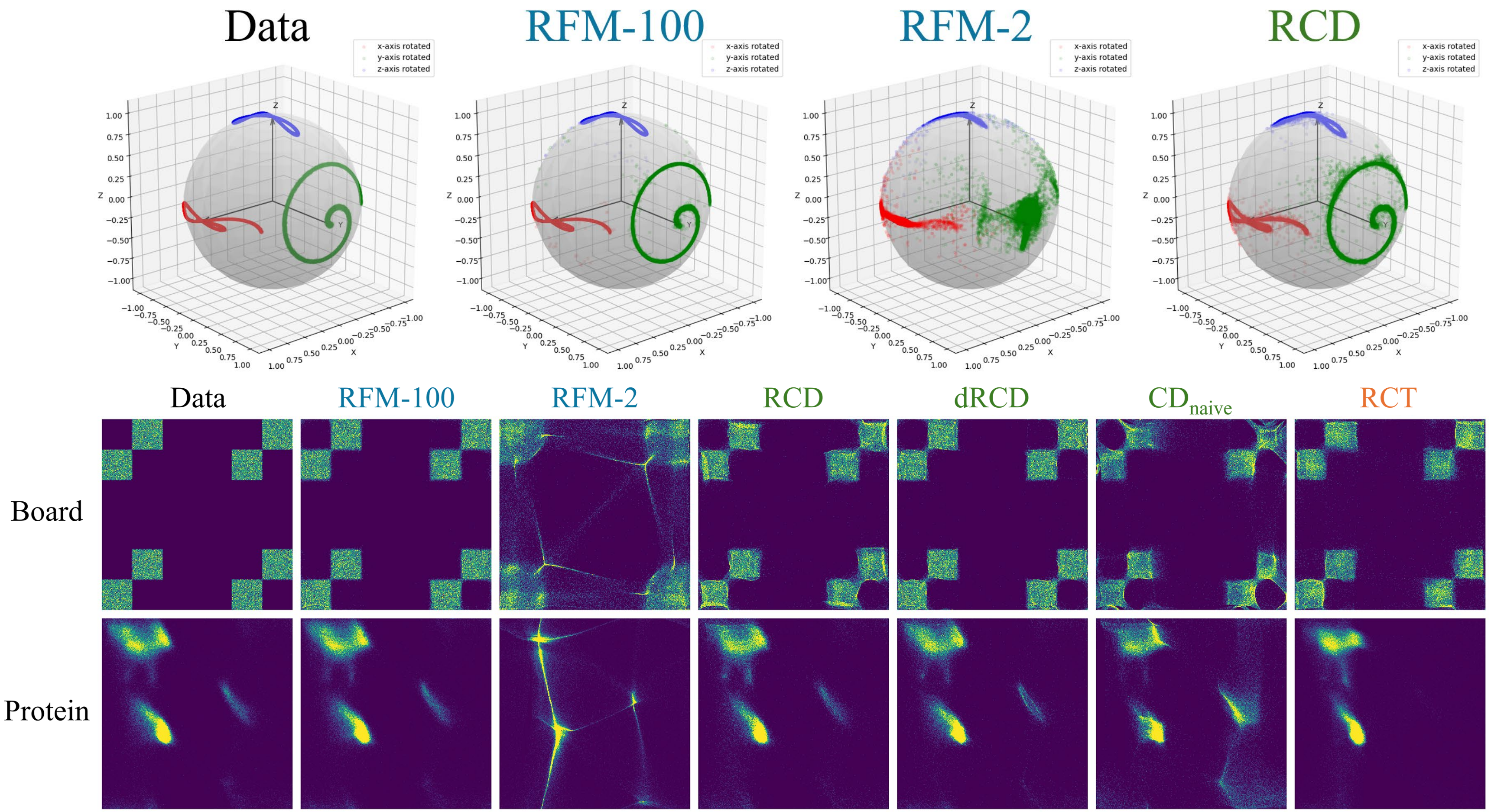
- For both distillation and training, RCM variants consistently achieve better generation quality in the 2-step generation setup.



UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Result

- For both distillation and training, RCM variants consistently achieve better generation quality in the 2-step generation setup.



Thanks

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN