UCLouvain

MadNIS

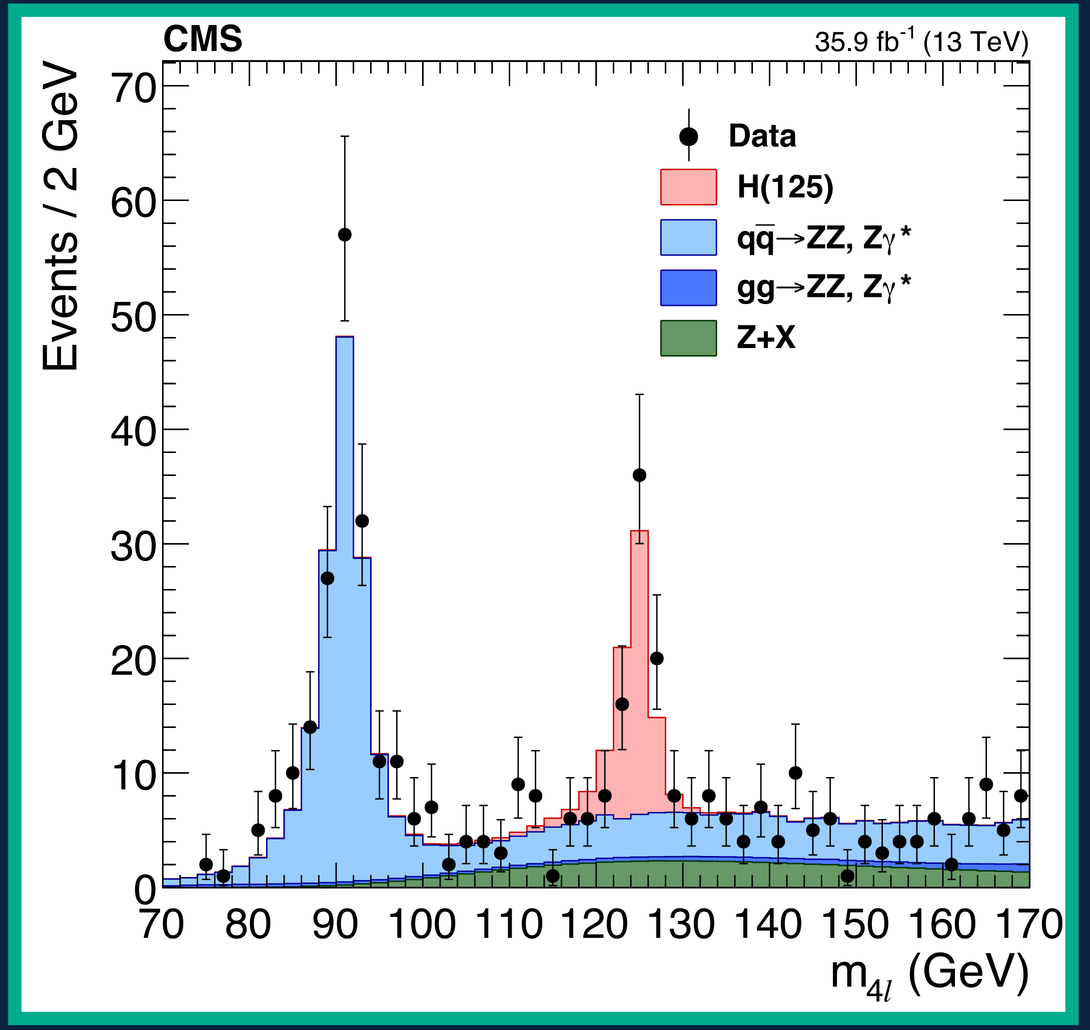
Neural networks for multi-channel importance sampling in MadGraph

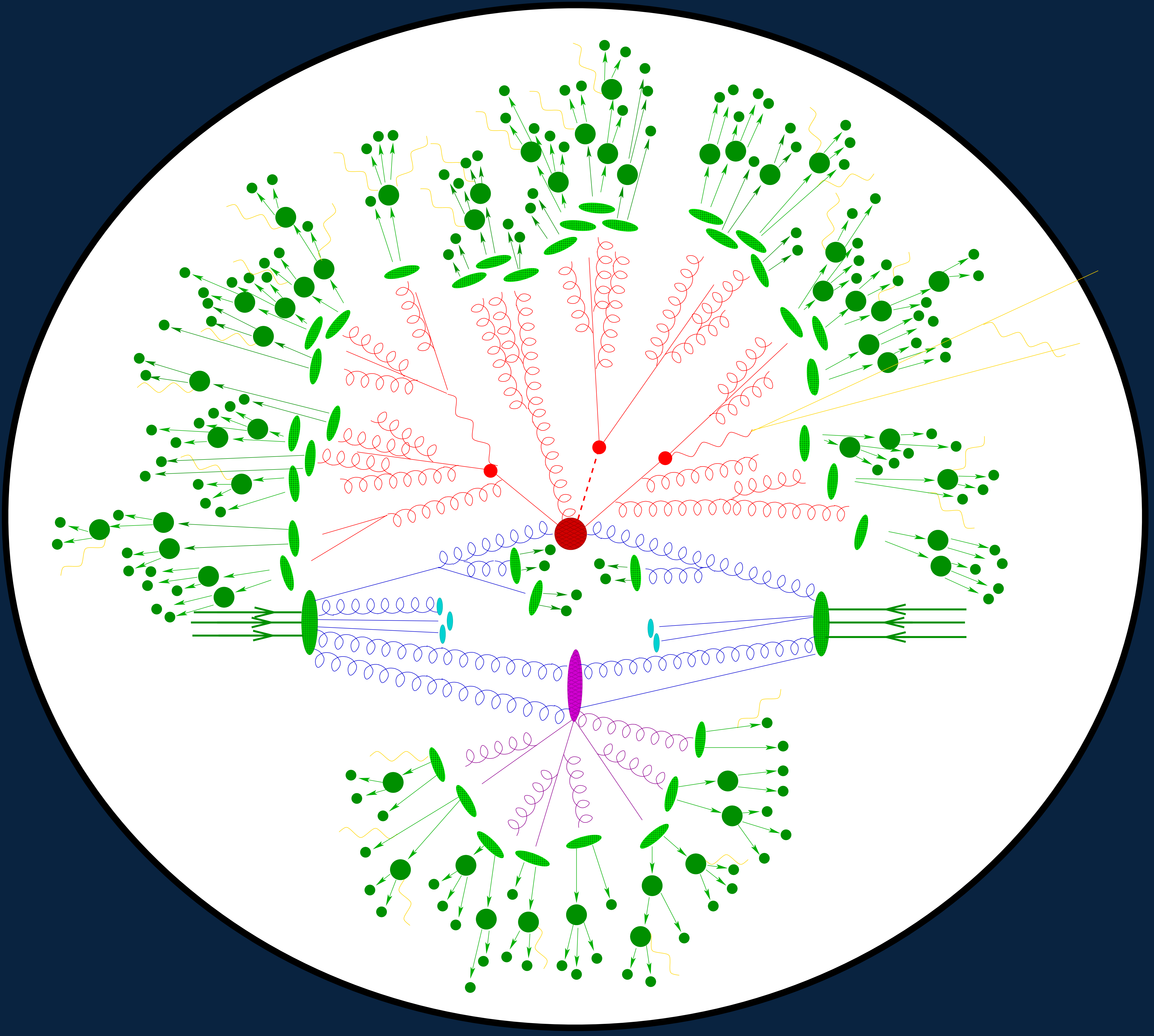
IML - Optimal transport and invertible algorithms

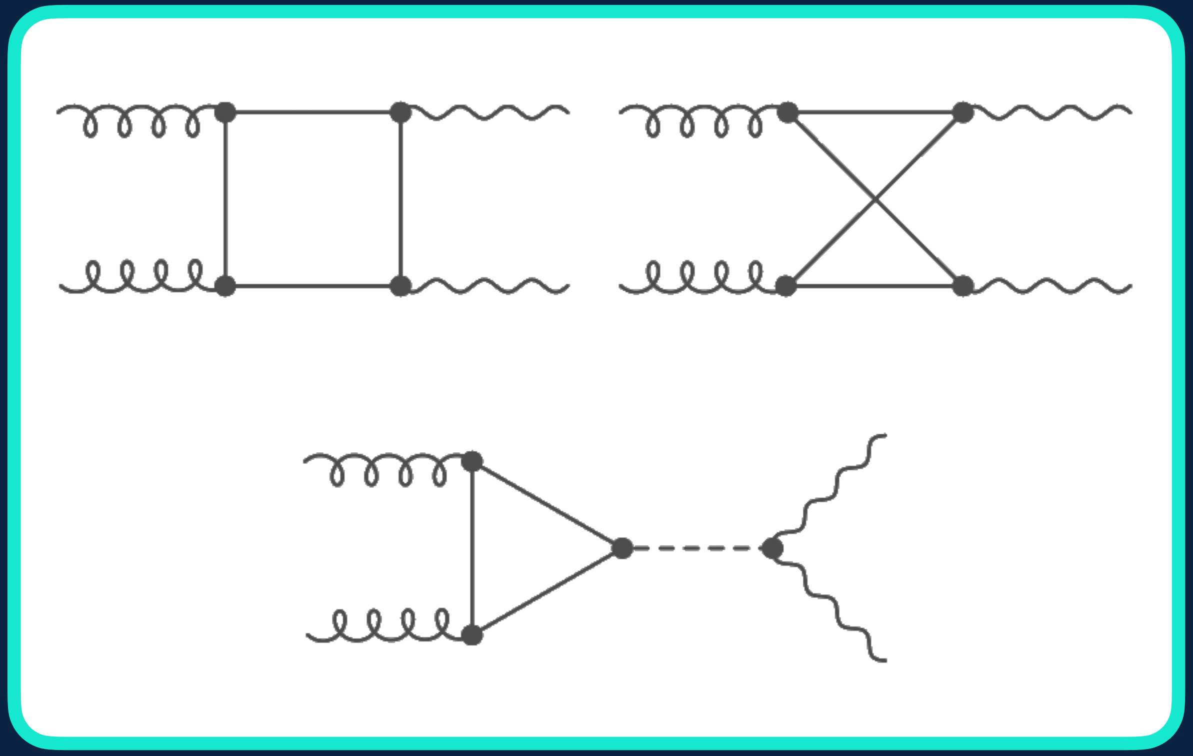
Ramon Winterhalder — UC Louvain

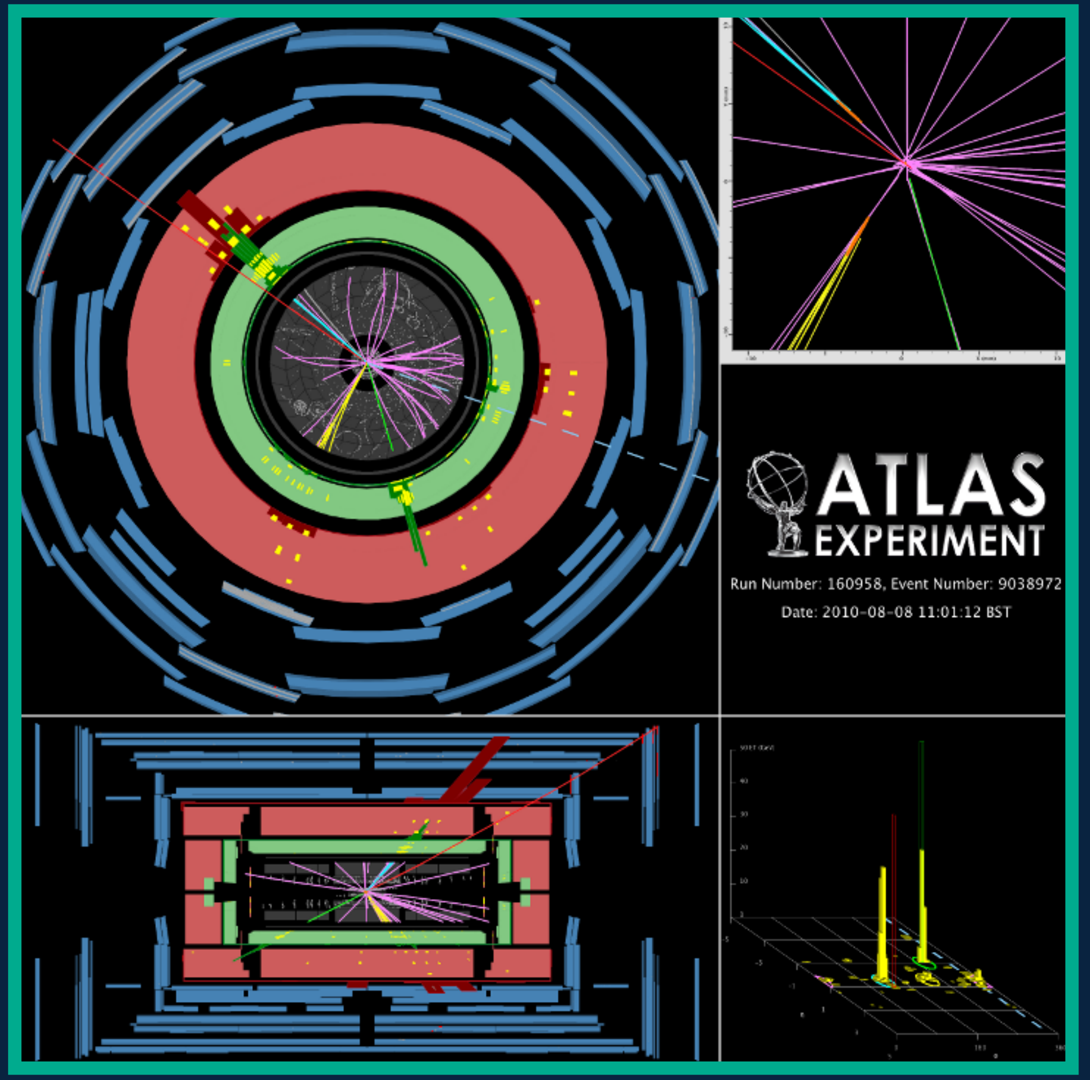
In collaboration with Anja Butter, Theo Heimel, Joshua Isaacson, Claudius Krause, Fabio Maltoni, Olivier Mattelaer and Tilman Plehn

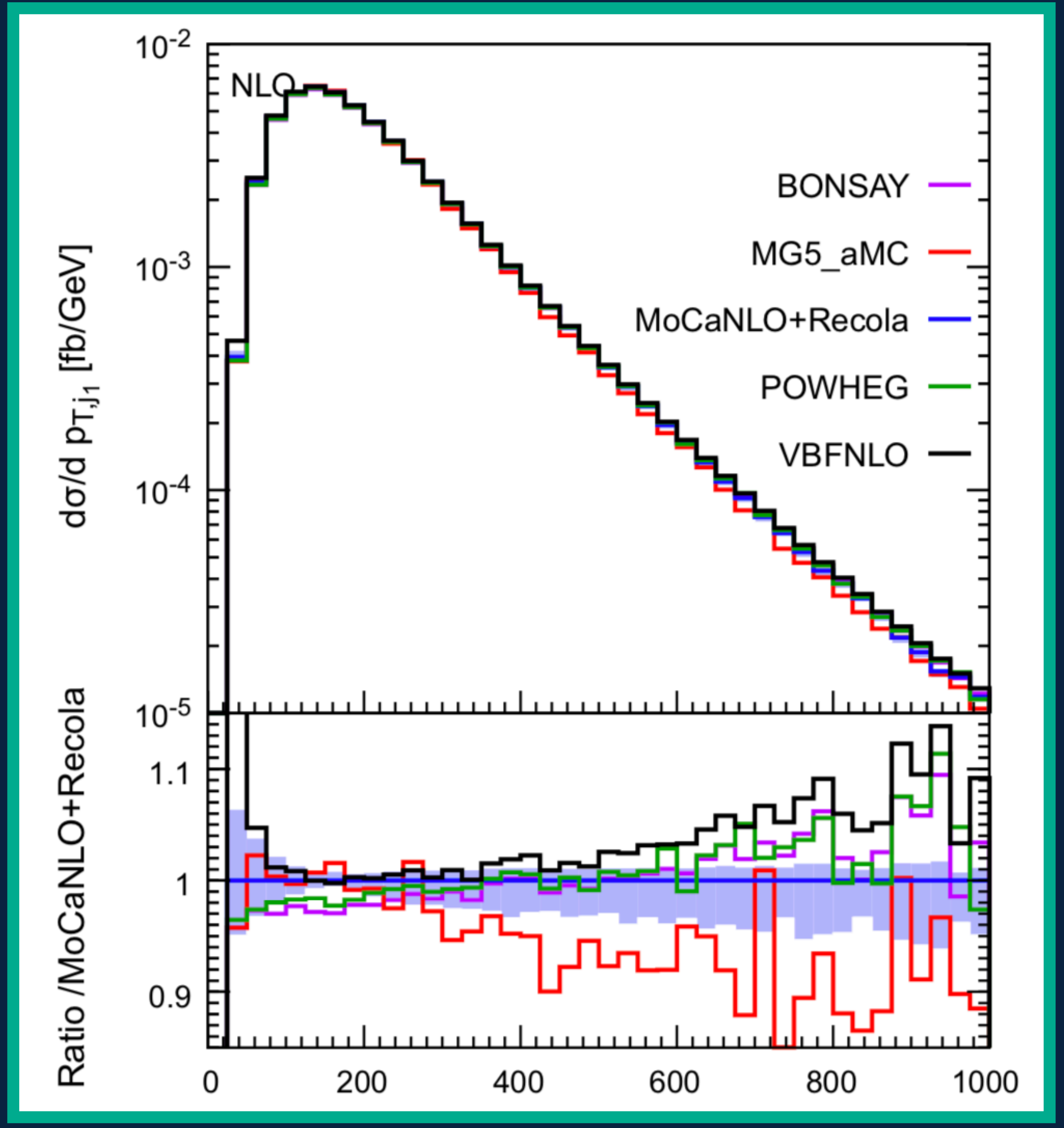
Data analysis in HEP



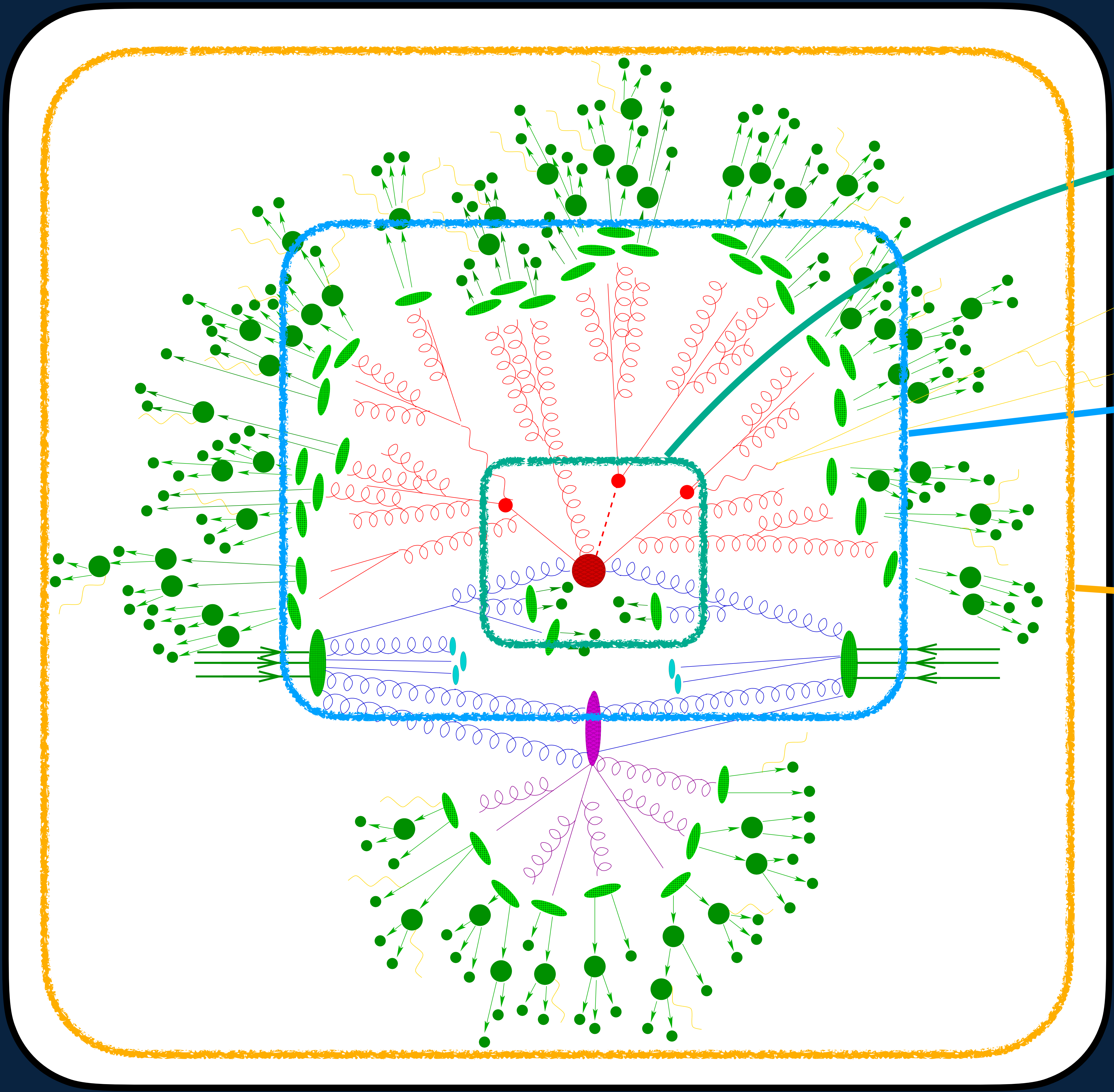








Simulations for the LHC



✦ **Hard Process**

✤ Depends on the Model (SM, SUSY,…)

✤ Perturbative QCD

✦ **Parton Showering**

✤ Universal (QCD)

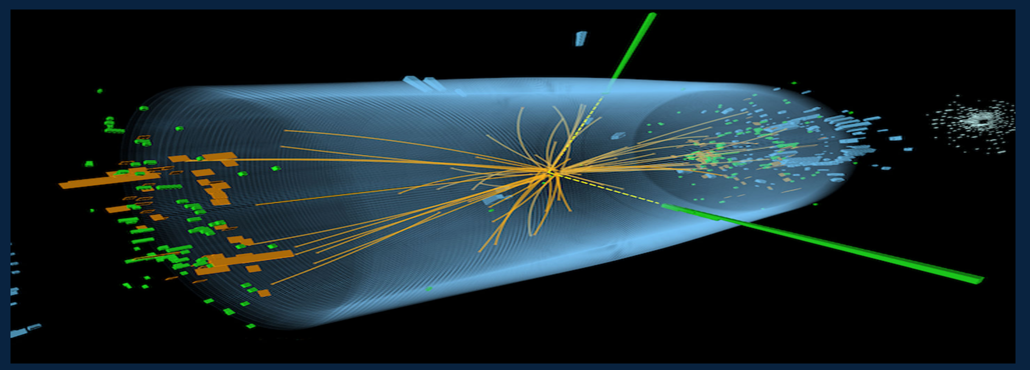
✦ **Hadronization**

✤Model-based, universal

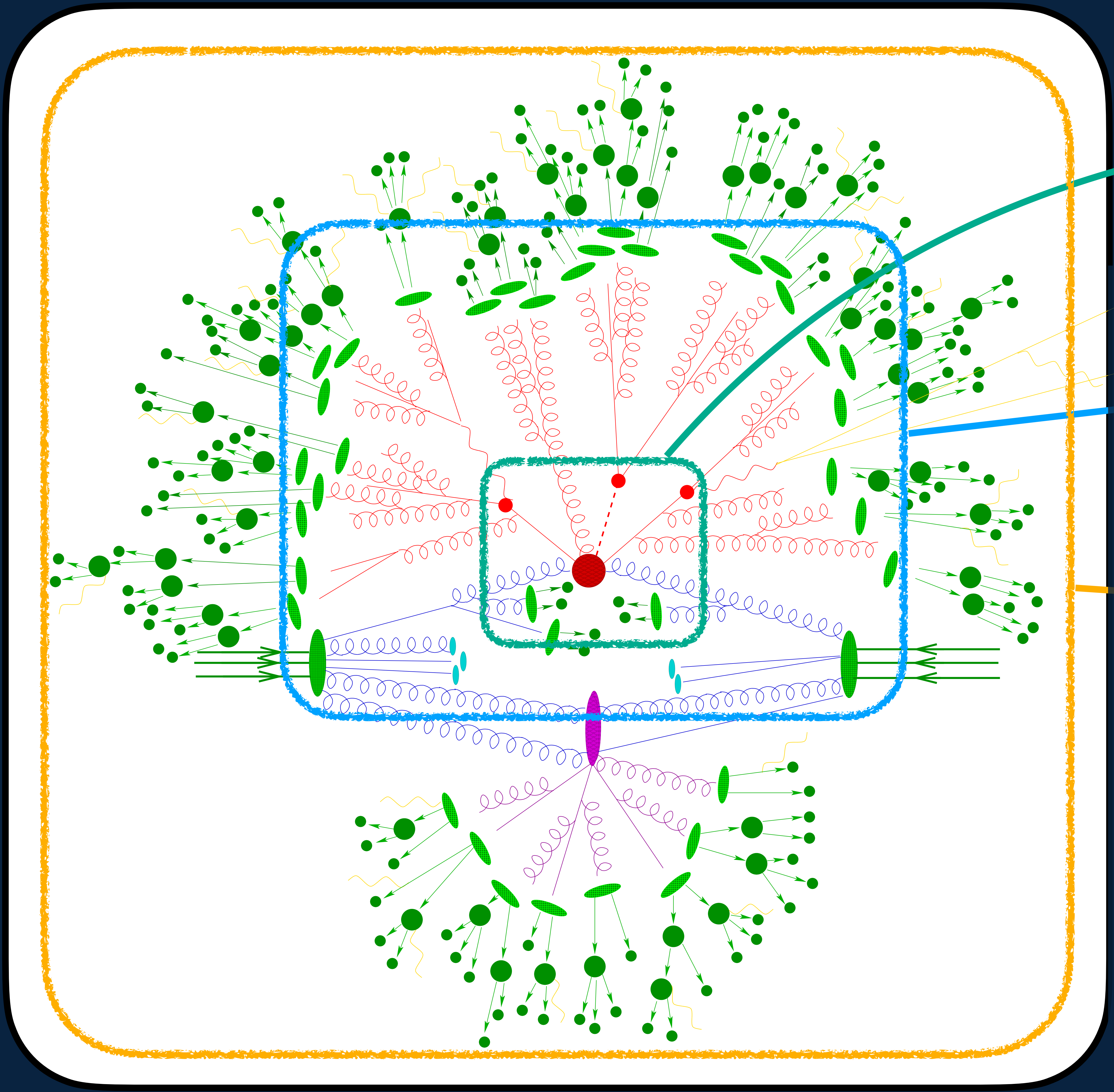
✦ **Underlying Event**

✤Model-based, non-universal

✦ **Detector Simulation**



Simulations for the LHC



✦ Hard Process

- ✤ Depends on the Model (SM, SUSY,…)
- ✤ Perturbative QCD

✦ Parton Showering

- ✤ Universal (QCD)

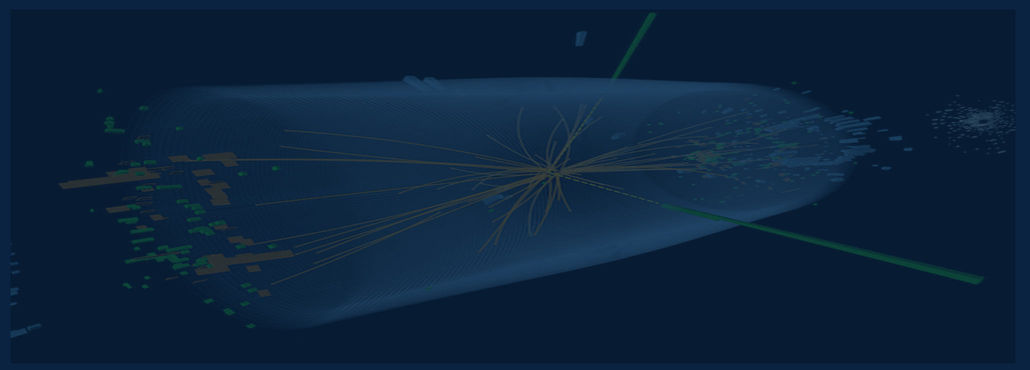
✦ Hadronization

- ✤ Model-based, universal

✦ Underlying Event

- ✤ Model-based, non-universal

✦ Detector Simulation



Theory predictions in HEP

$$M_{\lambda,c,\ldots}(p_a, p_b \,|\, p_1, \ldots, p_n) : \mathbb{M} \to \mathbb{C}$$

Quantum numbers: spin, colour charge etc.

Kinematics: Momenta in Minkowski space, masses, etc.

$$\sigma = \frac{1}{\text{flux}} \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b\, f(x_a) f(x_b) \int \mathrm{d}\Phi_n \left\langle \left| M_{\lambda,c,\ldots}(p_a, p_b \,|\, p_1, \ldots, p_n) \right|^2 \right\rangle$$

Cross section: more generally, differential observables*

PDFs: convolution over all possible initial state configurations

Phase-space integral: over final state kinematics

Squared amplitude: summed over final states, averaged over initial states

Theory predictions in HEP

$$M_{\lambda,c,\ldots}(p_a, p_b \,|\, p_1, \ldots, p_n) : \mathbb{M} \to \mathbb{C}$$

Quantum numbers:
spin, colour charge etc.

Kinematics:
Momenta in Minkowski space, masses, etc.

$$\sigma = \frac{1}{\text{flux}} \sum_{a,b} \int \mathrm{d}x_a \mathrm{d}x_b\, f(x_a) f(x_b) \int \mathrm{d}\Phi_n \langle\, |M_{\lambda,c,\ldots}(p_a, p_b \,|\, p_1, \ldots, p_n)|^2 \,\rangle$$

Cross section:
more generally, differential observables*

PDFs:
convolution over all possible initial state configurations

Phase-space integral:
over final state kinematics

Squared amplitude:
summed over final states, averaged over initial states

Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x\, f(x) \simeq \frac{1}{N} \sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2 \rangle_x - \langle f \rangle_x^2}{N-1}}$$

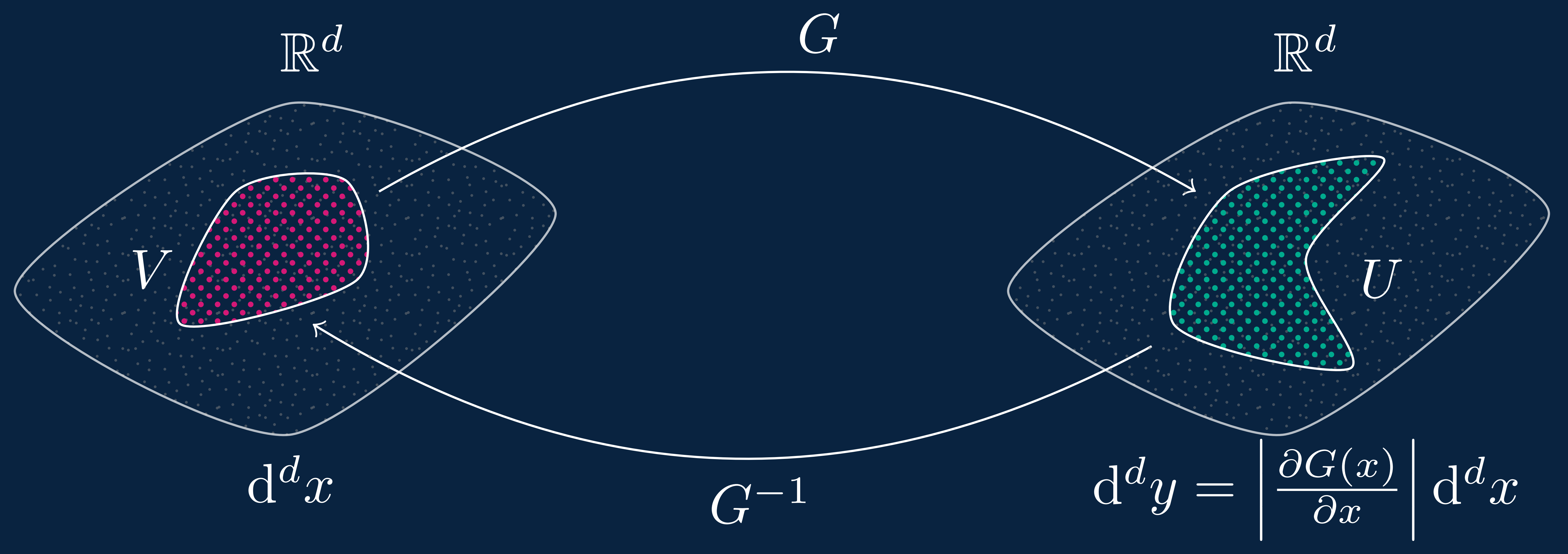
Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x\, f(x) \simeq \frac{1}{N}\sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2 \rangle_x - \langle f \rangle_x^2}{N-1}}$$

Mapping

$$y = G(x) \qquad g(x) = \left|\frac{\partial G(x)}{\partial x}\right|$$



Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x \, f(x) \simeq \frac{1}{N} \sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2 \rangle_x - \langle f \rangle_x^2}{N-1}}$$

Mapping

$$y = G(x) \qquad g(x) = \left| \frac{\partial G(x)}{\partial x} \right|$$

Importance Sampling

$$I = \int_U \mathrm{d}^d y \, \frac{f(x)}{g(x)} \Bigg|_{x=G^{-1}(y)} \simeq \langle f/g \rangle_y \qquad \sigma_I \simeq \sqrt{\frac{\langle (f/g)^2 \rangle_y - \langle f/g \rangle_y^2}{N-1}}$$

Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x\, f(x) \simeq \frac{1}{N}\sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2 \rangle_x - \langle f \rangle_x^2}{N-1}}$$

Mapping

$$y = G(x) \qquad g(x) = \left| \frac{\partial G(x)}{\partial x} \right|$$

Importance Sampling

$$I = \int_U \mathrm{d}^d y\, \left.\frac{f(x)}{g(x)}\right|_{x=G^{-1}(y)} \simeq \langle f/g \rangle_y \quad \sigma_I \simeq \sqrt{\frac{\langle (f/g)^2 \rangle_y - \langle f/g \rangle_y^2}{N-1}}$$

Wanted

$$\sigma_I \to 0$$

Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x\, f(x) \simeq \frac{1}{N}\sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2\rangle_x - \langle f \rangle_x^2}{N-1}}$$

Importance Sampling

$$I = \int_U \mathrm{d}^d y\, \frac{f(x)}{g(x)}\bigg|_{x=G^{-1}(y)} \simeq \langle f/g \rangle_y \qquad \sigma_I \simeq \sqrt{\frac{\langle (f/g)^2\rangle_y - \langle f/g \rangle_y^2}{N-1}}$$

Mapping

$$y = G(x) \qquad g(x) = \left|\frac{\partial G(x)}{\partial x}\right|$$

Wanted

$\sigma_I \to 0$

Needs

$f/g \approx$ const

Monte Carlo Integration

Standard Monte Carlo integration

$$I = \int_V \mathrm{d}^d x\, f(x) \simeq \frac{1}{N}\sum_{j=1}^{N} f(x_j) = \langle f \rangle_x \qquad \sigma_I \simeq \sqrt{\frac{\langle f^2\rangle_x - \langle f\rangle_x^2}{N-1}}$$

Mapping

$$y = G(x) \qquad g(x) = \left|\frac{\partial G(x)}{\partial x}\right|$$

Importance Sampling

$$I = \int_U \mathrm{d}^d y\, \frac{f(x)}{g(x)}\bigg|_{x=G^{-1}(y)} \simeq \langle f/g \rangle_y \quad \sigma_I \simeq \sqrt{\frac{\langle (f/g)^2\rangle_y - \langle f/g\rangle_y^2}{N-1}}$$

Multi-Channel Integration → $\sum \alpha_i(x) = 1$

$$I = \int_V \mathrm{d}^d x\, f(x) = \sum_i \int_V \mathrm{d}^d x\, \alpha_i(x) f(x) = \sum_i \int_{U_i} \mathrm{d}^d y_i\, \alpha_i(x) \frac{f(x)}{g_i(x)}\bigg|_{x\equiv x(y_i)}$$

Channel Mappings

$$y_i = G_i(x) \qquad g_i(x) = \left|\frac{\partial G_i(x)}{\partial x}\right|$$

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \mathrm{d}G_i(x)$$

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \mathrm{d}G_i(x)$$

Neural Channel Mappings

$$G_i(x) \to G_i(x \,|\, \varphi), \quad g_i(x \,|\, \varphi) = \left| \frac{\partial G_i(x \,|\, \varphi)}{\partial x} \right|$$

n-dimensional remapping

- tractable Jacobian with **normalizing flow**

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \mathrm{d}G_i(x)$$

Neural Channel Mappings

$$G_i(x) \to G_i(x\,|\,\varphi), \quad g_i(x\,|\,\varphi) = \left| \frac{\partial G_i(x\,|\,\varphi)}{\partial x} \right|$$

n-dimensional remapping

- tractable Jacobian with **normalizing flow**

→ *So far*: improvement factors of ~2-4 are achieved [2001.05486, 2001.05478, 2001.10028, 2112.09145]

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \mathrm{d}G_i(x)$$

Neural Channel Mappings

$$G_i(x) \to G_i(x\,|\,\varphi), \quad g_i(x\,|\,\varphi) = \left| \frac{\partial G_i(x\,|\,\varphi)}{\partial x} \right|$$

n-dimensional remapping

- tractable Jacobian with **normalizing flow**

→ *So far*: improvement factors of ~2-4 are achieved [2001.05486, 2001.05478, 2001.10028, 2112.09145]

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \mathrm{d}G_i(x)$$

Neural Channel Weights

$$\alpha_i(x) \to \alpha_i(x|\theta), \quad \sum_i \alpha_i(x|\theta) = 1$$

k-dimensional regression

- with boundary condition

Neural Channel Mappings

$$G_i(x) \to G_i(x|\varphi), \quad g_i(x|\varphi) = \left| \frac{\partial G_i(x|\varphi)}{\partial x} \right|$$

n-dimensional remapping

- tractable Jacobian with **normalizing flow**

⟶ *So far*: improvement factors of ~2-4 are achieved [2001.05486, 2001.05478, 2001.10028, 2112.09145]

Neural Multi-Channel Monte Carlo

Multi-Channel Importance Sampling

$$I = \sum_i \int_{U_i} \alpha_i(x) \frac{f(x)}{g_i(x)} \, dG_i(x)$$

Further Improvements?

Neural Channel Weights

$$\alpha_i(x) \to \alpha_i(x \,|\, \theta), \quad \sum_i \alpha_i(x \,|\, \theta) = 1$$

k-dimensional regression

- with boundary condition

Neural Channel Mappings

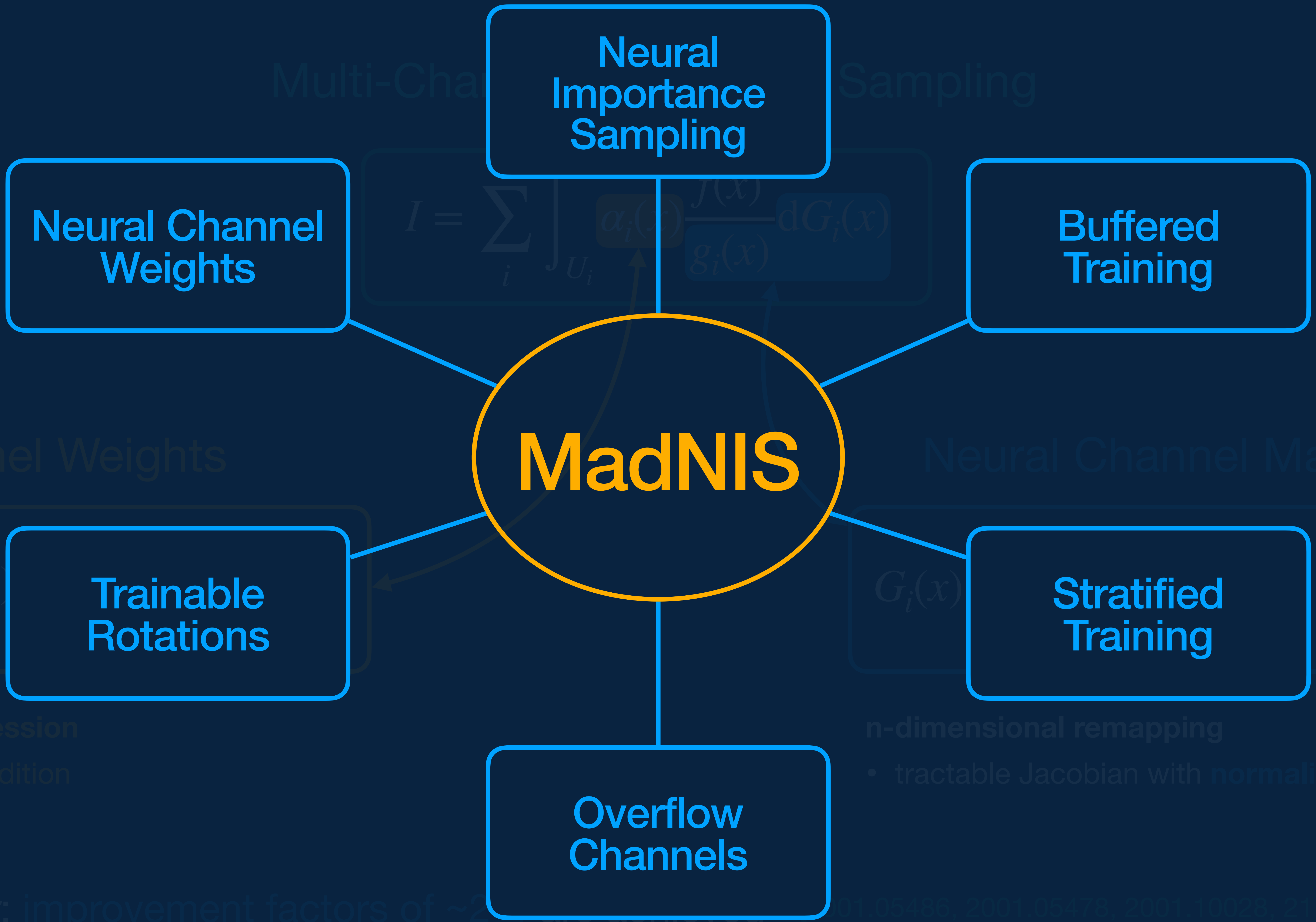
$$G_i(x) \to G_i(x \,|\, \varphi), \quad g_i(x \,|\, \varphi) = \left| \frac{\partial G_i(x \,|\, \varphi)}{\partial x} \right|$$

n-dimensional remapping

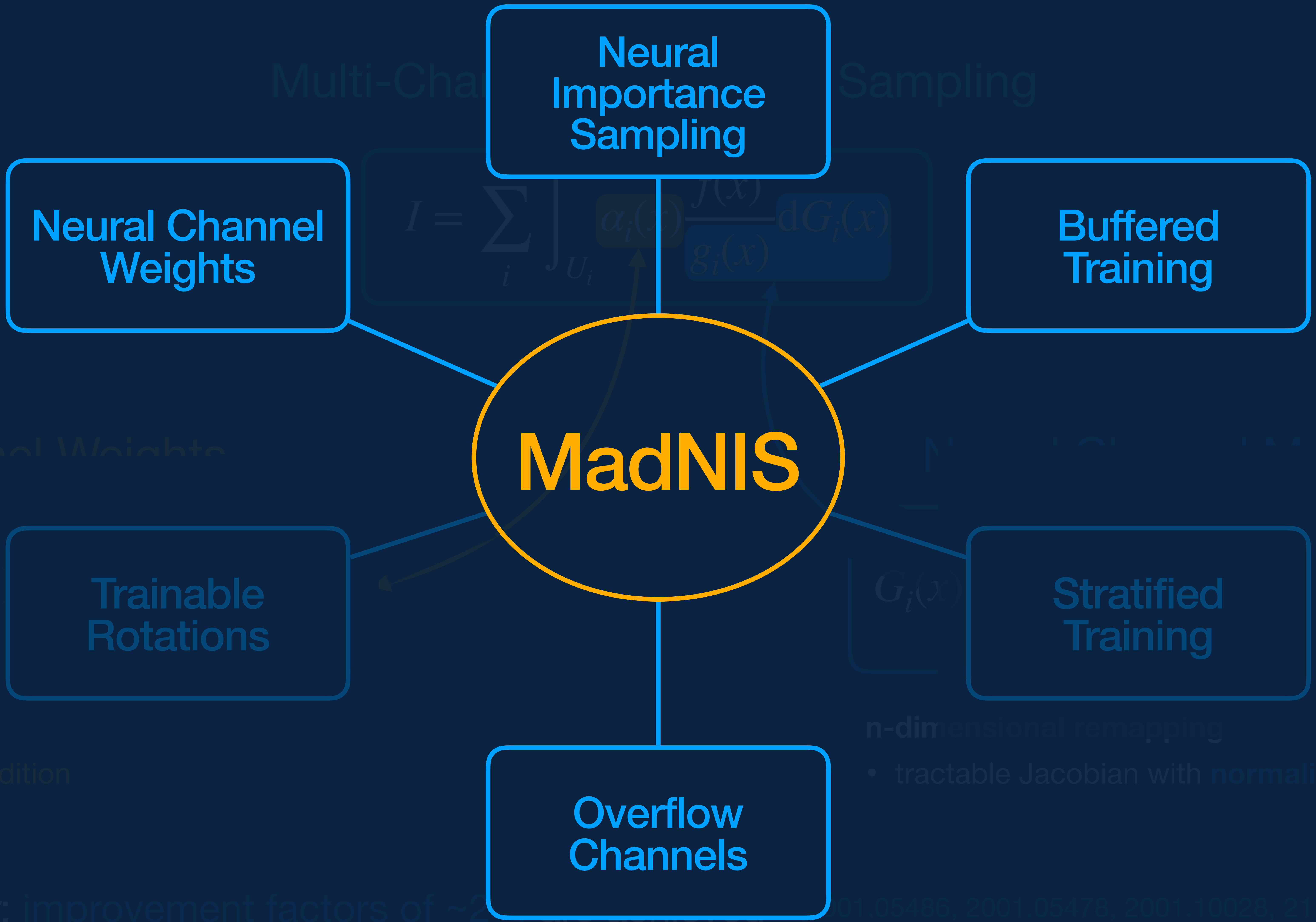
- tractable Jacobian with normalizing flow

→ *So far*: improvement factors of ~2-4 are achieved [2001.05486, 2001.05478, 2001.10028, 2112.09145]

Neural Multi-Channel Monte Carlo



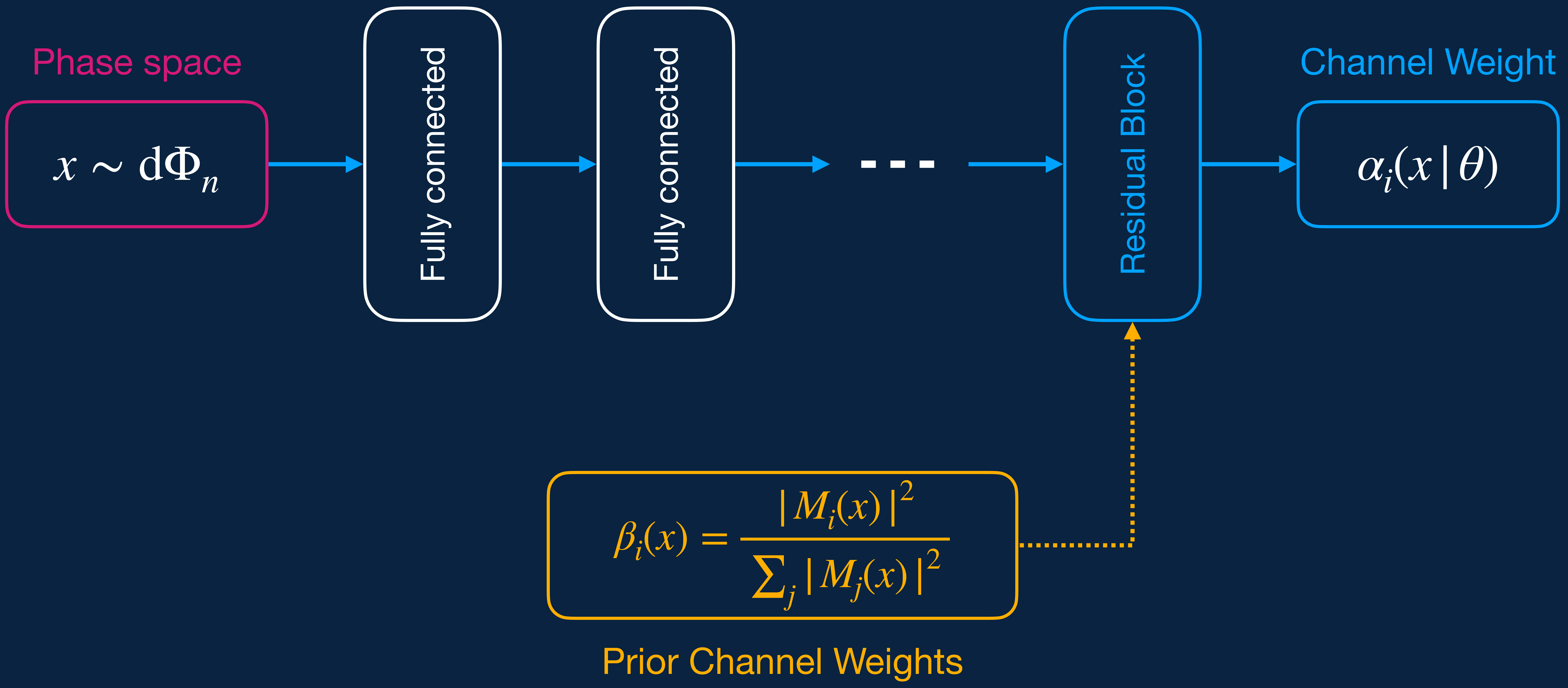
Neural Multi-Channel Monte Carlo



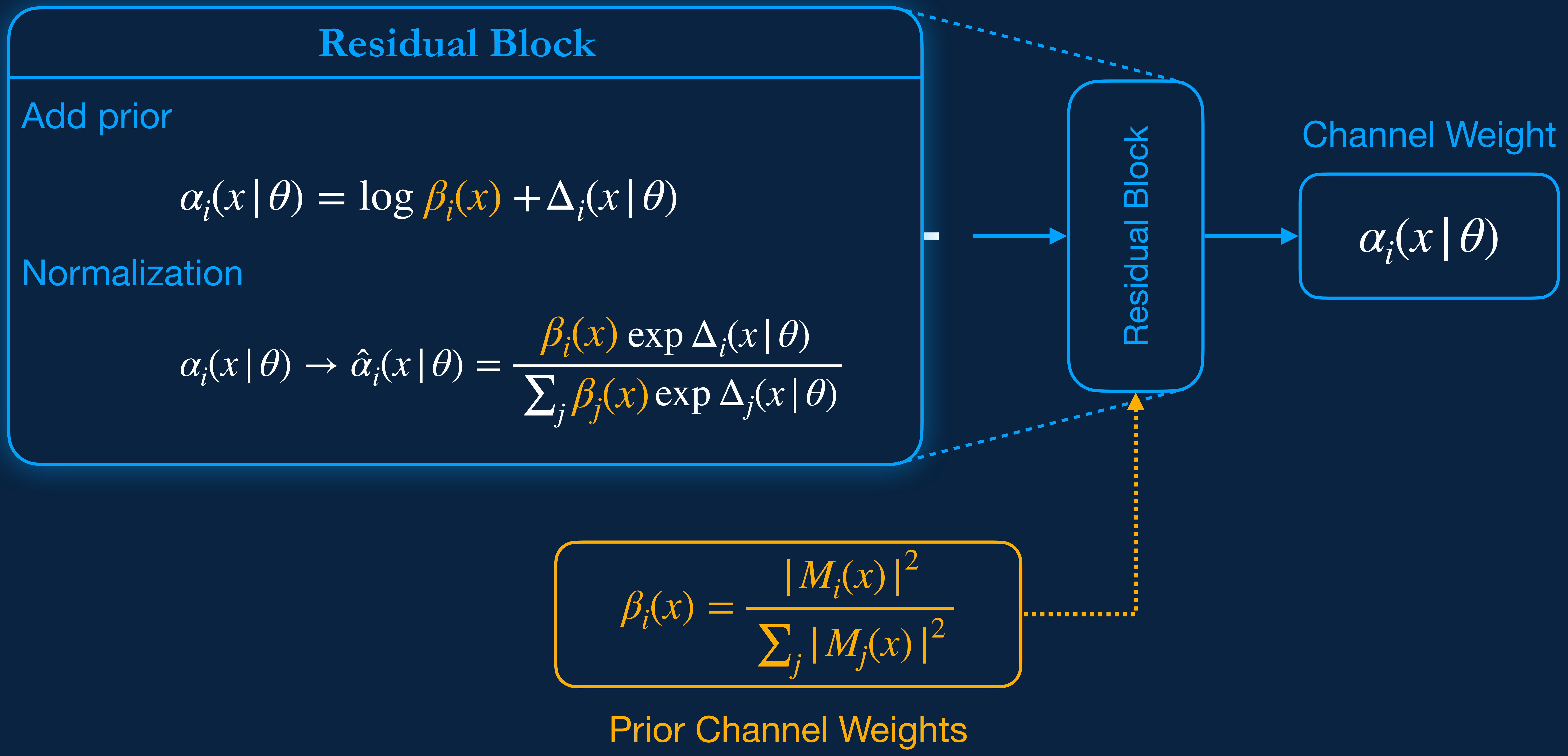
MadNIS

Neural Channel Weights

Neural Channel Weights



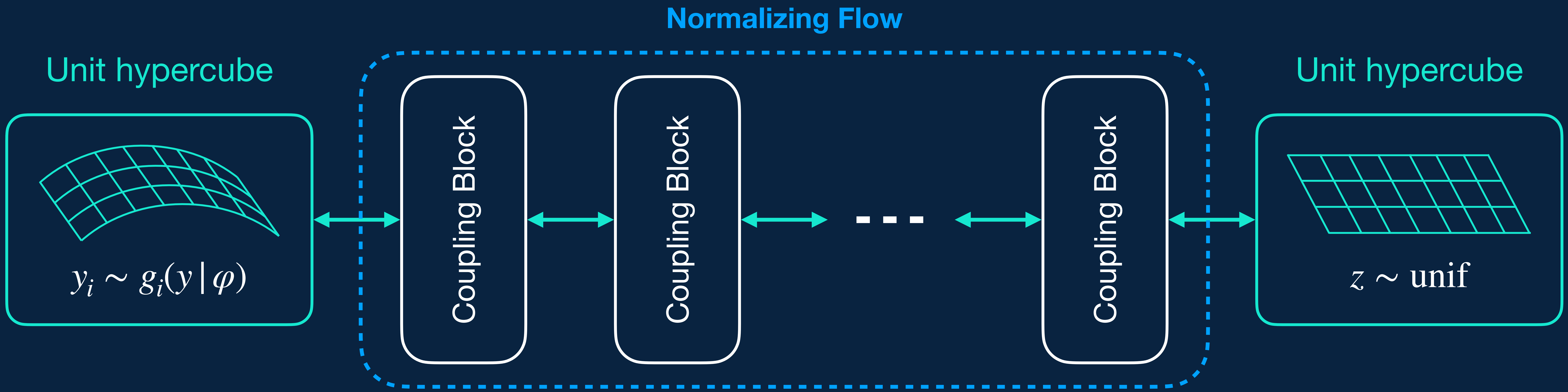
Neural Channel Weights



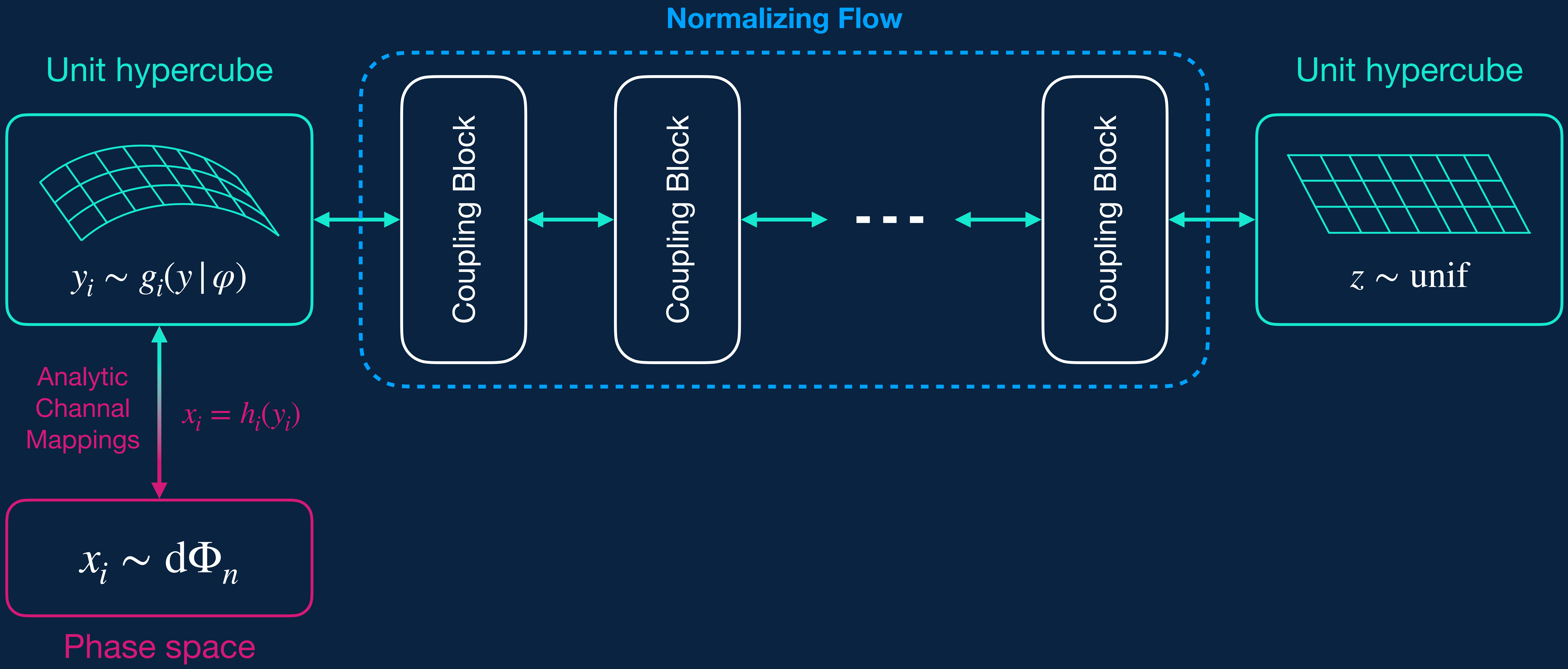
MadNIS

Neural Importance Sampling

Neural Importance Sampling



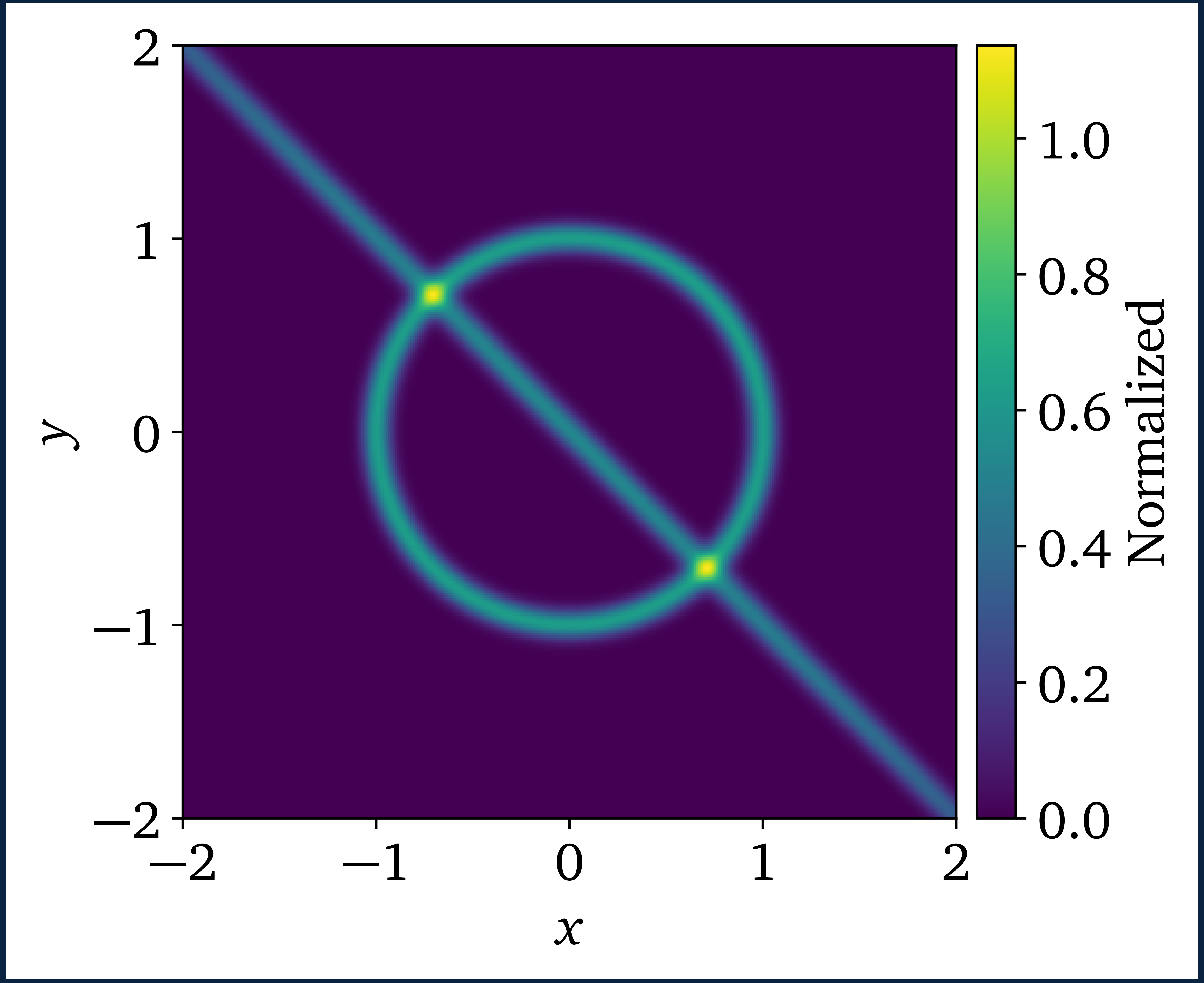
Neural Importance Sampling



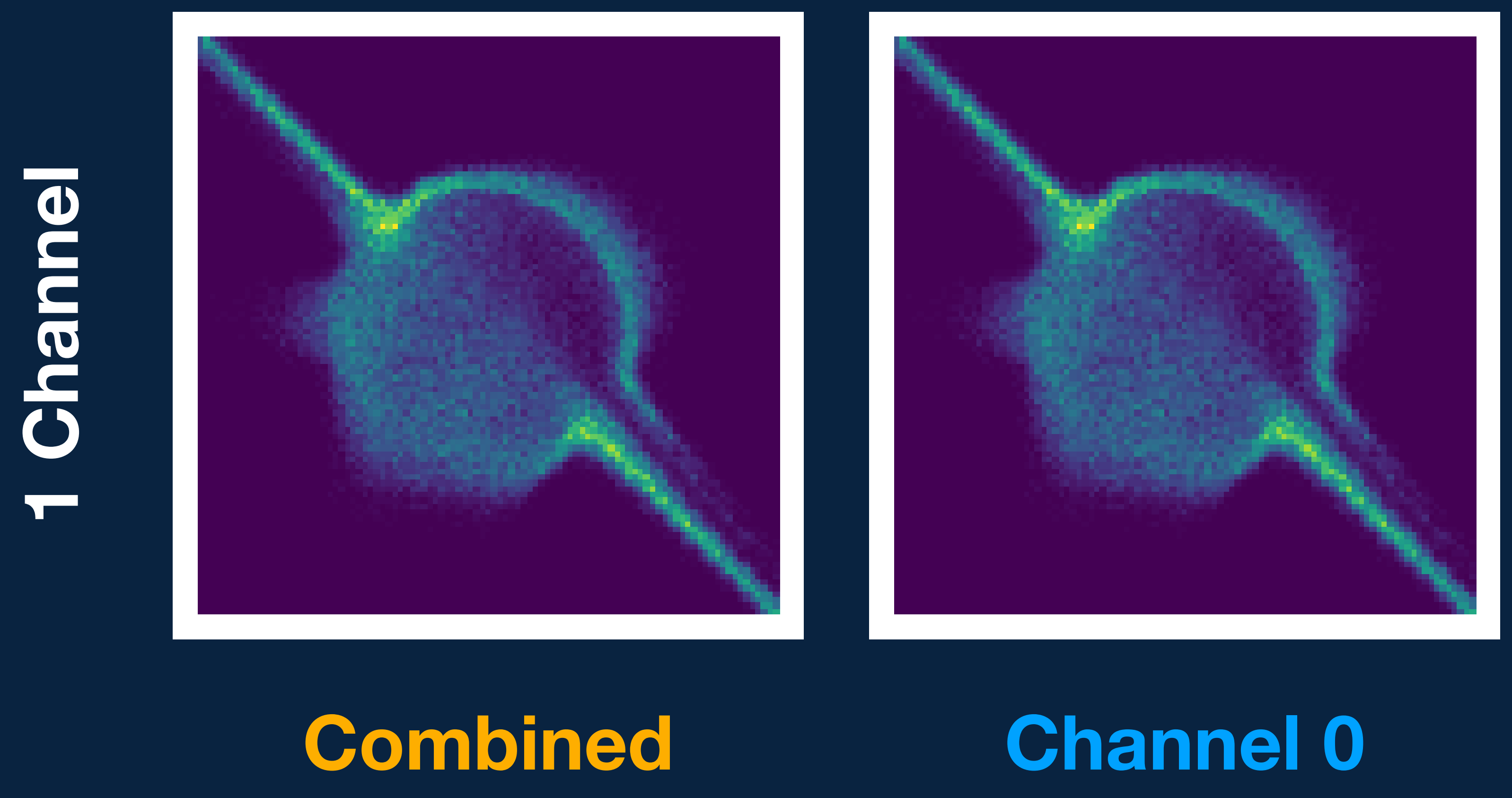
Example I

Crossed Ring

Toy Example: Crossed Ring

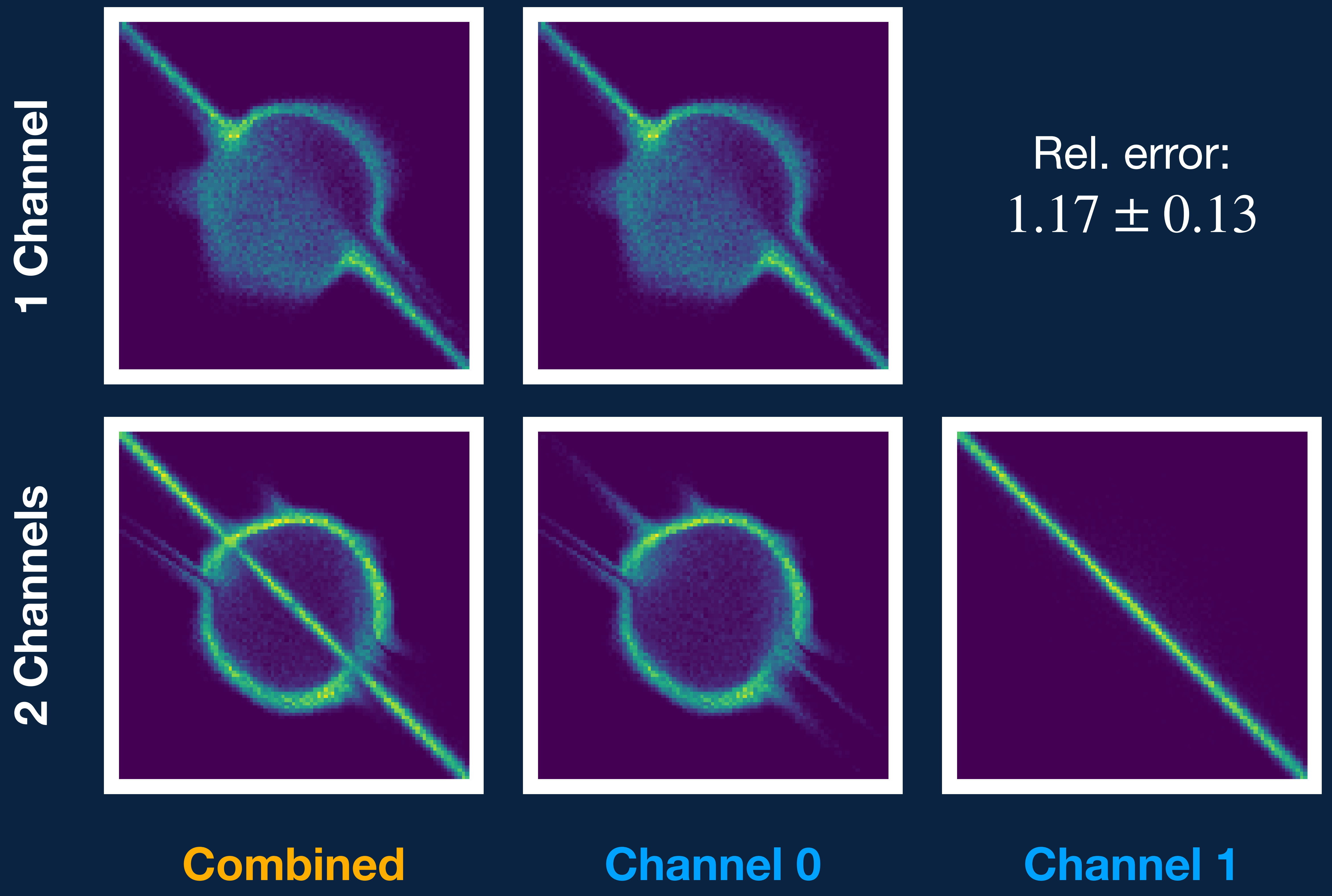


Example I: Crossed Ring



Rel. error:
1.17 ± 0.13

Example I: Crossed Ring

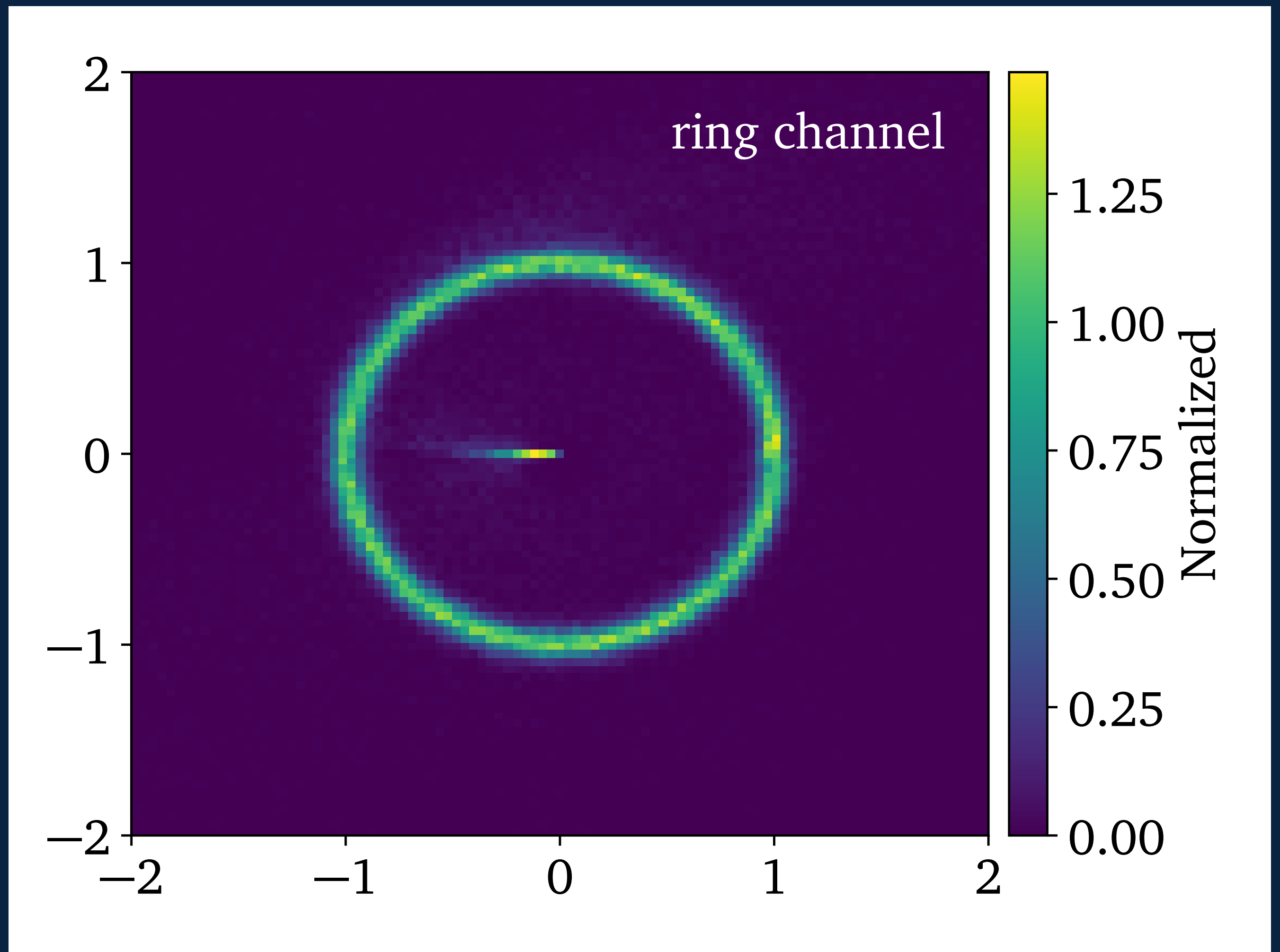


Rel. error:
1.17 ± 0.13

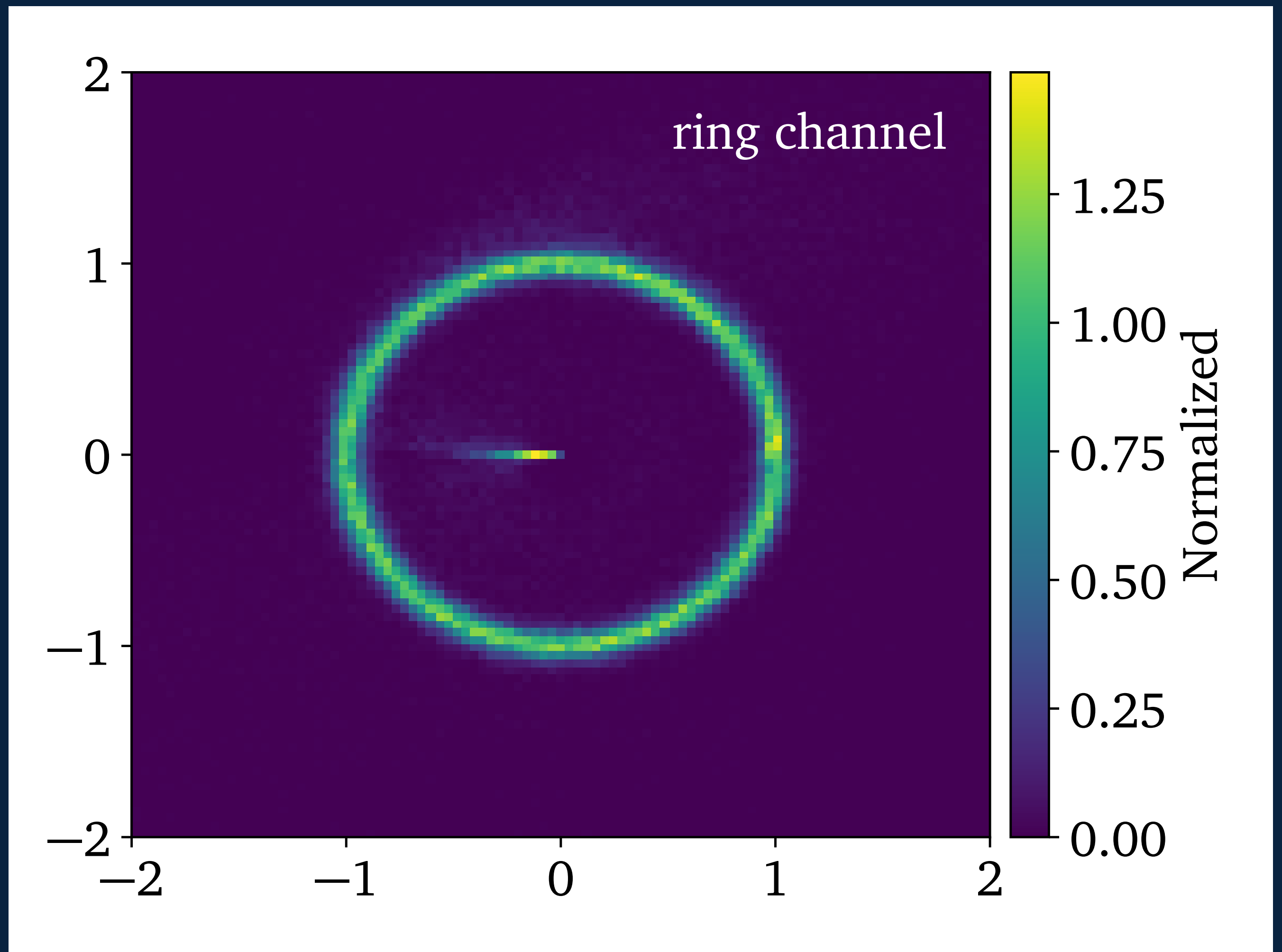
Rel. error:
0.71 ± 0.15

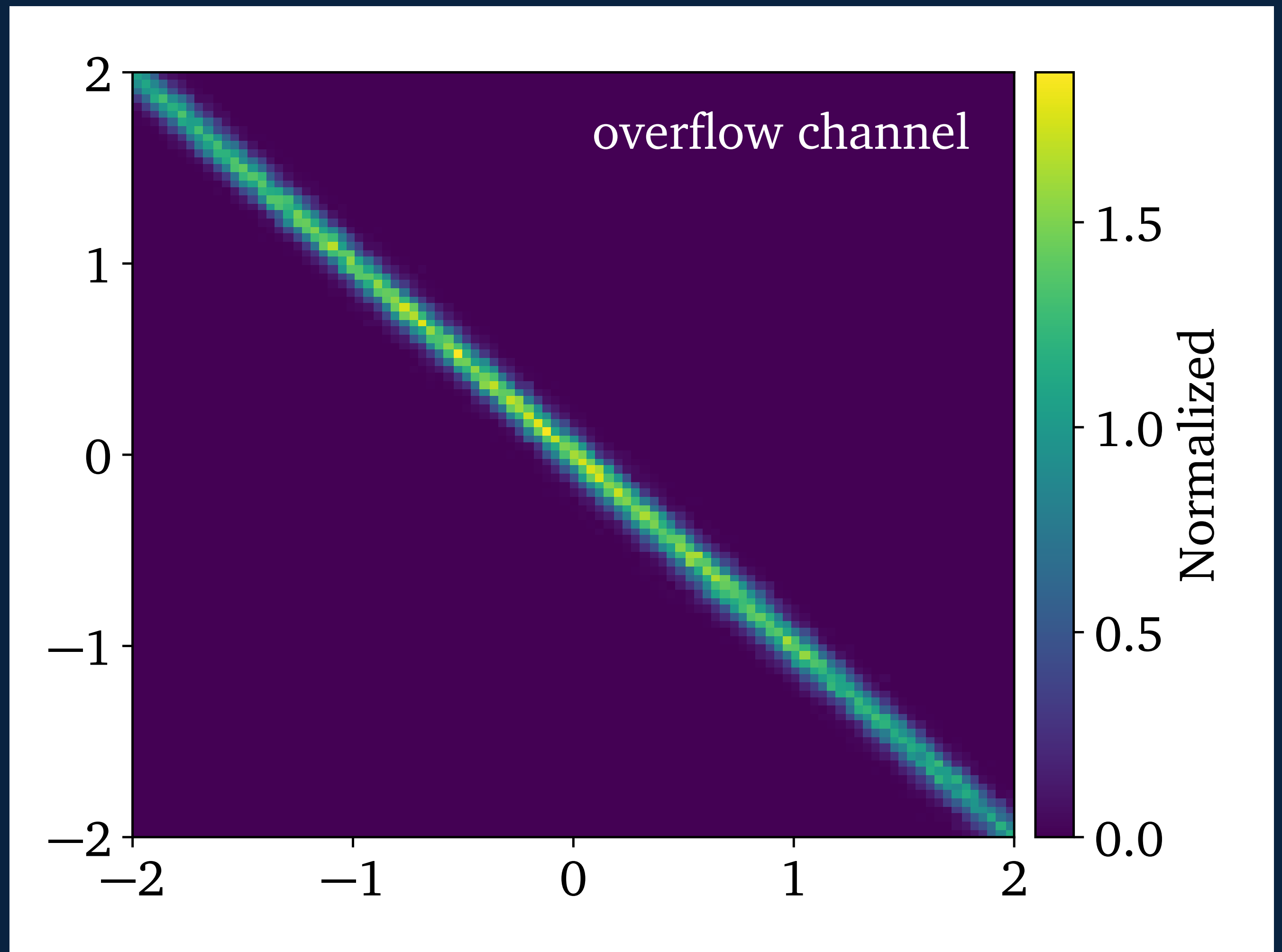
Example I: Crossed Ring

Example I: Crossed Ring

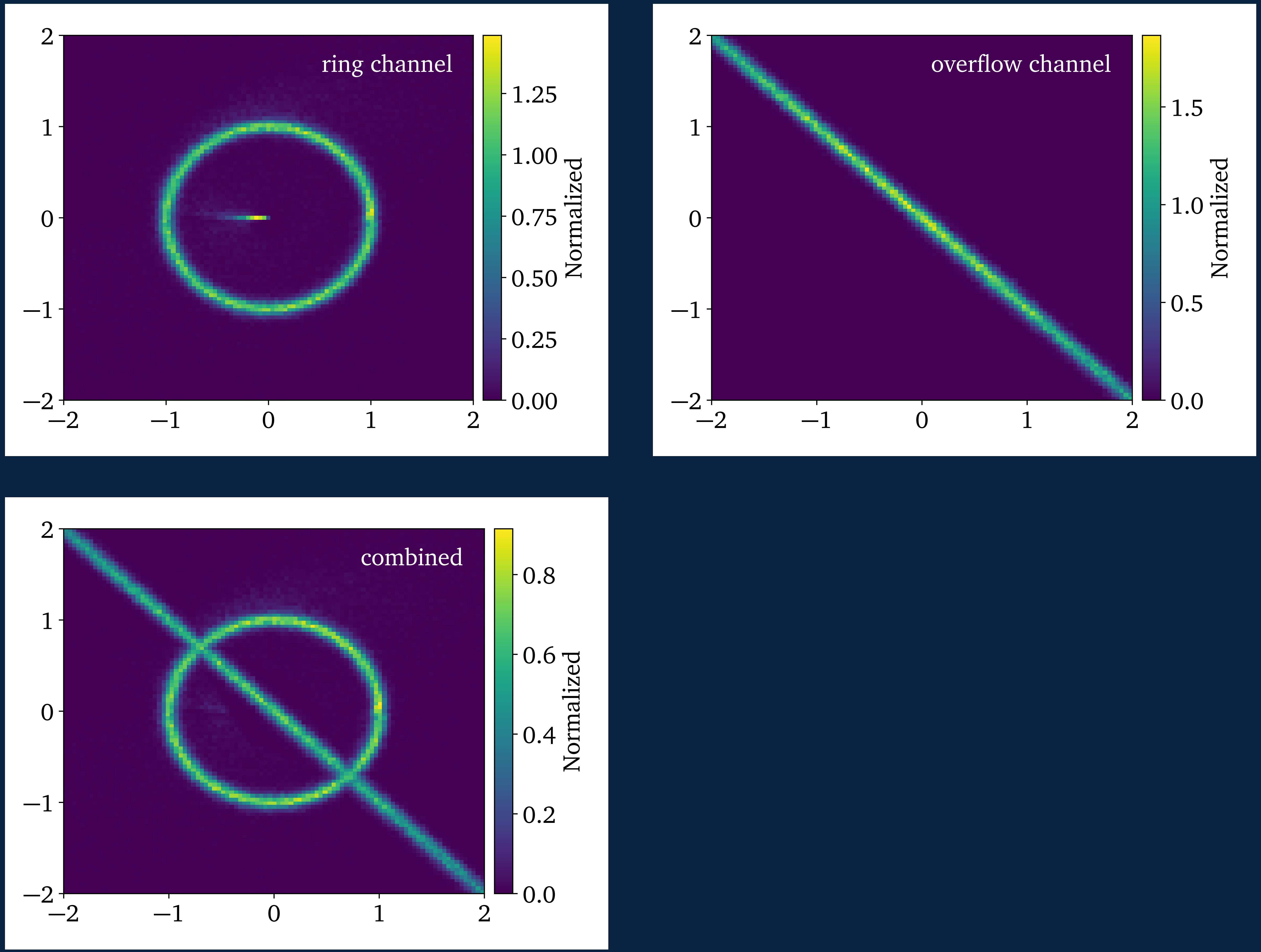


Example I: Crossed Ring

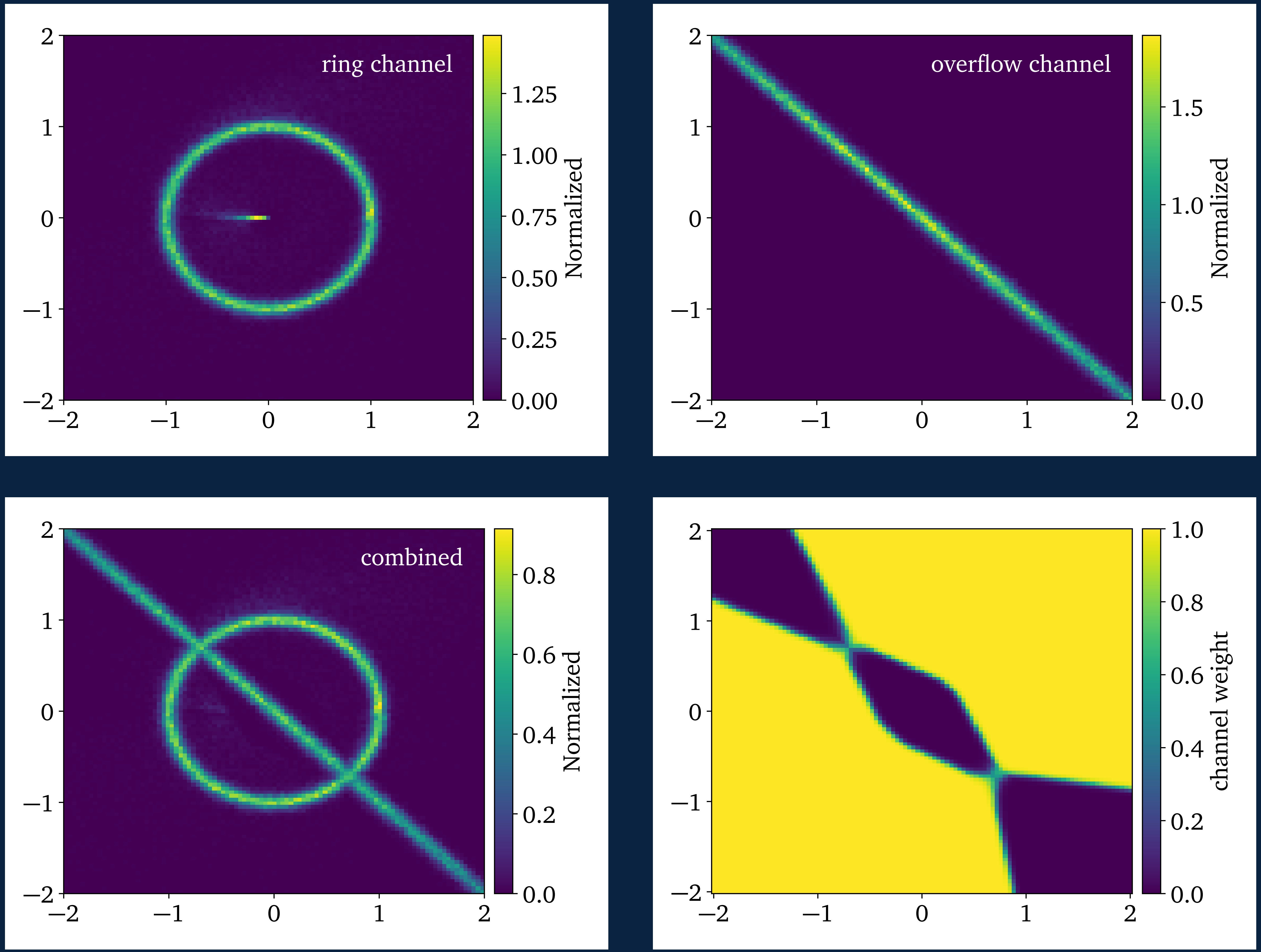




Example I: Crossed Ring



Example I: Crossed Ring



Rel. error:
0.37 ± 0.05

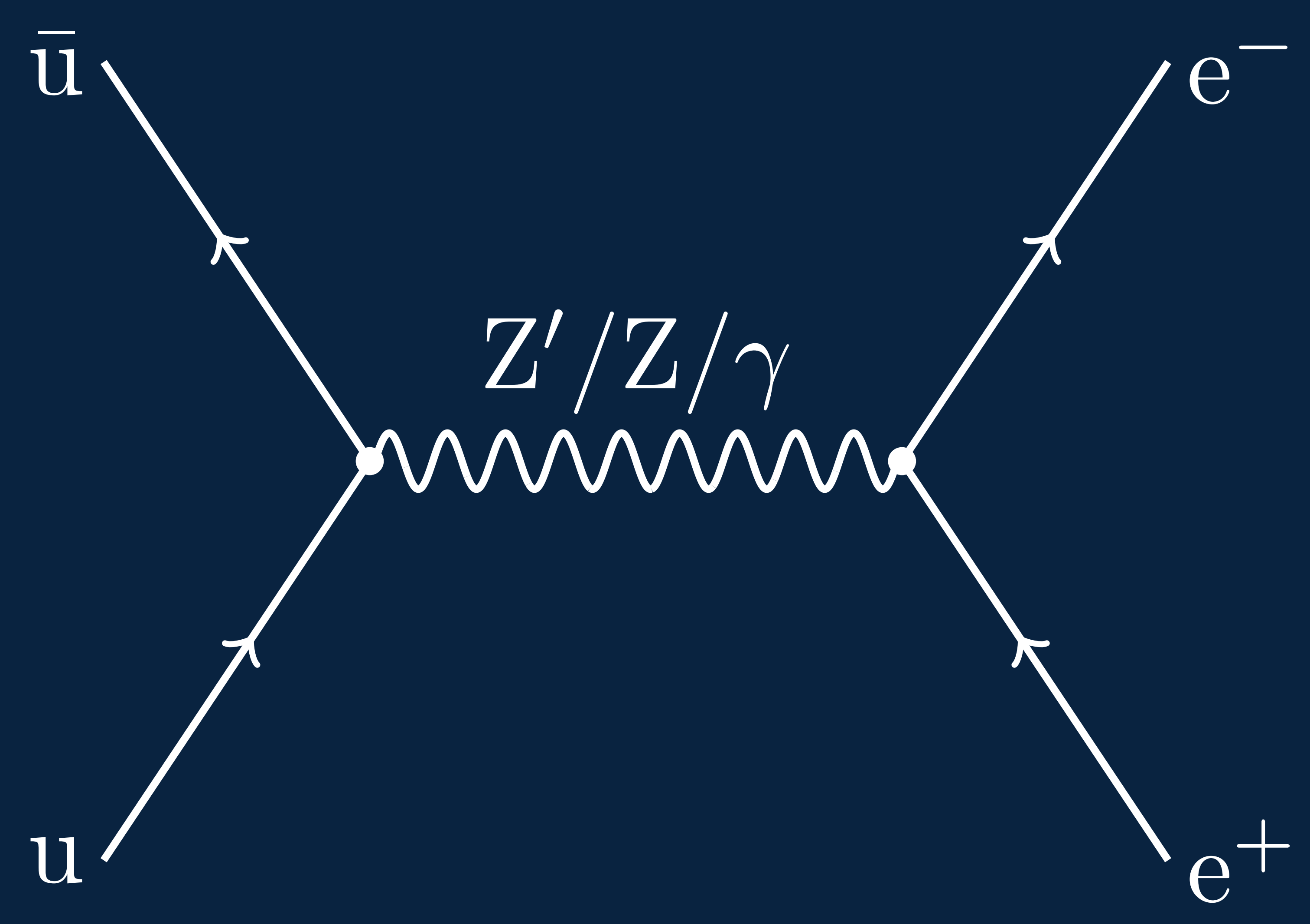
Example II

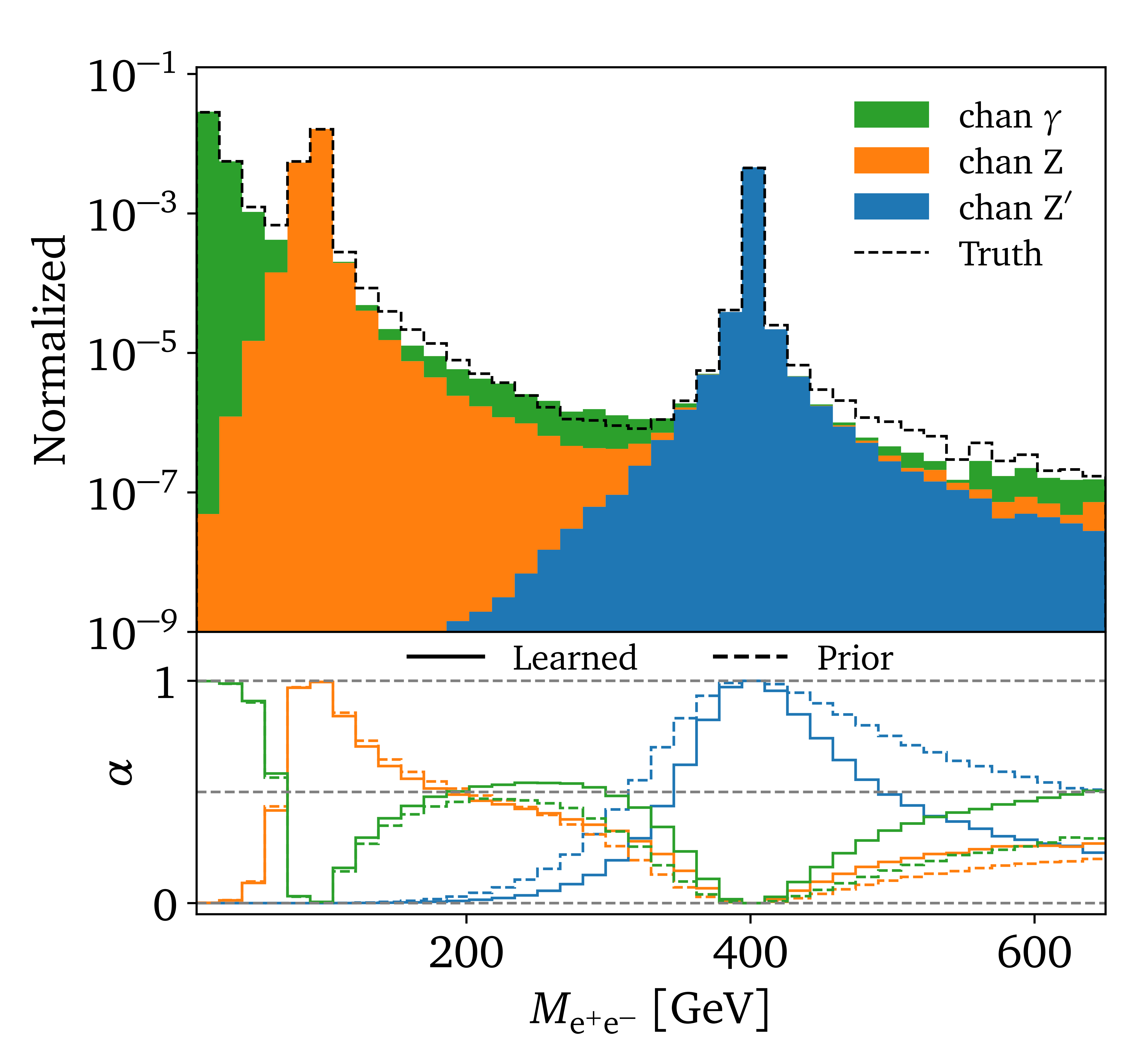
Drell-Yan + Z’

Example II: Drell-Yan + Z'

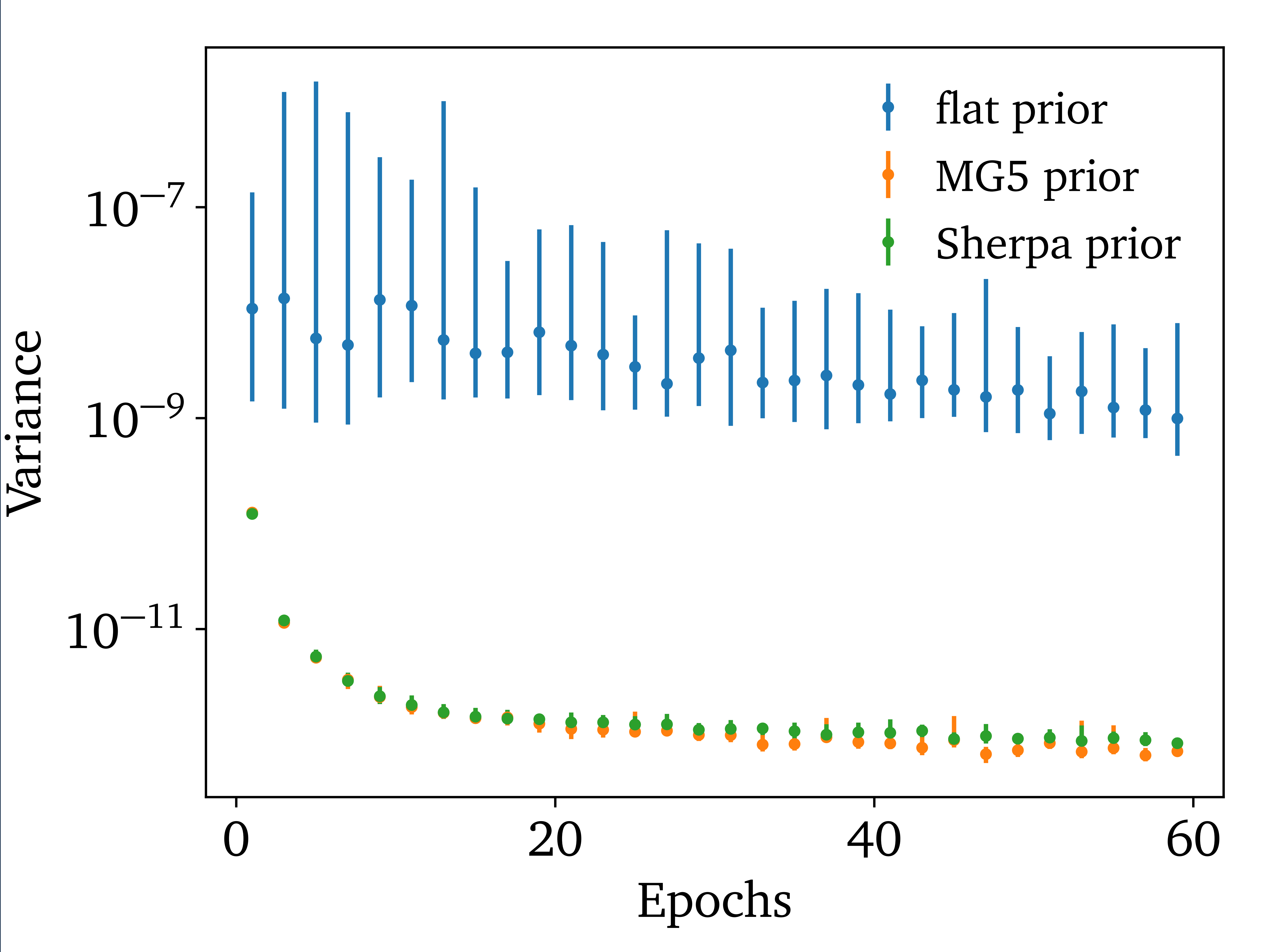
Implementation

- Custom amplitude in TensorFlow2
- Custom PS mappings in TensorFlow2
- PDFs from LHAPDF [1412.7420]

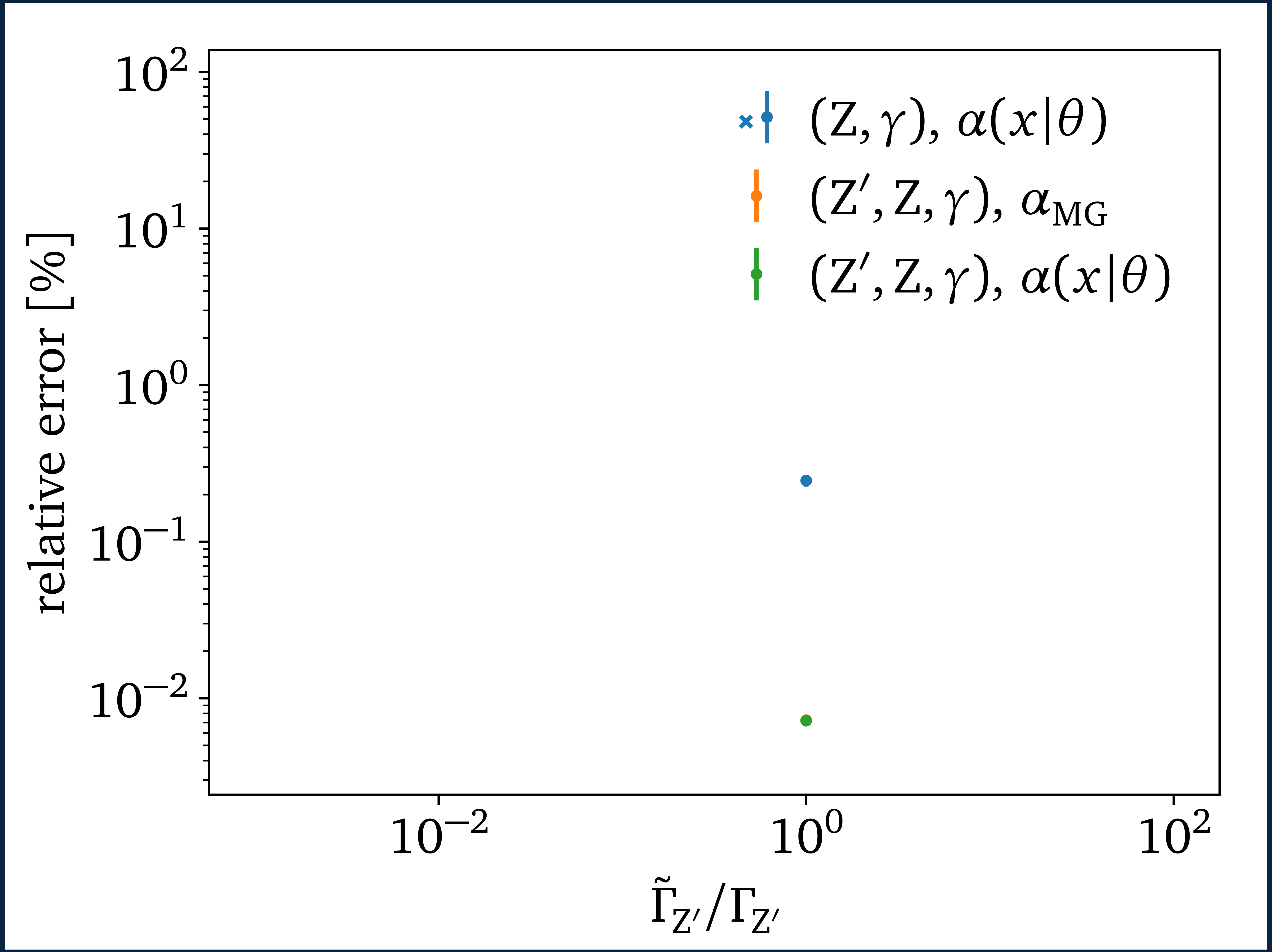




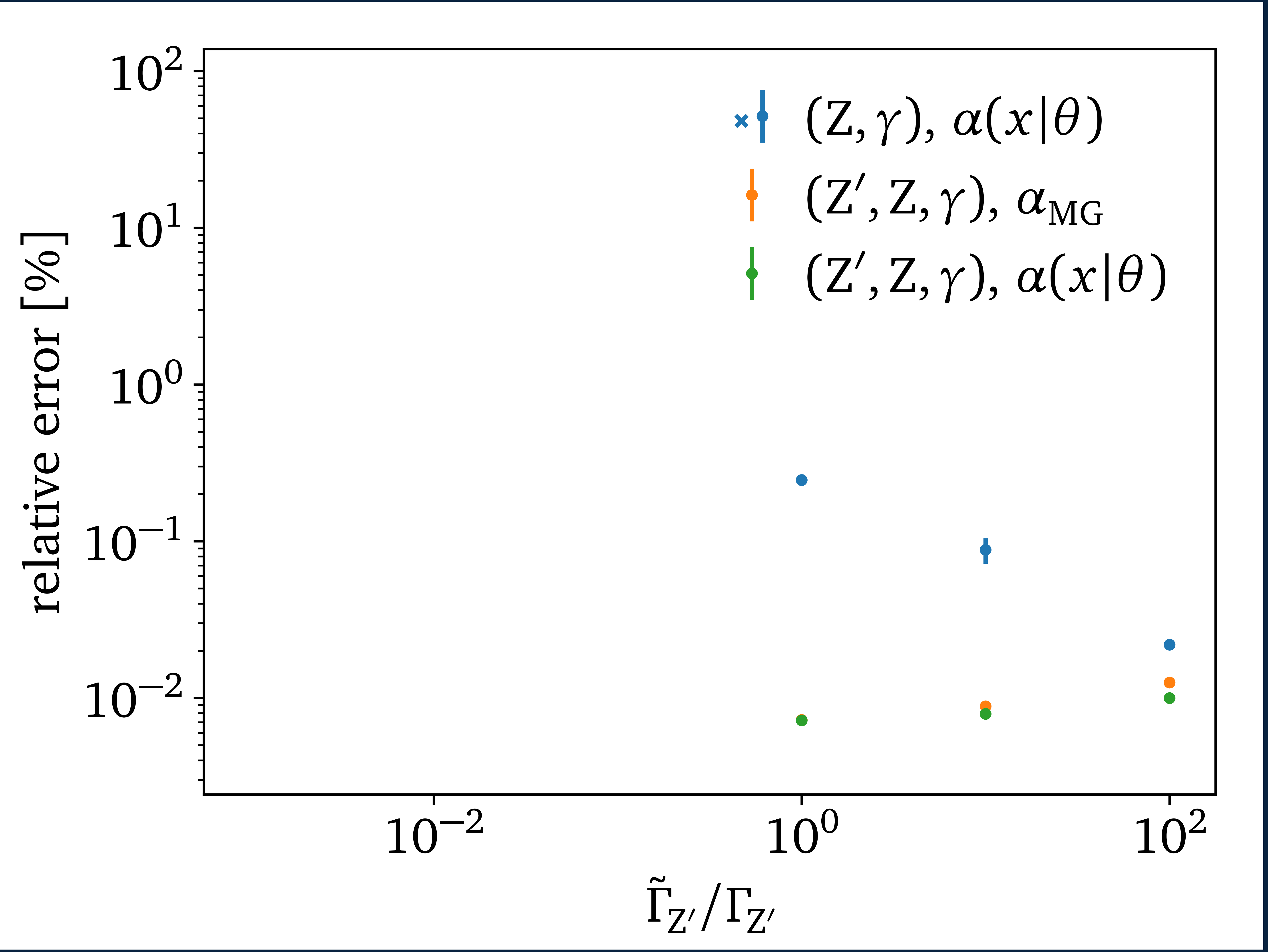
Example II: Drell-Yan + Z’



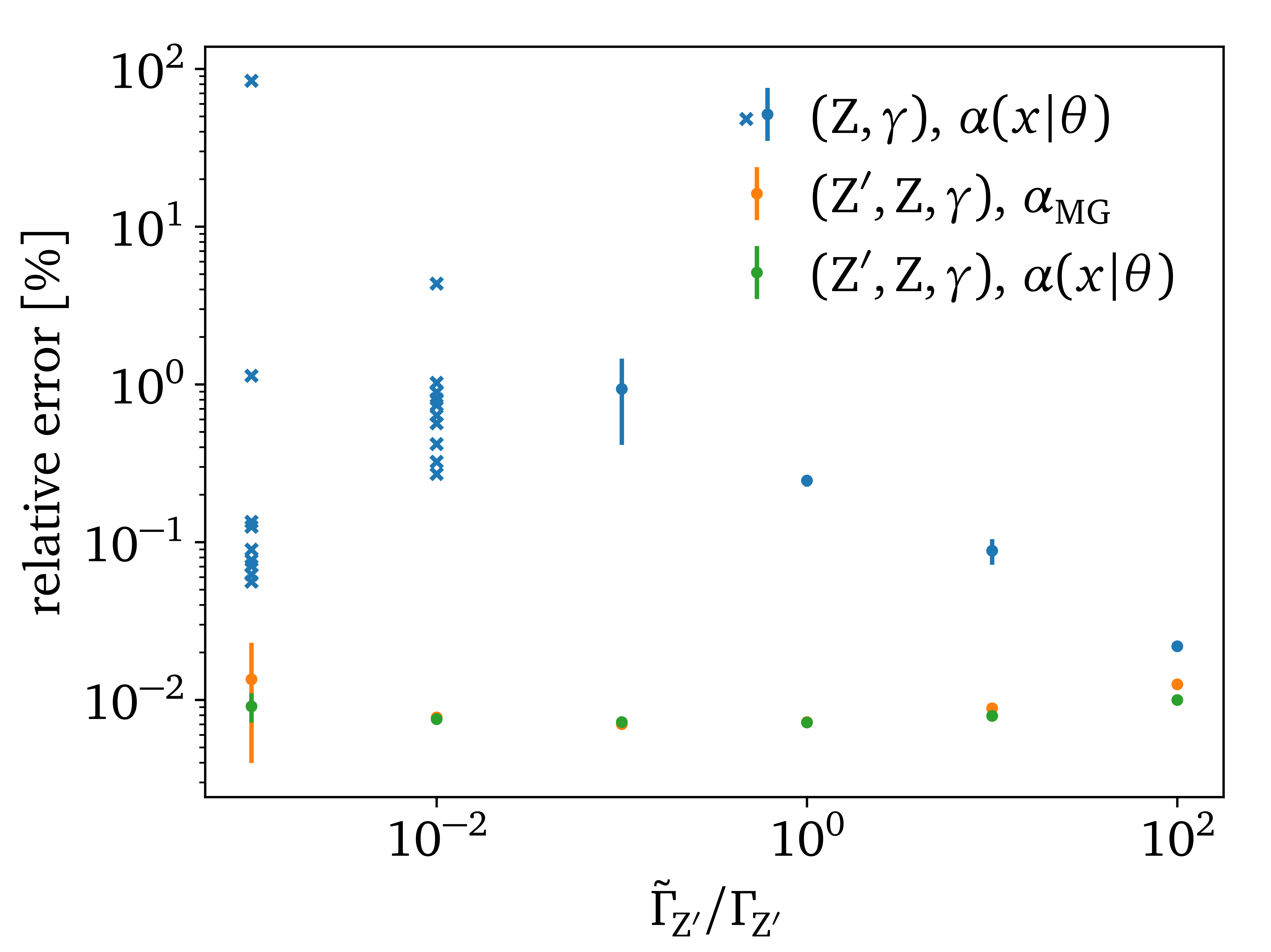
Example II: Drell-Yan + Z'



Example II: Drell-Yan + Z’

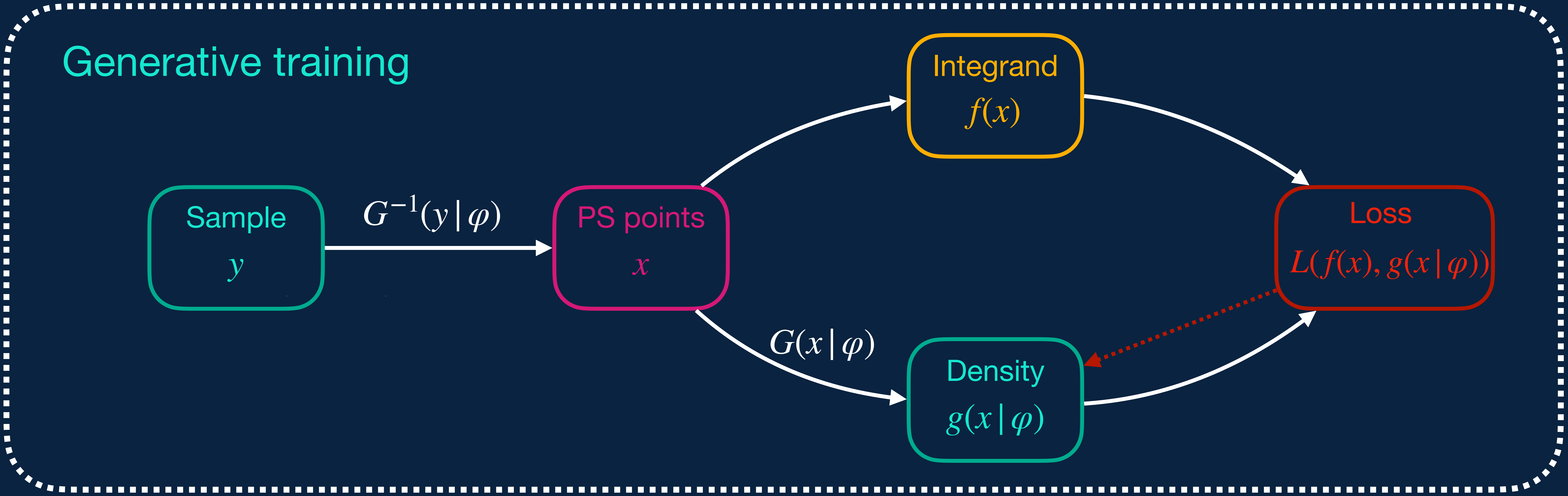


Example II: Drell-Yan + Z’

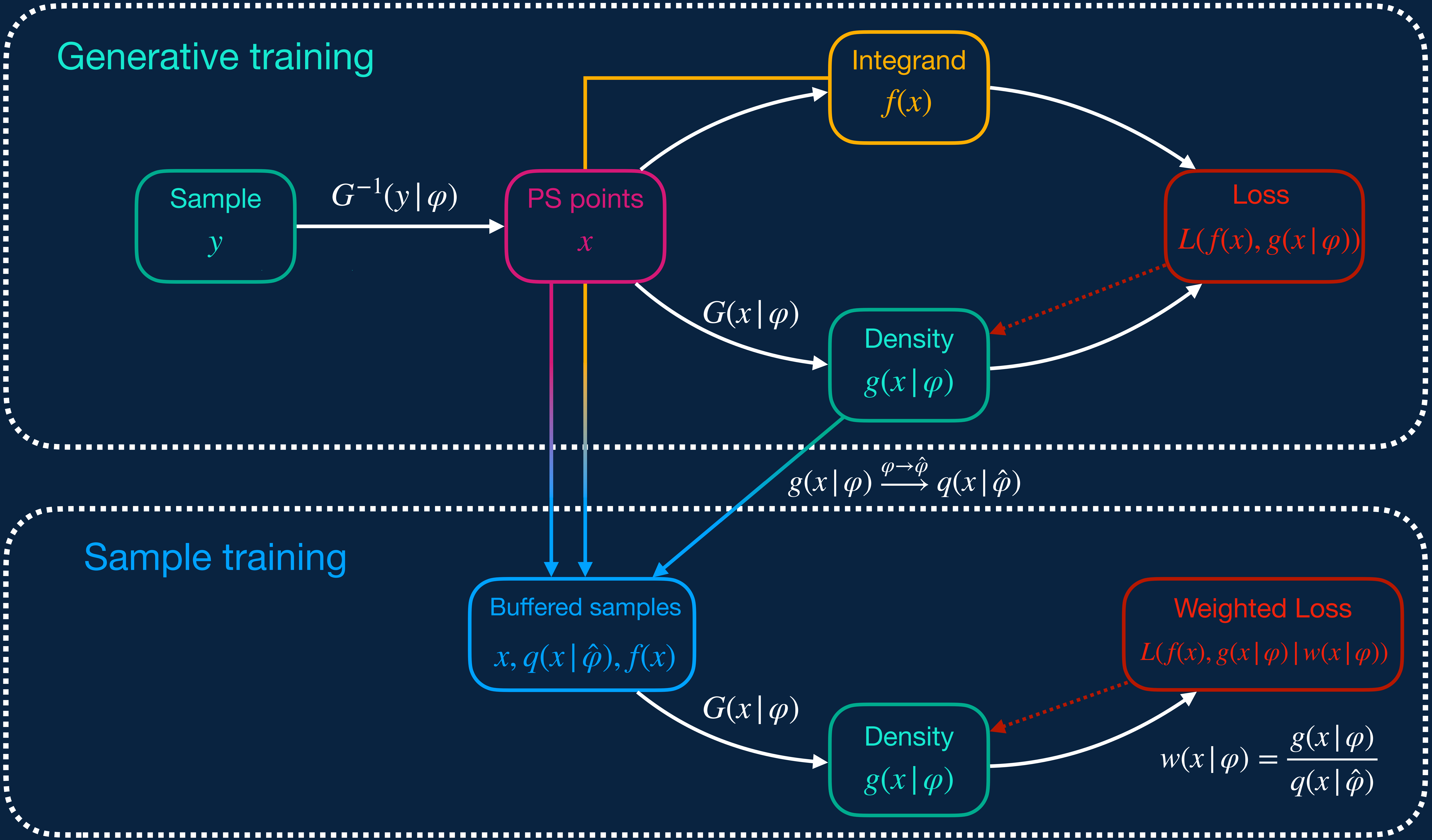


MadNIS

Buffered Training

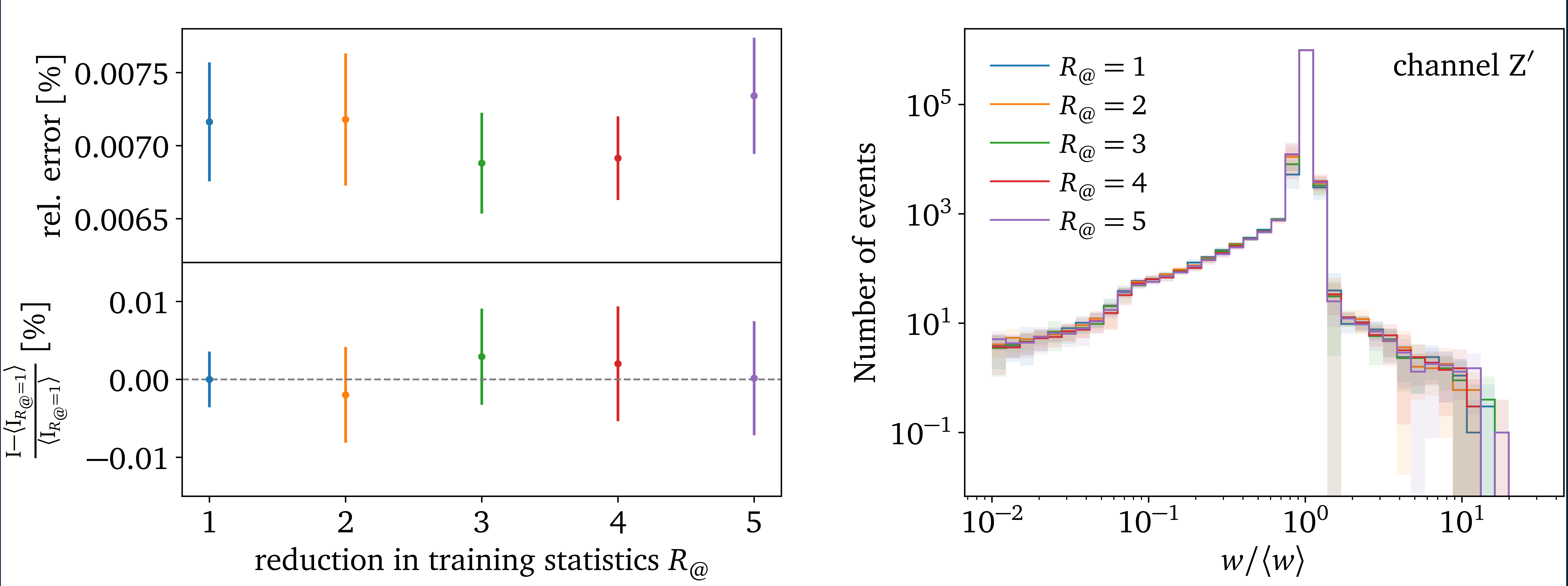


Generative training
Sample
y
$G^{-1}(y|\varphi)$
PS points
x
Integrand
$f(x)$
$G(x|\varphi)$
Density
$g(x|\varphi)$
Loss
$L(f(x), g(x|\varphi))$

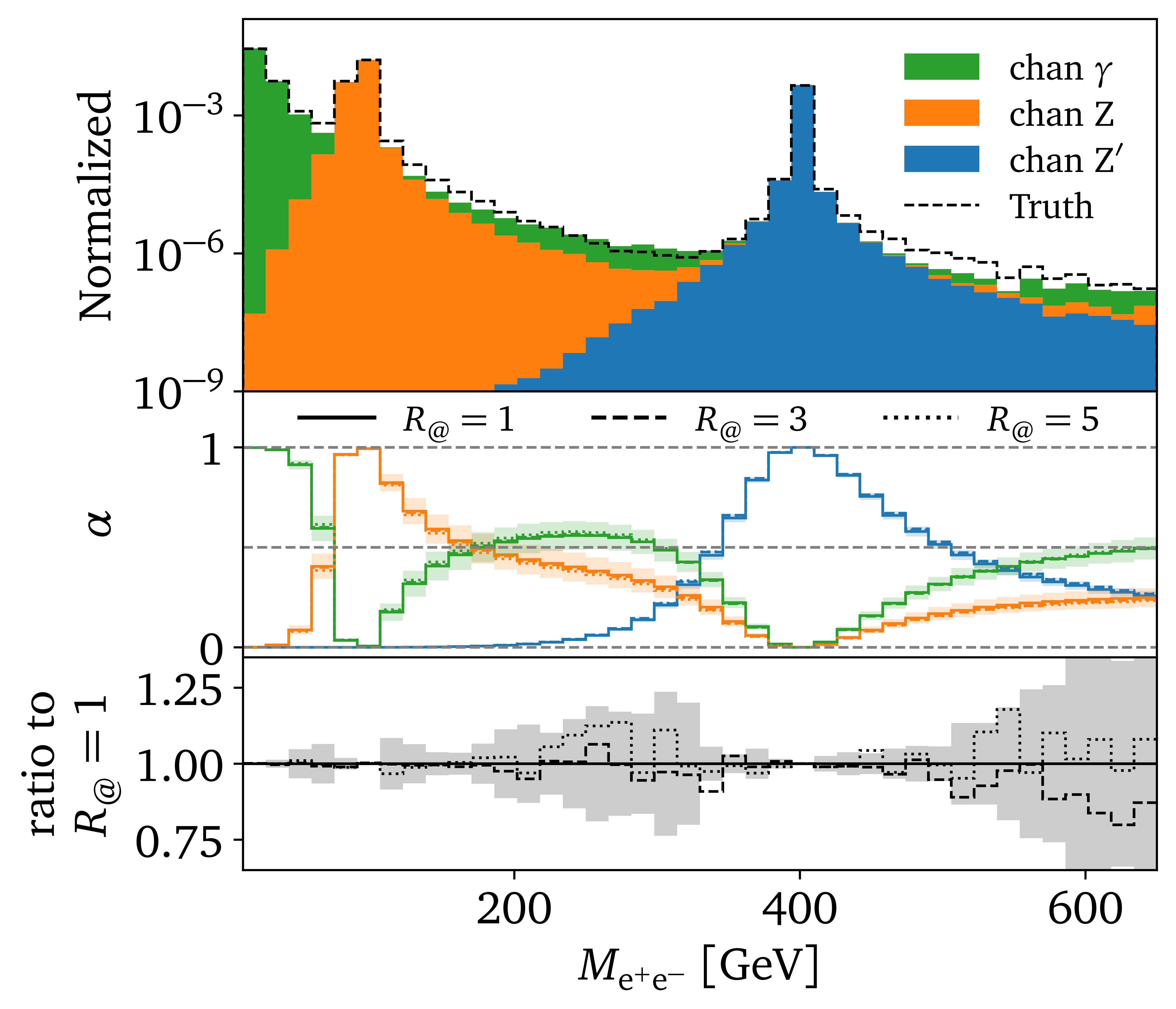


Generative training
Sample
y
$G^{-1}(y\,|\,\varphi)$
PS points
x
Integrand
$f(x)$
Loss
$L(f(x), g(x\,|\,\varphi))$
$G(x\,|\,\varphi)$
Density
$g(x\,|\,\varphi)$
$g(x\,|\,\varphi) \xrightarrow{\varphi \to \hat{\varphi}} q(x\,|\,\hat{\varphi})$
Sample training
Buffered samples
$x, q(x\,|\,\hat{\varphi}), f(x)$
Weighted Loss
$L(f(x), g(x\,|\,\varphi)\,|\,w(x\,|\,\varphi))$
$G(x\,|\,\varphi)$
Density
$g(x\,|\,\varphi)$
$w(x\,|\,\varphi) = \frac{g(x\,|\,\varphi)}{q(x\,|\,\hat{\varphi})}$

Buffered Training



Buffered Training



Summary and Outlook

Summary

- Channel mappings are important
- Multi-channel is more efficient when trained simultanously with the flow
- Buffered training reduces computational overhead and preserves precision

Summary and Outlook

Summary

- Channel mappings are important
- Multi-channel is more efficient when trained simultanously with the flow
- Buffered training reduces computational overhead and preserves precision

Outlook

- Implementation of MadNIS into MadGraph
- Test performance on real LHC examples: (eg. multi-leg, NLO, complicated cuts, …)
- Make everything run on the GPU and differentiable [MadJax 2203.00057]

Summary and Outlook

Summary

- Channel mappings are important
- Multi-channel is more efficient when trained simultanously with the flow
- Buffered training reduces computational overhead and preserves precision

Outlook

- Implementation of MadNIS into MadGraph
- Test performance on real LHC examples: (eg. multi-leg, NLO, complicated cuts, …)
- Make everything run on the GPU and differentiable [MadJax 2203.00057]

Stay tuned: Paper on arxiv tomorrow!! [14.12.2022]