Differentiable MadNIS-Lite

Theo Heimel
November 2024

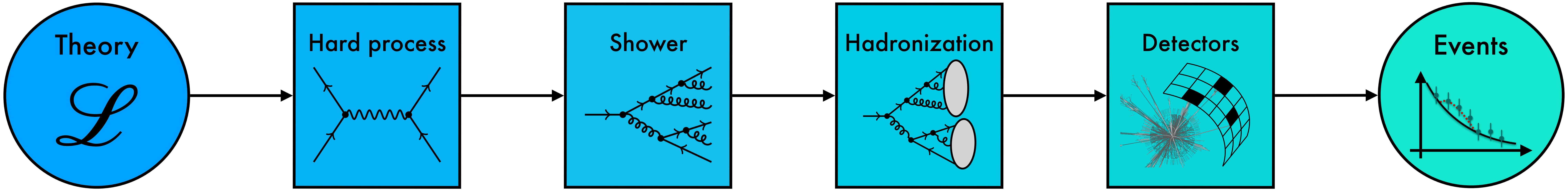
CP3, UCLouvain

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

[2408.01486] TH, Mattelaer, Plehn, Winterhalder

Differentiable programming @ LHC

Applications of **fully differentiable LHC simulation chain**



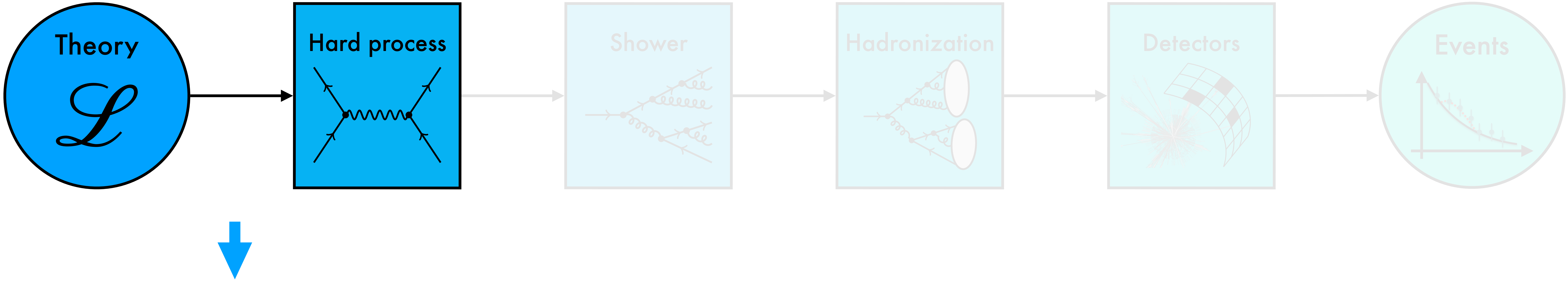
More sample-efficient SBI training in MadMiner
[1805.12244]

Tune model parameters directly on data
[see also MLHAD]

Optimized detector design
[MODE collaboration]

Differentiable programming @ LHC

Applications of **fully differentiable LHC simulation chain**

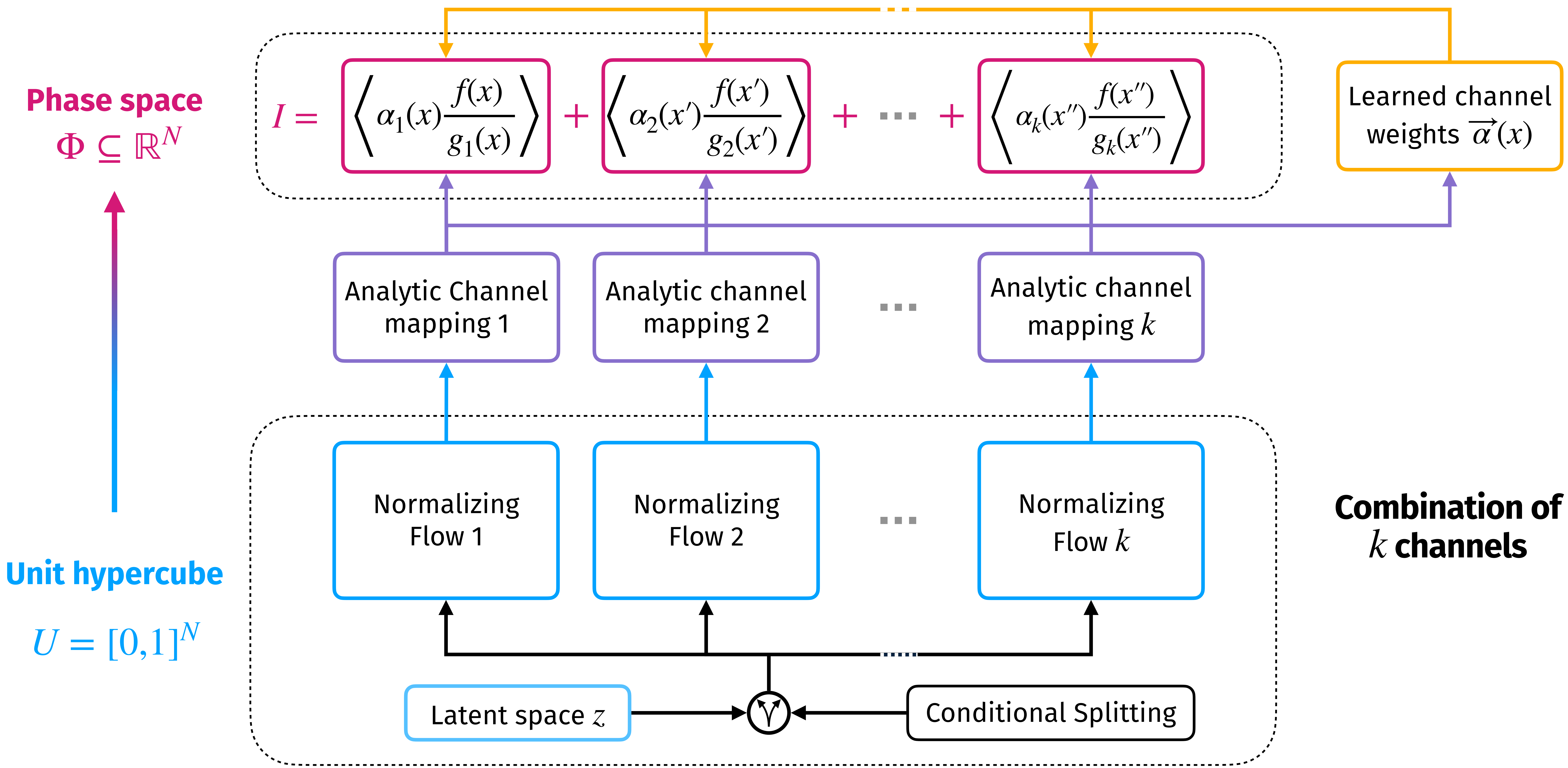


Previous study on differentiable matrix elements: MadJax [2203.00057] → promising results

①
Can our MadNIS training profit from gradient information?

②
Can we use differentiability to improve phase space mappings?

MadNIS: Neural Importance Sampling



Differentiable MadNIS

Phase space

$\Phi \subseteq \mathbb{R}^N$

$$\frac{d\sigma}{d\Phi} = \frac{1}{\text{flux}} f(x_a) f(x_b) \langle \, |M_{\lambda,c,\ldots}(p_a, p_b \mid p_1, \ldots, p_n)|^2 \, \rangle$$

Analytic Channel mapping

Differentiable phase space

Differentiable parton densities

Differentiable matrix elements

Normalizing Flow

Fully differentiable MadGraph using PyTorch

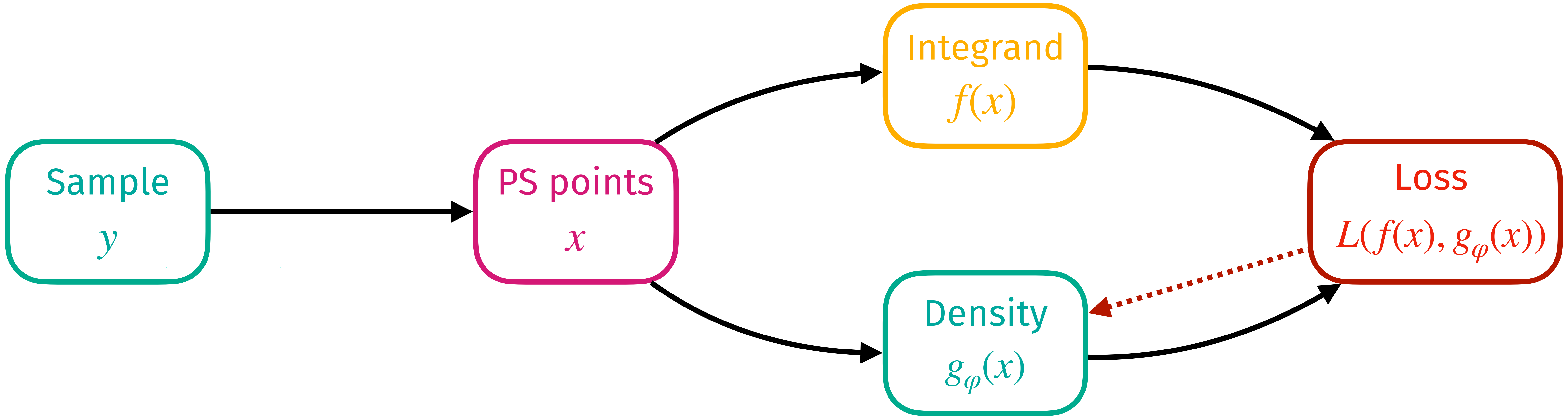
Unit hypercube

$U = [0,1]^N$

Based on previous implementations in Jax and TF:
PDFFlow [2009.06635], MadFlow [2106.10279],
MadJax [2203.00057]

Latent space z

MadNIS Training



Forward loss

Regular MadNIS training

$$L_F^{\text{fw}} = \left\langle \frac{g_\theta(x)}{q(x)} F\left(\frac{f(x)}{g_\theta(x)}\right)\right\rangle_{x\sim q(x)}$$

Inverse loss

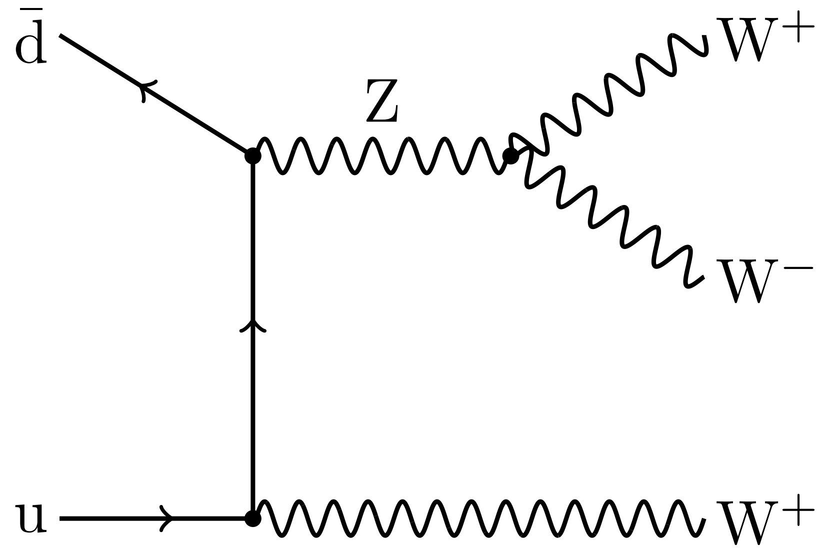
Only possible with differentiable integrand

$$L_F^{\text{inv}} = \left\langle F\left(\frac{f(\overline{G}_\theta(z))}{\overline{g}_\theta(z)}\right)\right\rangle_{z\sim p_0(z)}$$

Arbitrary f-divergence: KL, RKL, variance

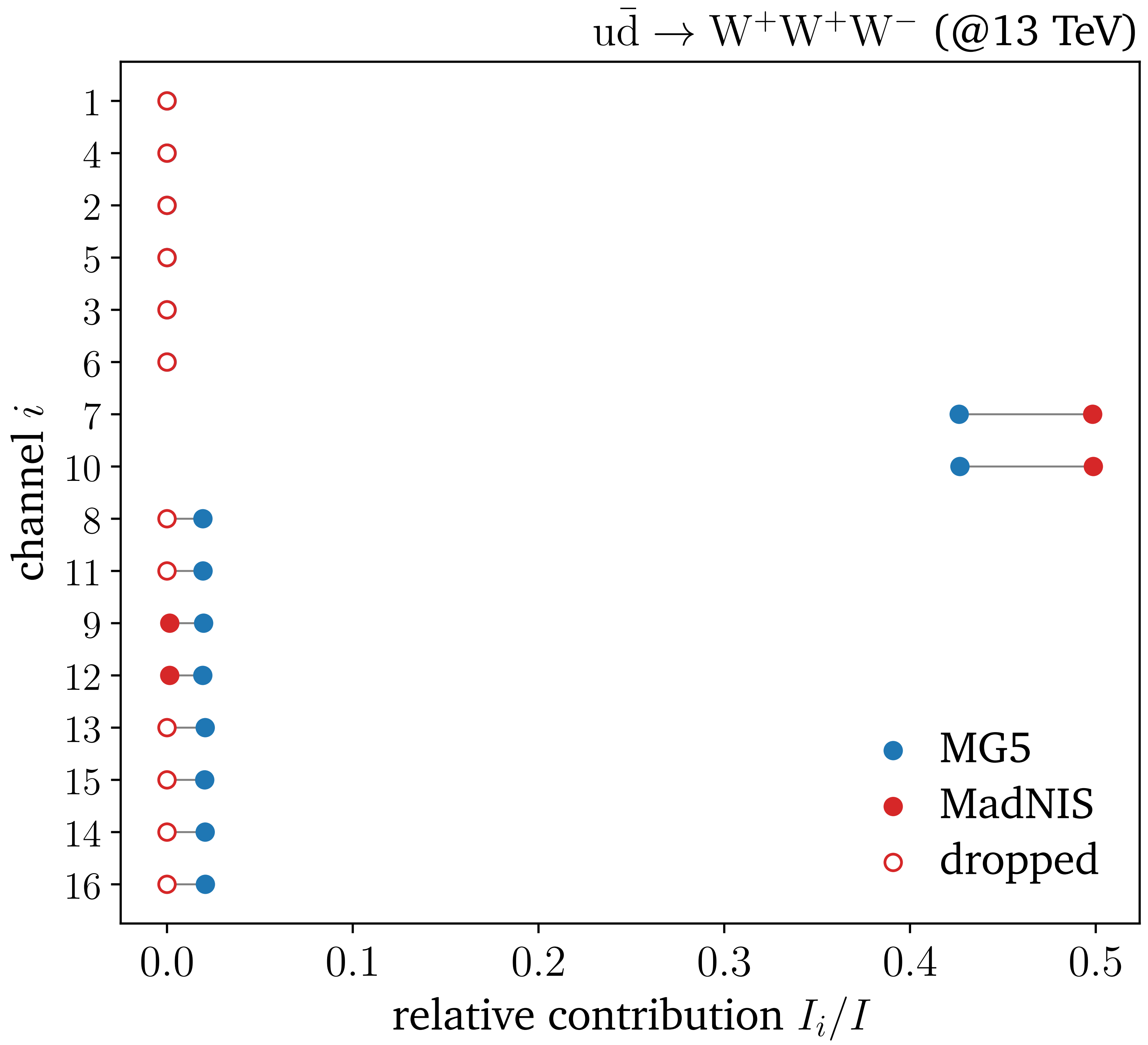
Example process

- Example process: $u\bar{d} \to W^+W^+W^-$

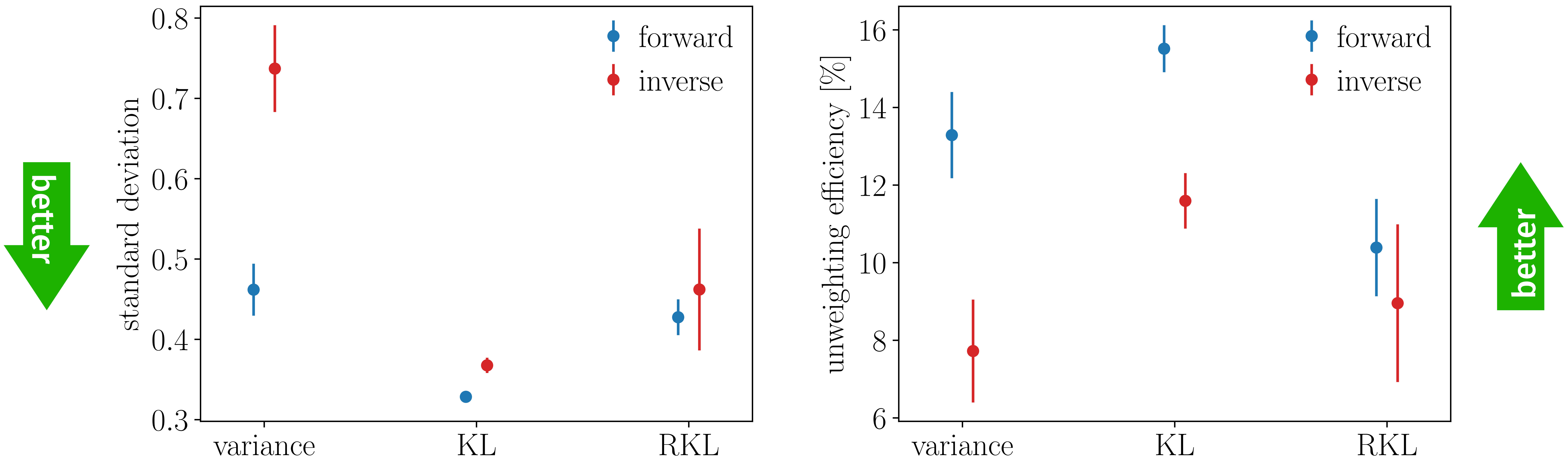


- Last MadNIS paper:
  single channel is sufficient

- Flat, differentiable phase space
  using RAMBO on diet

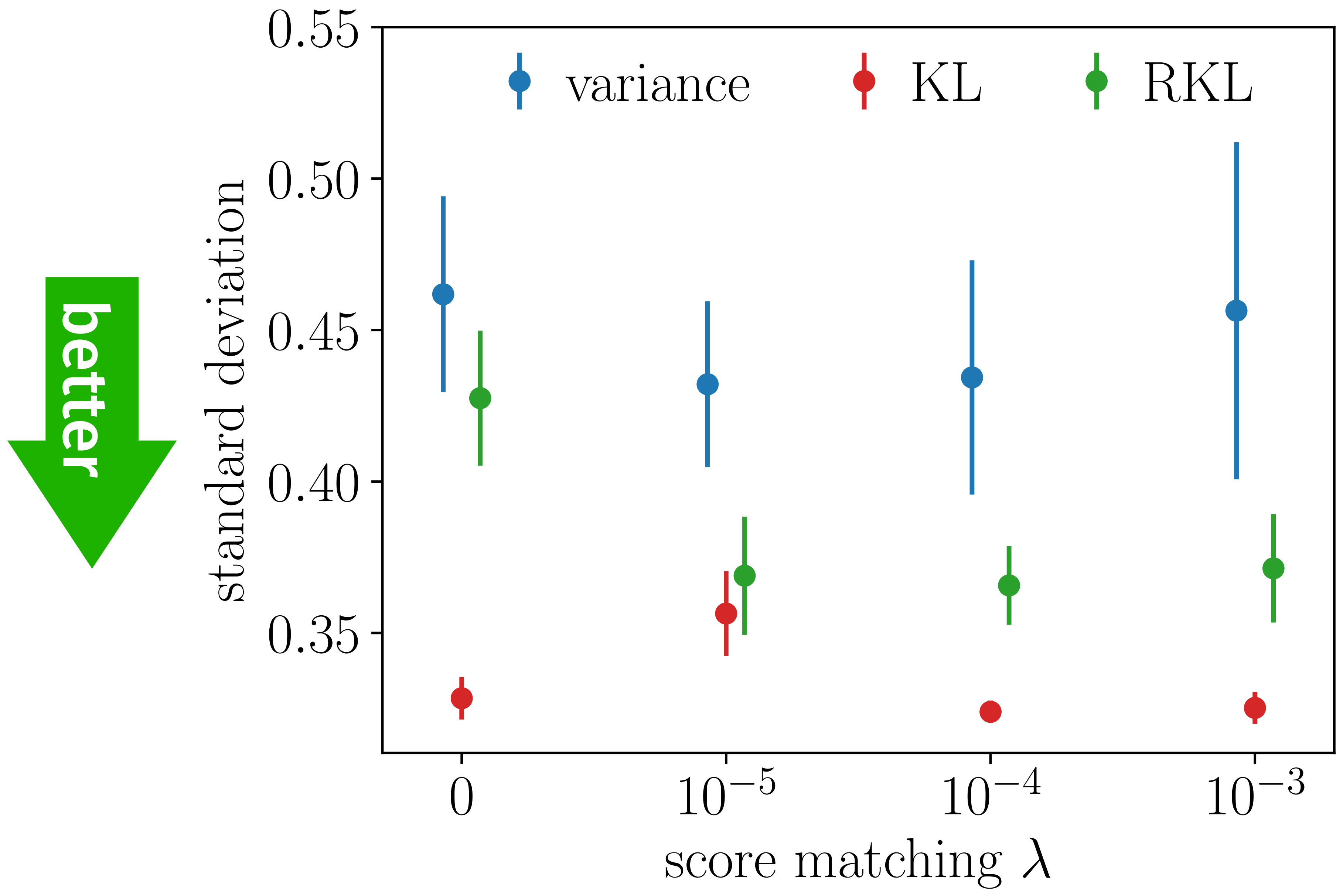


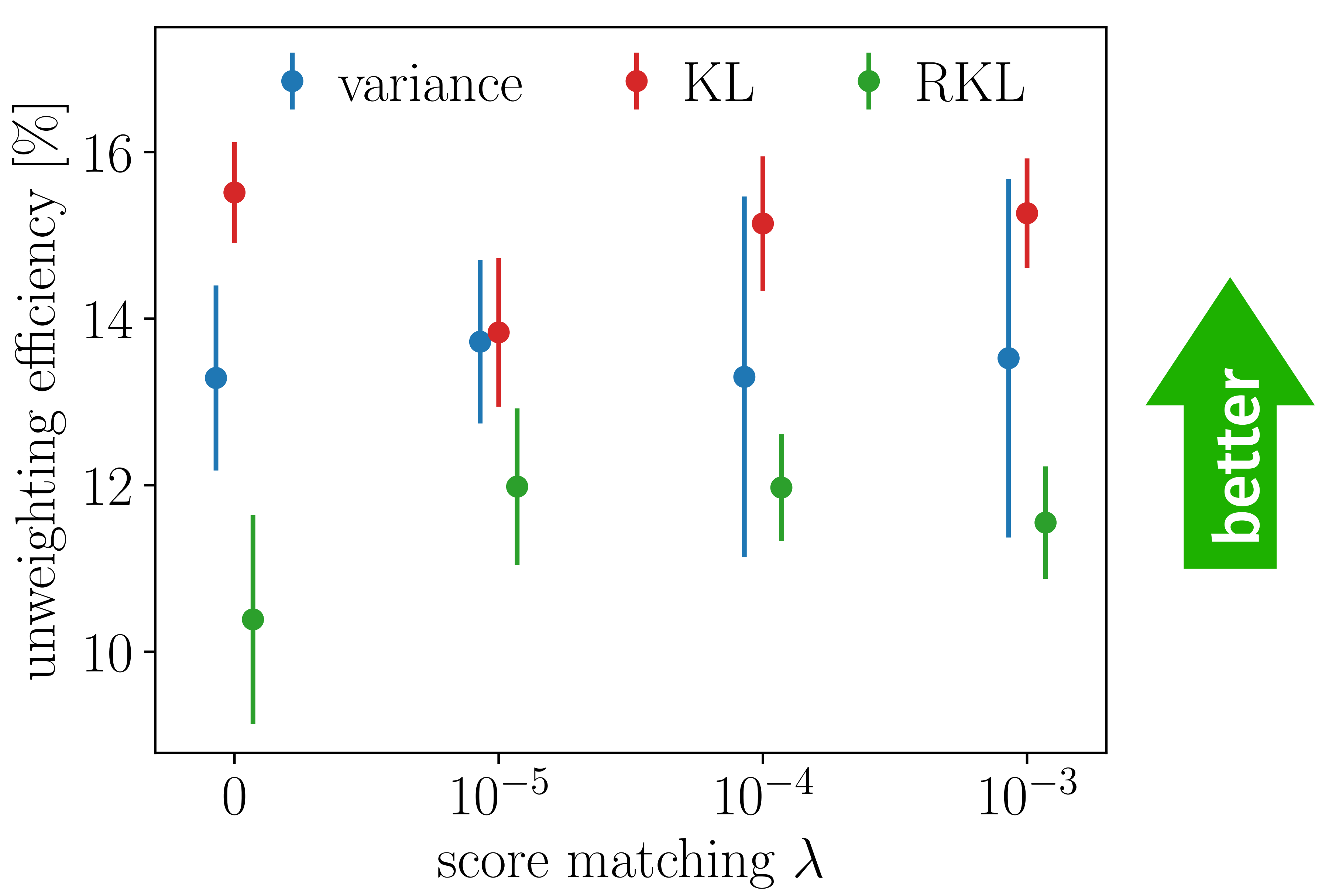
Forward vs inverse loss



- forward losses perform better than inverse losses
- Side result: KL performs better than variance,
  however not compatible with trainable channel mappings

Derivative matching loss



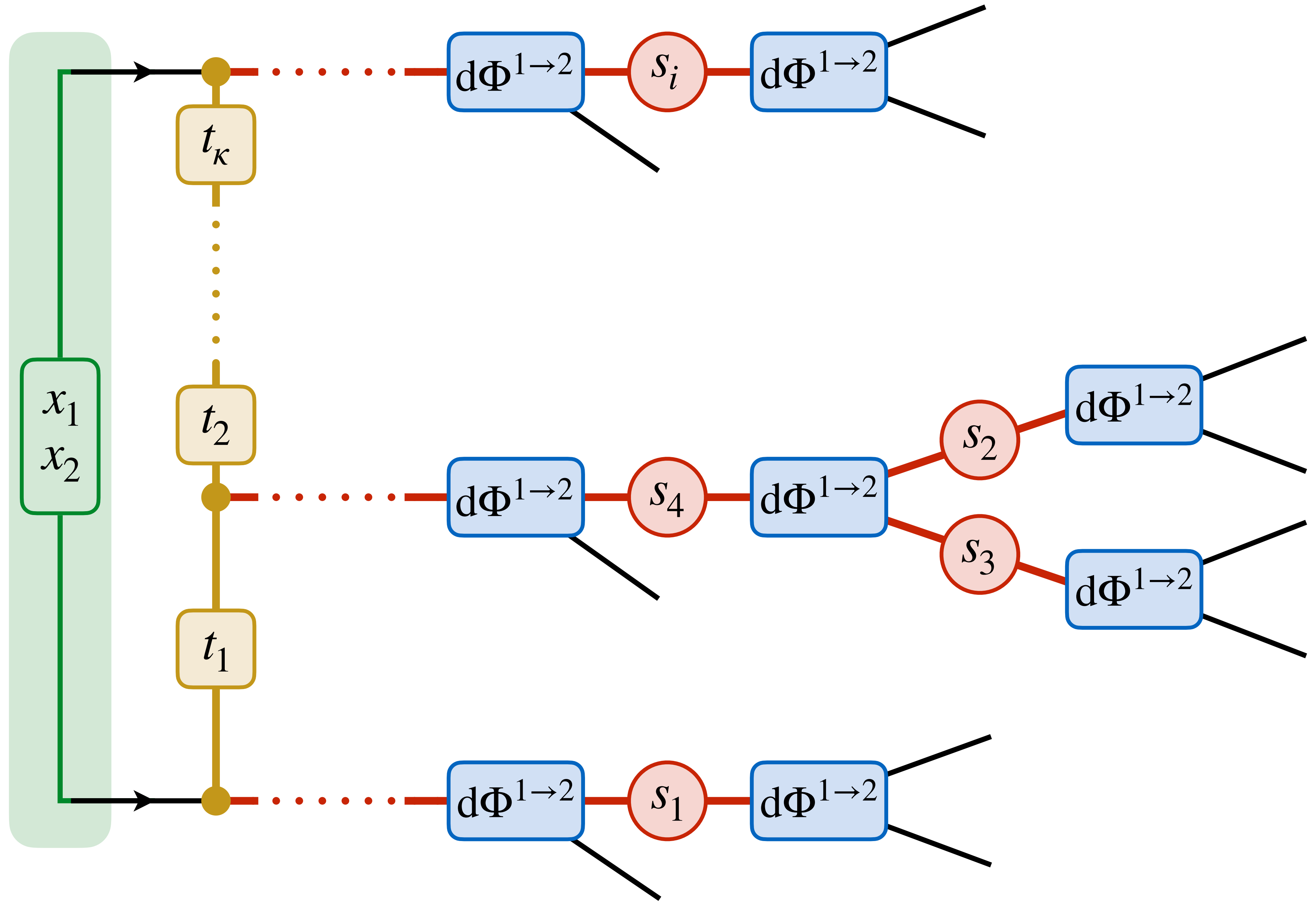


- Alternative use for gradients: derivative matching loss

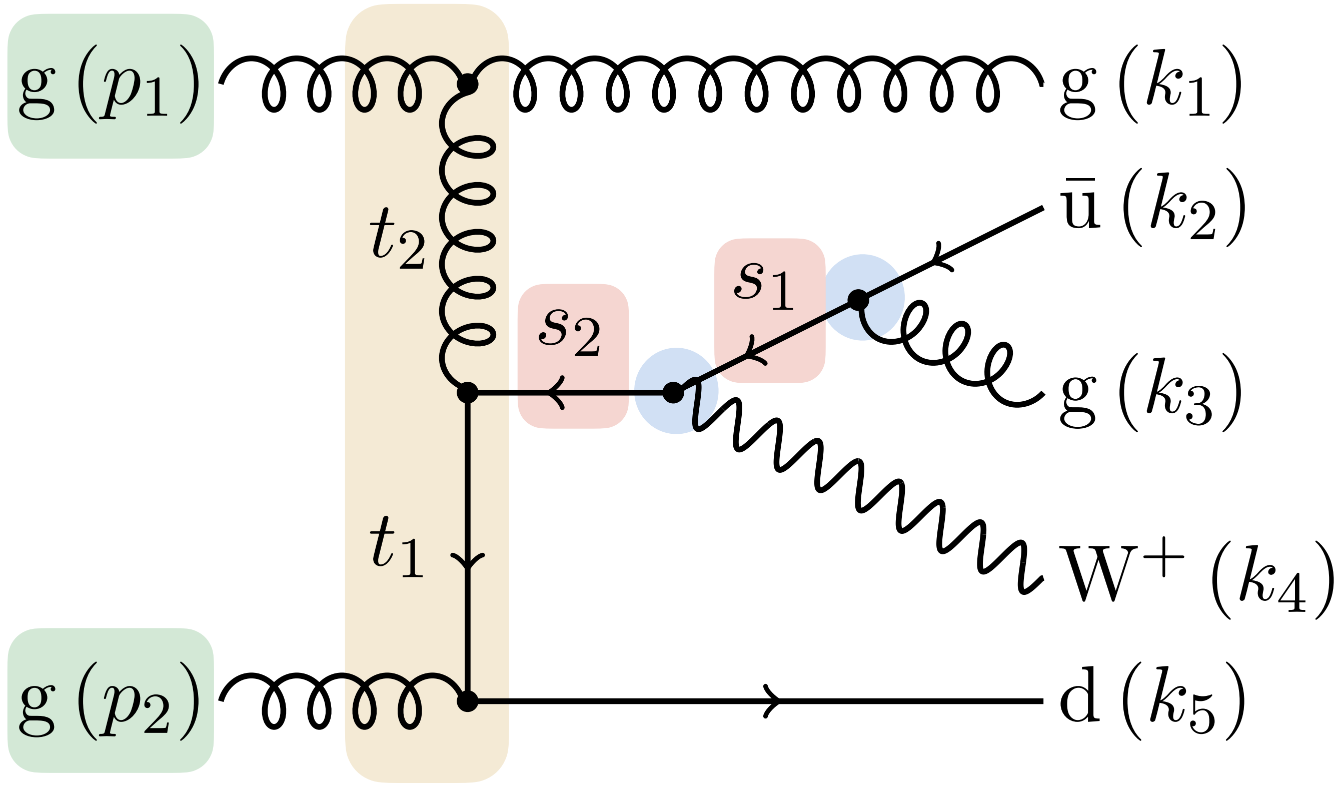
$$L^{\mathrm{fw}} \to L^{\mathrm{fw}} + \lambda \left\langle \left| \partial_x \log f(x) - \partial_x \log g_\theta(x) \right|^2 \right\rangle_{x\sim q(x)}$$

- Sometimes small improvements over normal forward loss
- Additional **cost of gradient evaluation not amortized**

MadNIS-Lite



- Standard construction of PS-mappings from Feynman diagrams

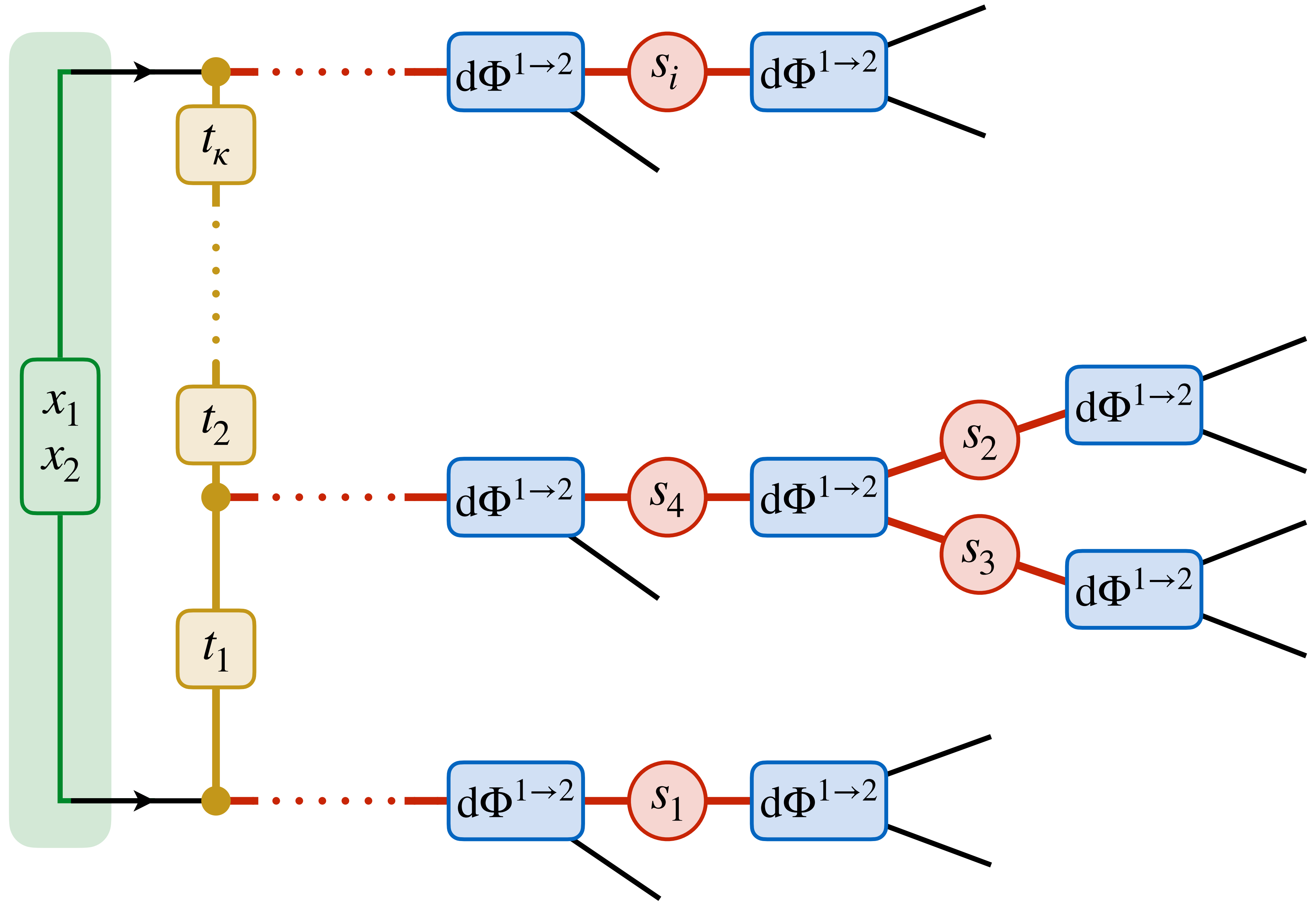


Example:
W+jets

- Phase space library based on PyTorch

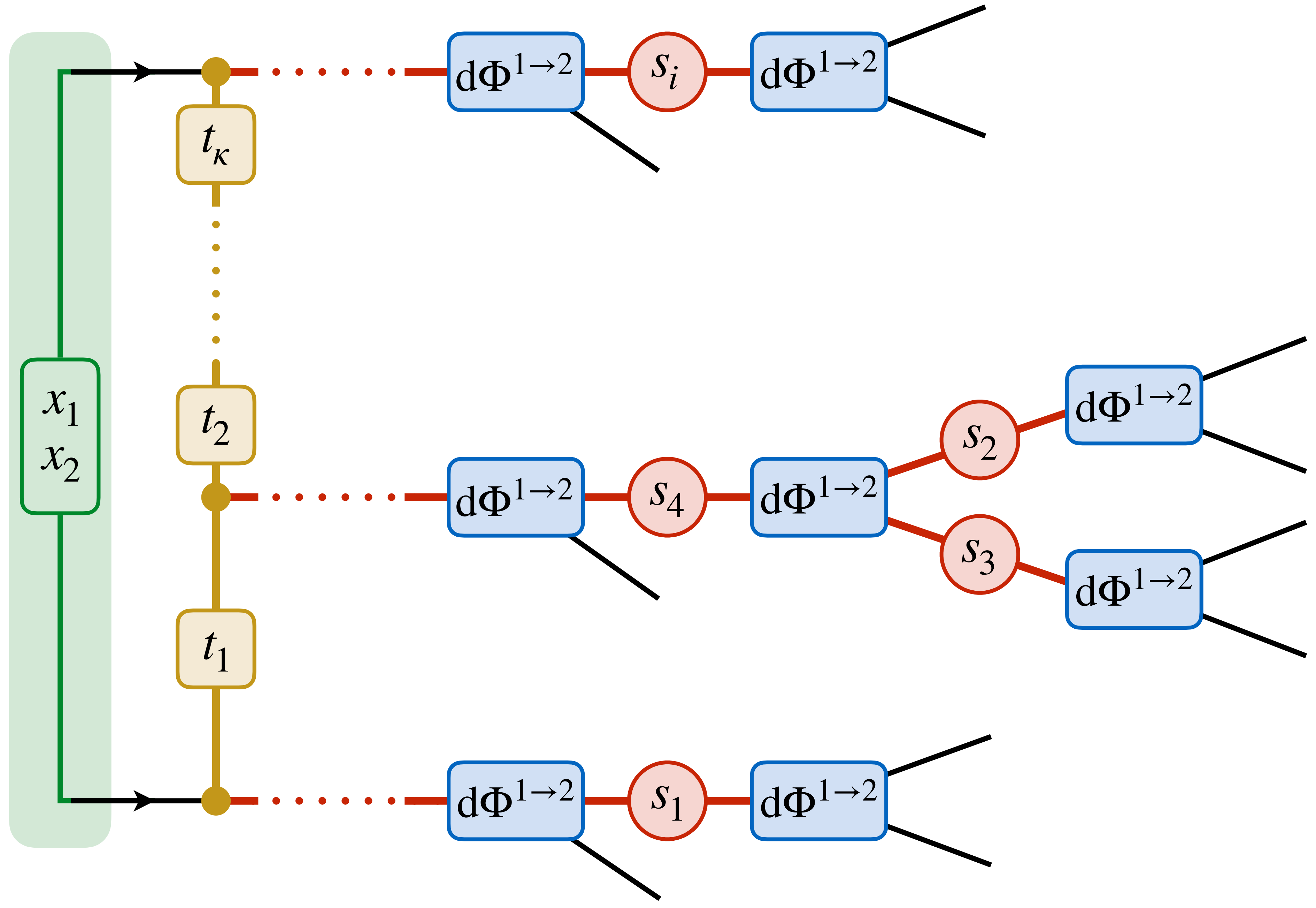
- **Fully differentiable + invertible**
  → can build in trainable components
  → include in MadNIS training

Trainable components



- Add small trainable components based on RQ spline transformations
- Condition on context
  → COM energy, decay energy, ...
- **Tiny number of parameter**: shared
  → between all components of same type
  → between all channels

Trainable components



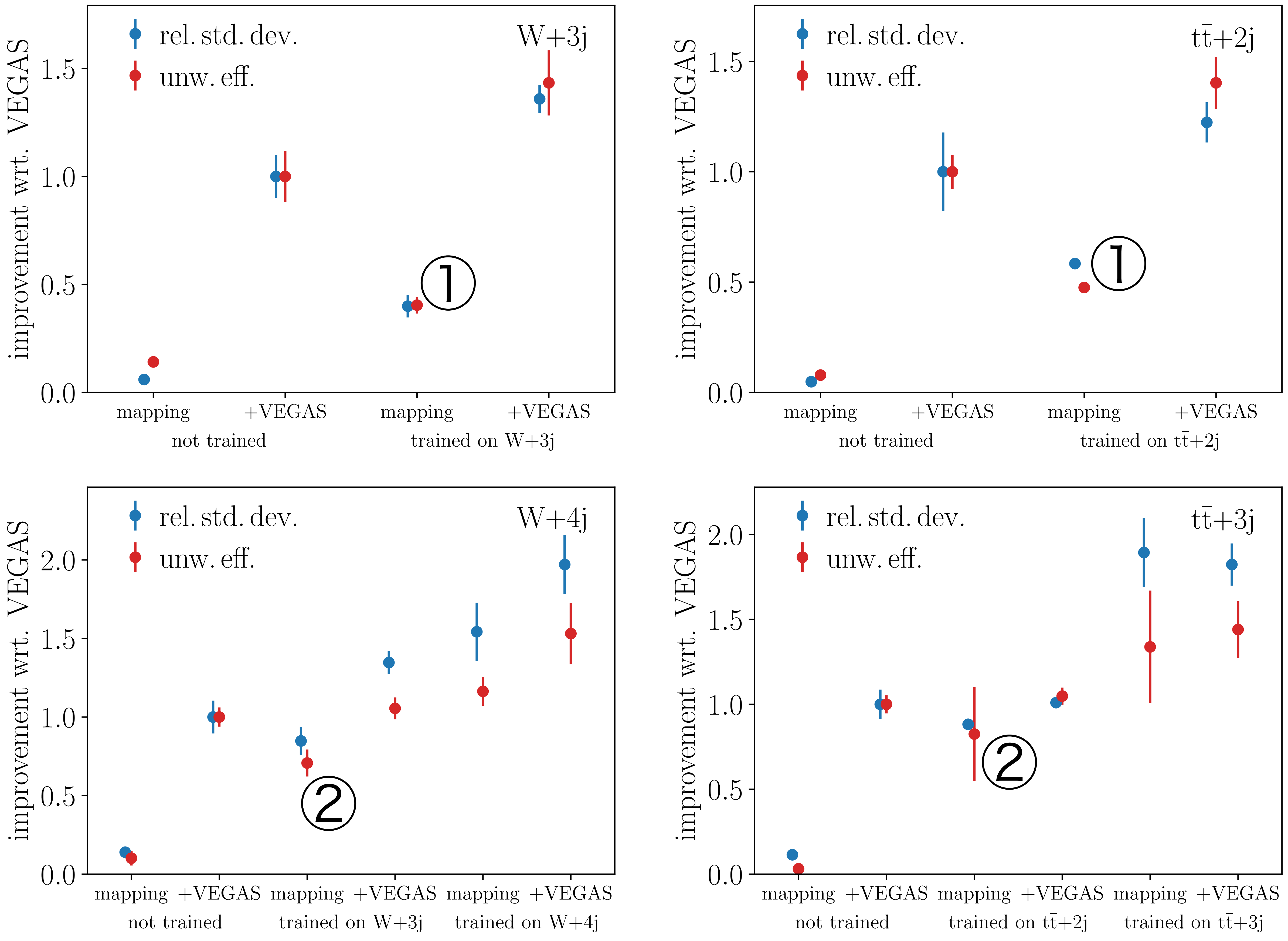
Mapping	Parameters	Conditions
Time-like invariants, Eqs.(39),(40) (separate for massless and massive propagators)	190	partonic CM energy $\sqrt{\hat{s}/s_{\text{lab}}}$ minimal decay CM energy $\sqrt{s_{\text{min}}/s_{\text{lab}}}$ maximal decay CM energy $\sqrt{s_{\text{max}}/s_{\text{lab}}}$
$2 \to 2$ scattering, Eq.(43)	798	correlations between z_t, z_ϕ partonic CM energy $\sqrt{\hat{s}/s_{\text{lab}}}$ scattering CM energy $\sqrt{p^2/s_{\text{lab}}}$ virtualities $\sqrt{k_{1,2}^2/s_{\text{lab}}}$
Time-like invariants for pseudo-particles, Eq.(45)	190	partonic CM energy $\sqrt{\hat{s}/s_{\text{lab}}}$ minimal energy $\sqrt{s_{\text{min}}/s_{\text{lab}}}$ maximal energy $\sqrt{s_{\text{max}}/s_{\text{lab}}}$
$1 \to 2$ decay, Eq.(46)	380	correlations between z_θ, z_ϕ partonic CM energy $\sqrt{\hat{s}/s_{\text{lab}}}$ decay CM energy $\sqrt{p^2/s_{\text{lab}}}$
PDF convolutions, Eq.(48)	114	correlations between z_τ, z_{x_1}

Total: 1672 parameters

Performance

W+jets

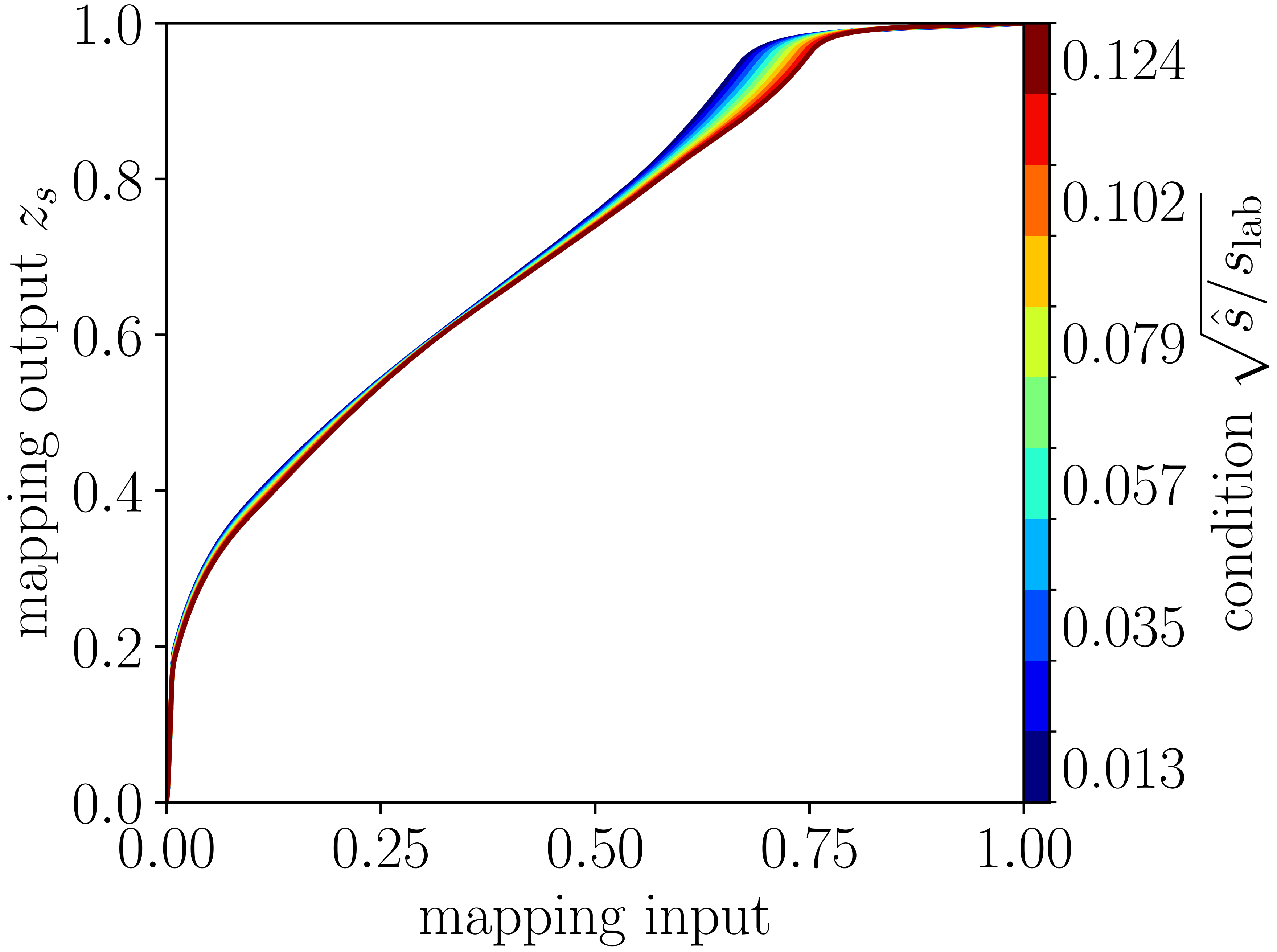
$t\bar{t}$+jets



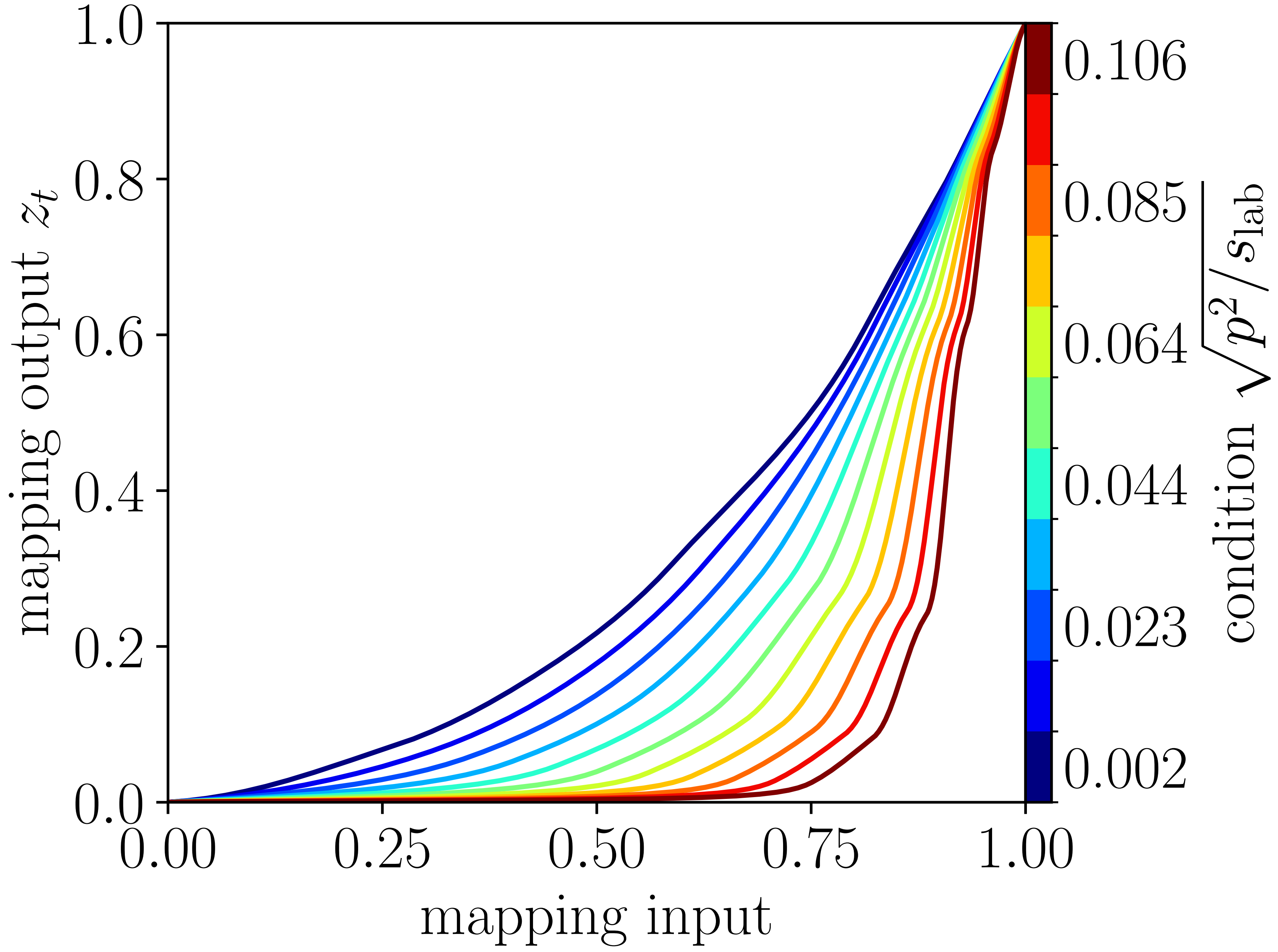
- good performance even though no channel-specific training ①
- **trained for n jets, used for n+1 jets**
  → performance like VEGAS ②
  → cheap training
- further improvements for VEGAS trained on top of MadNIS-Lite

Interpretability

Massless propagator
s-invariant



2→2 scattering
t-invariant



- s-invariant: small energy-dependence, easily learned by VEGAS, still room for improvement in underlying mapping
- t-invariant: large dependence on p^2

Outlook

- MadNIS training: only small benefits from differentiable ME
  → additional computational cost of gradients not amortized

- **MadNIS-Lite**: middle ground between VEGAS and MadNIS
  → **generalizes from n jets to n+1 jets**
  → interpretability to improve phase space mapping

- Many other applications of gradients, e.g. SBI, tuning, …
  → make **gradients easily available in future MadGraph versions**

$$\rightarrow \quad \partial\, 7 / \partial x$$