Neural Control Variates at LO and NLO

Rebecca Revelli - ITP Heidelberg

Theo Heimel, Tilman Plehn, Sophia Vent, Ramon Winterhalder



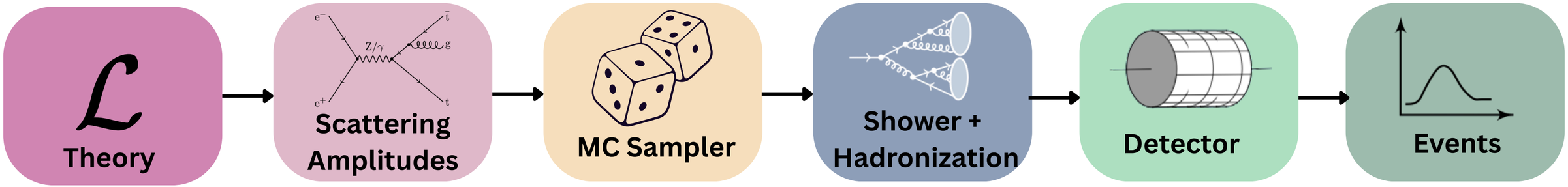
UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

IRN TERASCALE 2026 - 20.10.2026

IMPRS
for Precision Tests of Fundamental Symmetries
INTERNATIONAL MAX PLANCK RESEARCH SCHOOL

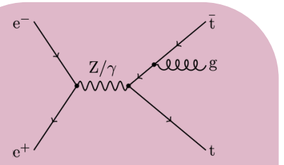


Motivation



NLO corrections in QCD are already expensive to compute
Future colliders and precision measurements demand **higher-order** predictions at even grater computational cost

The Integration Challenge



Scattering Amplitudes

$$\sigma = \int d\Phi_n \, |\mathcal{M}(\Phi_n)|^2$$

Integrate the squared scattering amplitude over phase space

High-dimensional phase space
Many final state particles

Complex integrand
Sharp peaks and correlations

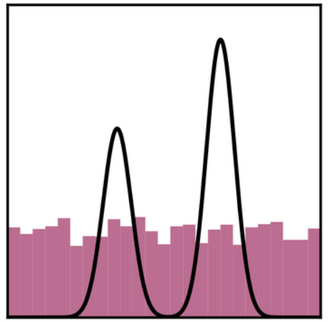


MC Sampler

Monte Carlo integration
Efficient sampling is essential for precise predictions

Monte Carlo 101

$$\sigma = \int f(x)\, dx$$



Importance Sampling

$$\int \frac{f(x)}{g(x)} g(x)\, dx$$

Multiply by 1:

adapt what we sample

Control Variates

$$\int \big(f(x) - c(x)\big) dx + \int c(x)\, dx$$

Add a zero:

reduce what we integrate

Monte Carlo 101

Neural Importance Sampling

$$\int \frac{f(x)}{g_\phi(x)} g_\phi(x)\, dx$$

- Changes the integral **measure**
- **Optimal sampling**
- Must be able to **sample** from g_ϕ

Neural Control Variates

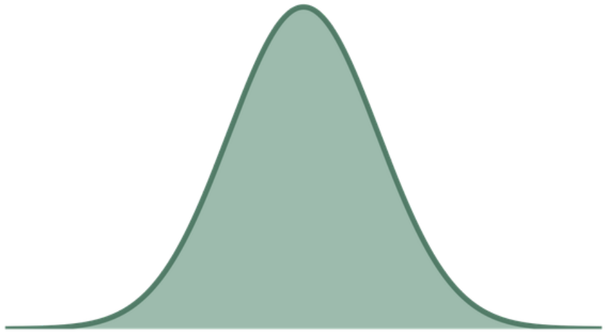
$$\int \big(f(x) - c_\theta(x)\big) dx \; + \int c_\theta(x)\, dx$$

- Changes the **integrand**
- **Uniform sampling**
- If the **integral** of c_θ is known analytically, only the residuals need numerical integration

Normalizing Flows

An invertible transformation maps a simple distribution to a flexible learned density

Latent

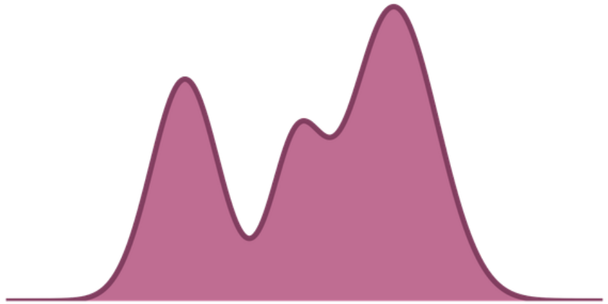


$p_Z(z)$

$$x = f_\theta(z)$$

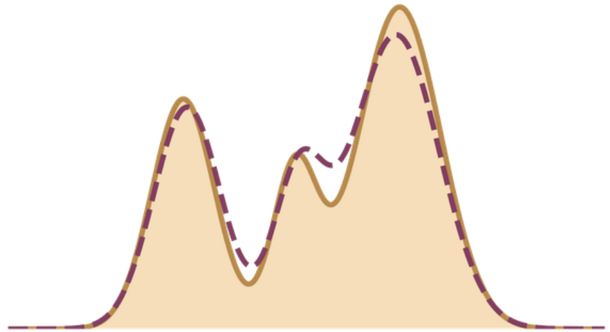
$$z = f_\theta^{-1}(x)$$

Learned



$p_\theta(x)$

Target



$p_{\text{target}}(x)$

Normalized with tractable likelihood $\int p_\theta(x)\, dx = 1$

Best of Both Worlds

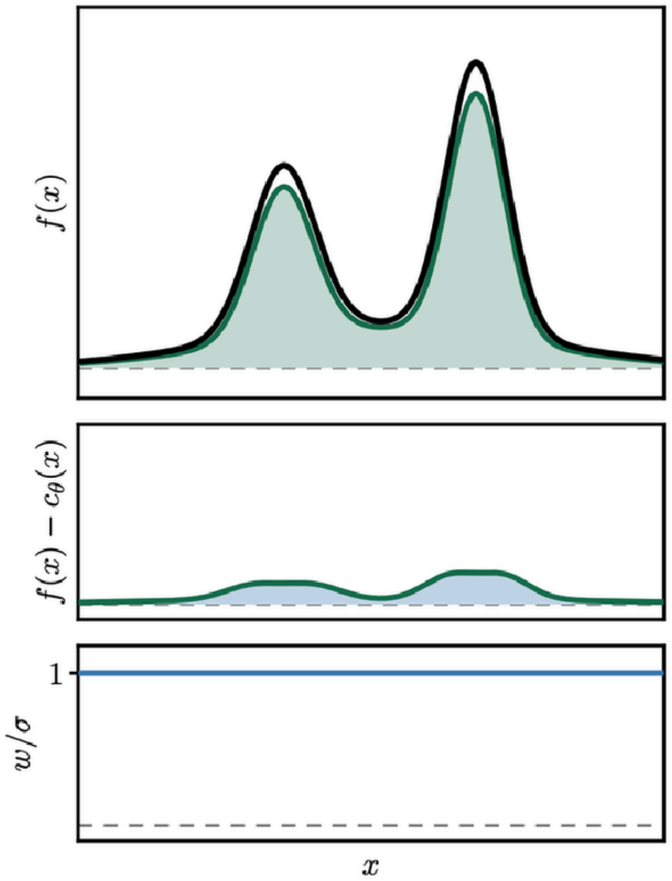
$$\sigma = \int (f(x) - c_\theta(x))\, dx + \int c_\theta(x)\, dx$$

parametrize the control variate with a normalizing flow times a constant

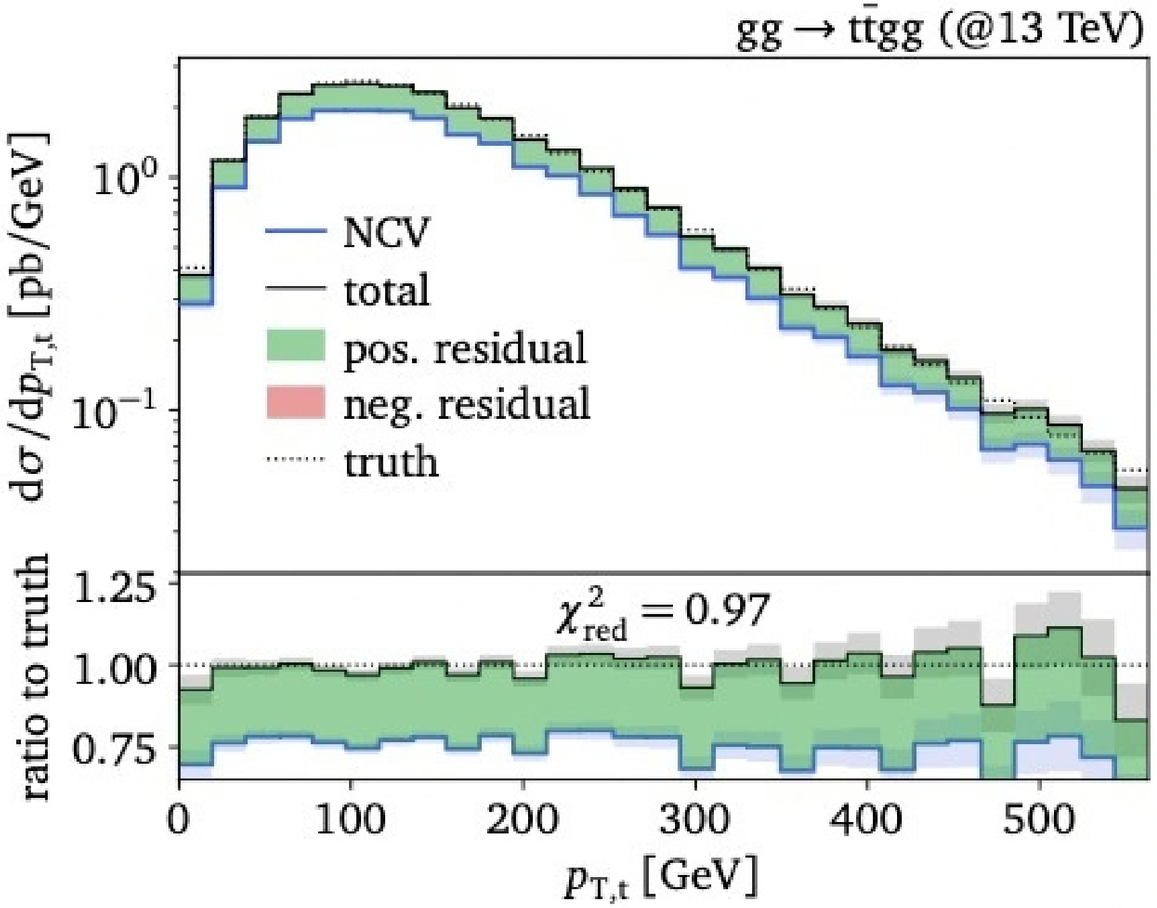
$$\sigma = \int (f(x) - C_\theta p_\theta(x))\, dx + \langle C_\theta \rangle$$

MadNIS samples the residual

$$\sigma = \left\langle \frac{f(x) - C_\theta p_\theta(x)}{g_\phi(x)} \right\rangle_{x \sim g_\phi} + \langle C_\theta \rangle$$



LO Results



$$gg \to t\bar{t}gg$$

$\sim \times 20$ speed up using MadNIS

$\sim \times 3$ Additional speedup with NCVs over MadNIS

NLO Calculations

At NLO the cross section includes virtual and real corrections:

$$\sigma_{\mathrm{NLO}} = \int d\Phi_n \, [B(\Phi_n) + V(\Phi_n)] + \int d\Phi_{n+1} \, R(\Phi_{n+1}) \,,$$

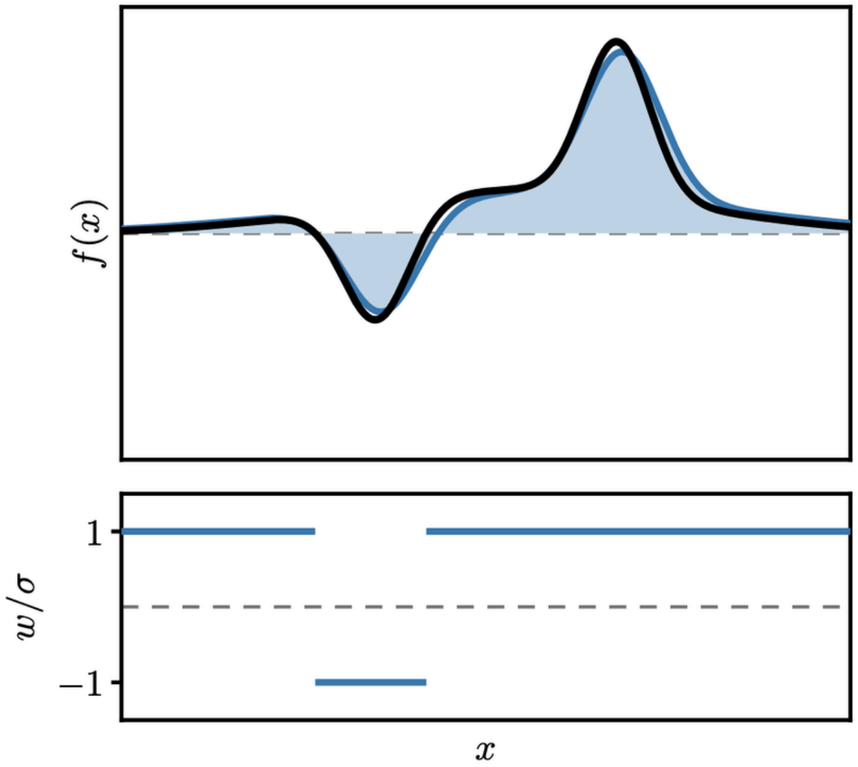
contain soft and collinear divergences

Subtraction scheme: Catani-Seymour Dipoles

$$\sigma_{\mathrm{NLO}} = \int d\Phi_n \, \left[B(\Phi_n) + V(\Phi_n) + I(\Phi_n)\right] + \int d\Phi_{n+1} \, \left[R(\Phi_{n+1}) - S(\Phi_{n+1})\right]$$

sign-changing **sign-changing**

The problem with negative weights



Inherent lower bound on the relative variance:

$$\mathrm{RV} \geq \mathrm{RV}_{\mathrm{min}} = \frac{\mathrm{Var}(w)}{\sigma^2} = \frac{4\rho(1-\rho)}{(2\rho-1)^2}$$

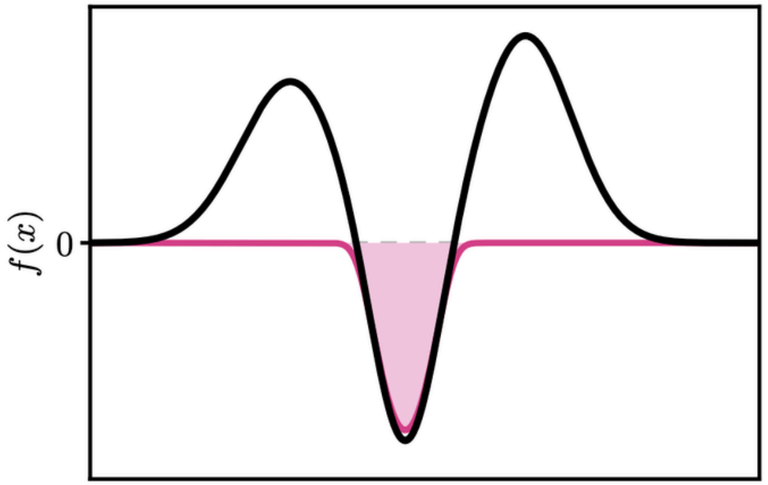
with $$\rho = \frac{\sigma^+}{\sigma^+ + \sigma^-}$$

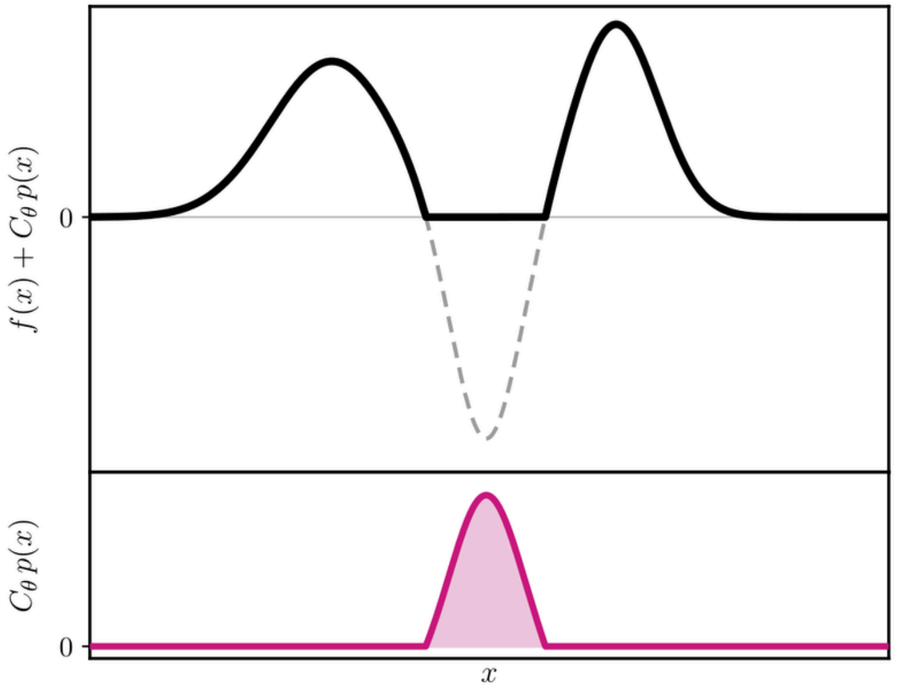
Even optimal importance sampling cannot overcome this bound!

Reducing Negative Weights with NCVs

$$\sigma = \int \big(f(x) + C_\theta p_\theta(x)\big) dx \; - \langle C_\theta \rangle$$

The flow learns the **shape** of the **negative region**



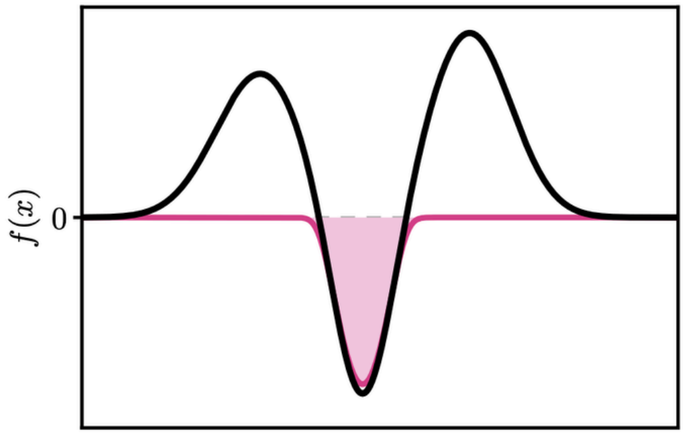


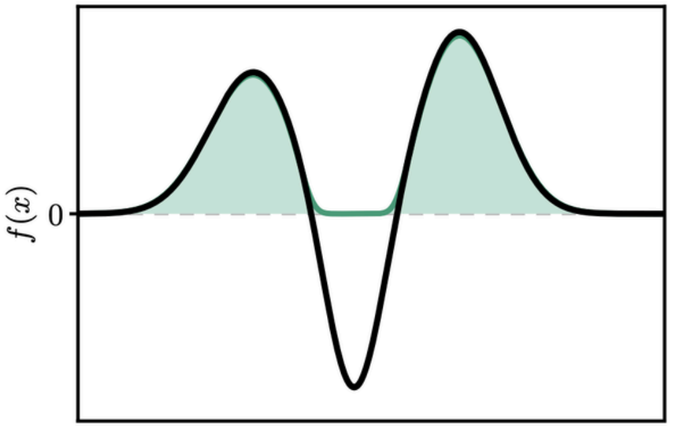
Signed NCVs

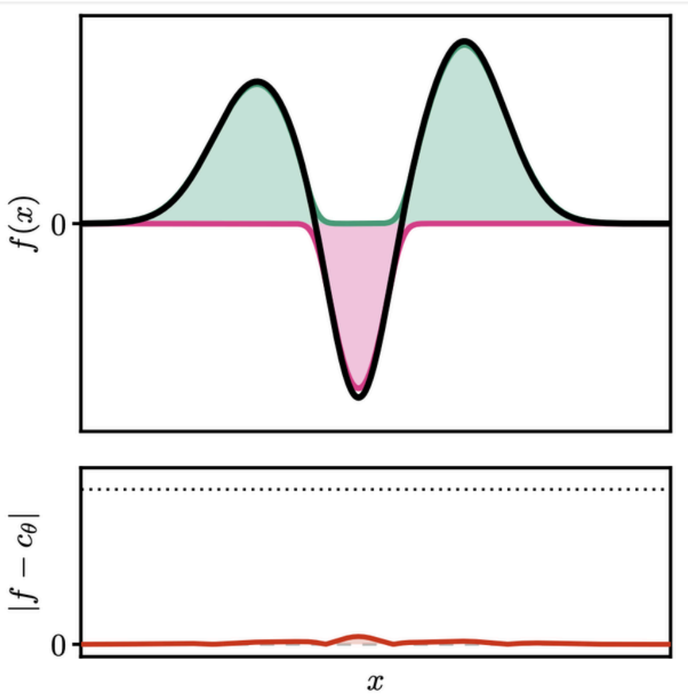
$$\sigma = \int \big(f(x) + C_\theta p_\theta(x)\big) dx \; - \langle C_\theta \rangle$$

Two normalizing flows: 1 positive NCV + 1 negative NCV

$$\sigma = \langle f(x) + C_\theta^- p_\theta(x) - C_\theta^+ p_\theta(x) \rangle - \langle C_\theta^- \rangle + \langle C_\theta^+ \rangle$$







Signed NCVs

$$\sigma = \int \big(f(x) + C_\theta p_\theta(x)\big) dx \; - \langle C_\theta \rangle$$

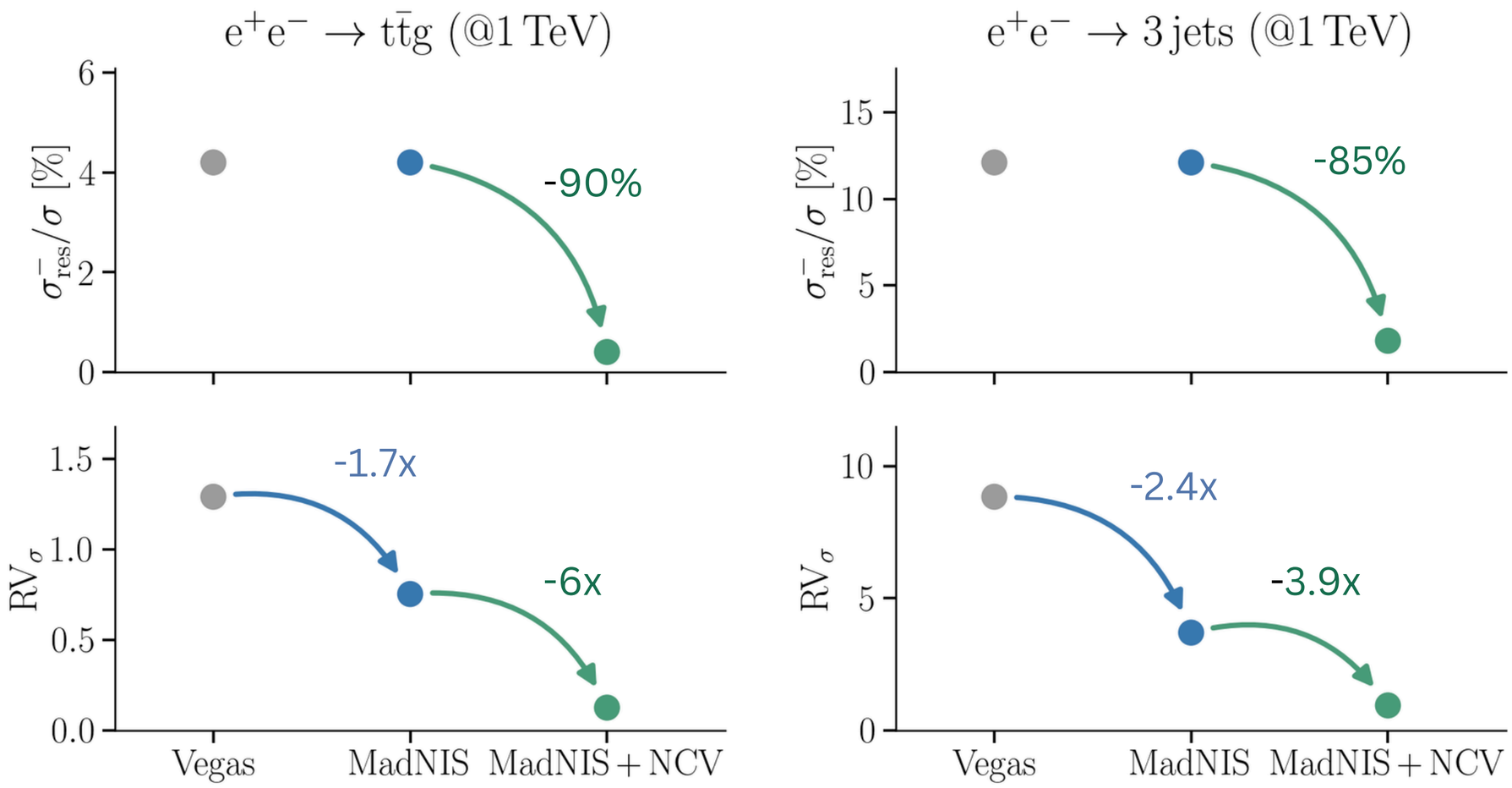
Two normalizing flows: 1 positive NCV + 1 negative NCV

$$\sigma = \langle f(x) + C_\theta^- p_\theta(x) - C_\theta^+ p_\theta(x) \rangle - \langle C_\theta^- \rangle + \langle C_\theta^+ \rangle$$

MadNIS on residual integral

$$\sigma = \left\langle \frac{f(x) + C_\theta^- p(x) - C_\theta^+ p(x)}{g_\phi(x)} \right\rangle_{x \sim g_\phi} - \langle C_\theta^- \rangle + \langle C_\theta^+ \rangle$$

Results


$e^+e^- \to t\bar{t}g$ (@1 TeV)
$e^+e^- \to 3\,\text{jets}$ (@1 TeV)
$\sigma^-_{\text{res}}/\sigma$ [%]
RV$_\sigma$
-90%
-85%
-1.7x
-6x
-2.4x
-3.9x
Vegas
MadNIS
MadNIS + NCV

Conclusions

NCVs

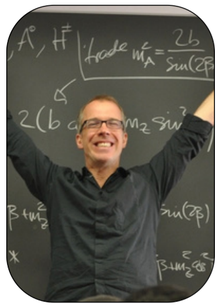
- Improve variance reduction for (N)LO
- Predictions effectively reduce negative weights
- Can reshape the integrand


Rebecca


Theo


Ramon

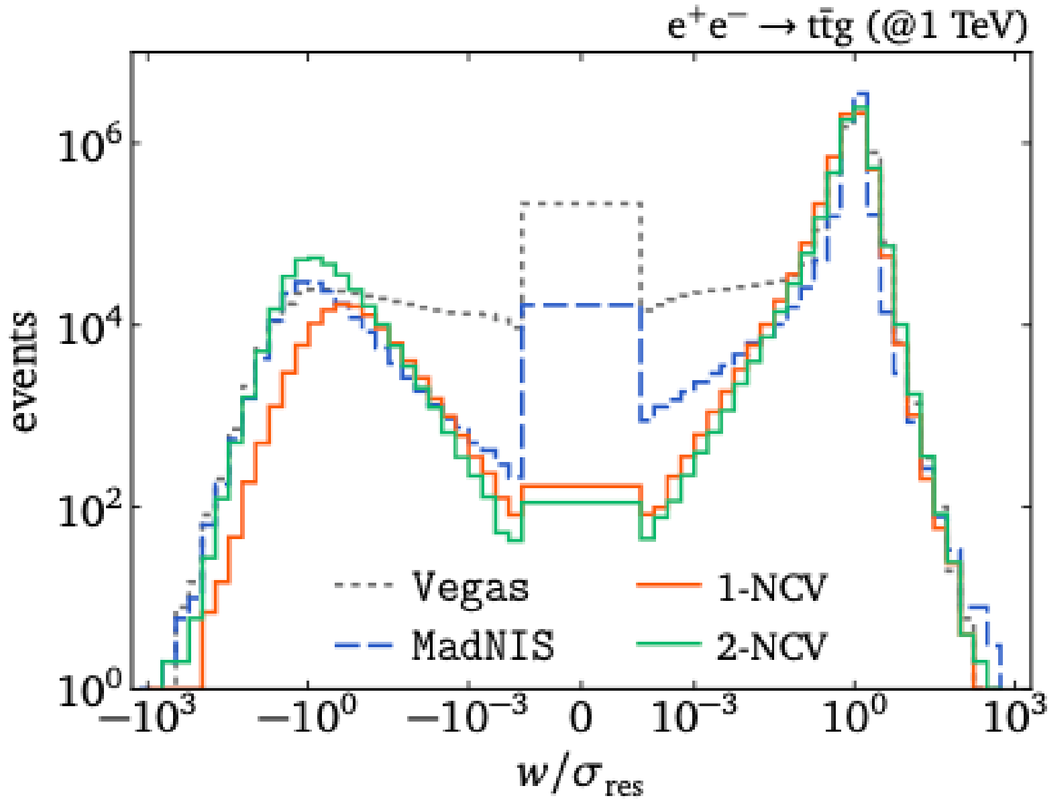

Sophia

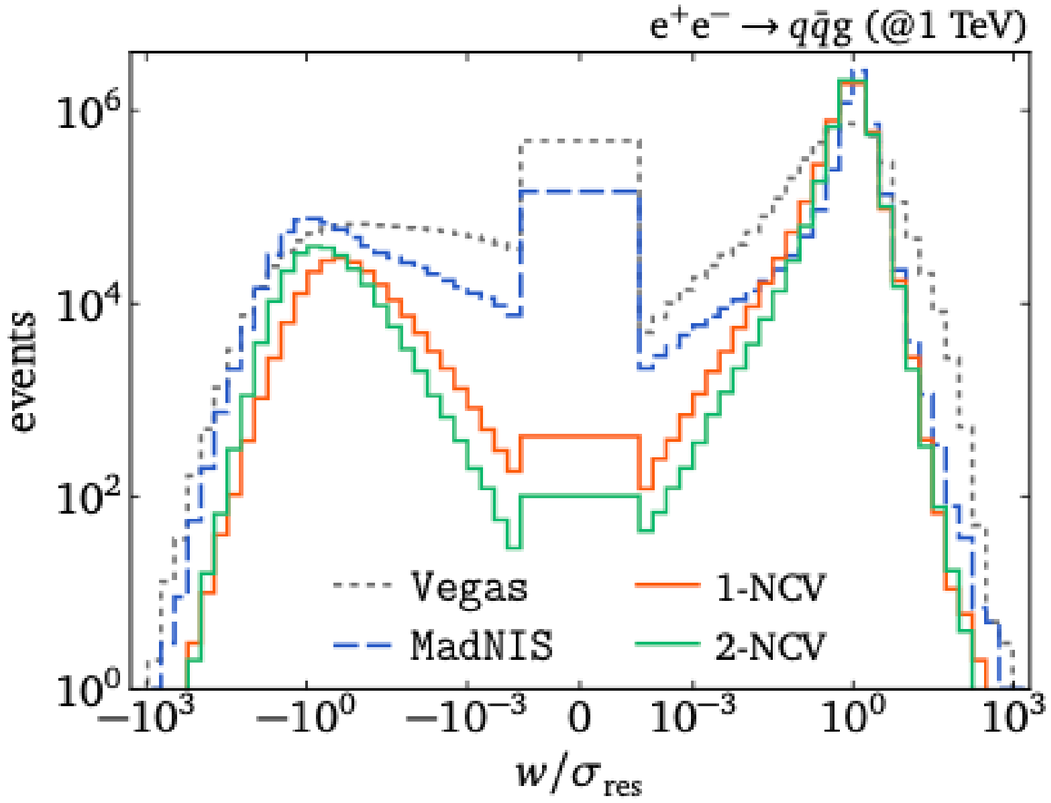

Tilman

read the paper!

[2607.23591]

Results


$e^+e^- \to t\bar{t}g$ (@1 TeV)
events
10^6
10^4
10^2
10^0
Vegas
MadNIS
1-NCV
2-NCV
-10^3
-10^0
-10^{-3}
0
10^{-3}
10^0
10^3
w/σ_{res}


$e^+e^- \to q\bar{q}g$ (@1 TeV)
events
10^6
10^4
10^2
10^0
Vegas
MadNIS
1-NCV
2-NCV
-10^3
-10^0
-10^{-3}
0
10^{-3}
10^0
10^3
w/σ_{res}

Loss functions

learns the **normalization constant**

keeps the residual **positive**

$$\mathcal{L} = \sum_{s=\pm} \left[\mathcal{L}_{\text{shape},s} + \lambda_{\text{int}}\, \mathcal{L}_{\text{int},s} \right] + \mathcal{L}_{\text{NIS},\phi} + \lambda_{\text{neg}}\, \mathcal{L}_{\text{neg},\theta}$$

learns the **shape** of the integrand

NIS learns the **shape** of the residual