PROCEEDINGS OF SCIENCE

# Applying Neural Importance Sampling to gluon scattering

**Enrico Bothmann,**[a,*] **Timo Janßen,**[a] **Max Knobbe,**[a] **Tobias Schmale**[a] **and Steffen Schumann**[a]

[a] *Institut für Theoretische Physik, Georg-August-Universität Göttingen*
*Friedrich-Hund-Platz 1, D-37077 Göttingen, Germany*

*E-mail:* enrico.bothmann@uni-goettingen.de

In this talk we report on a novel approach for the integration of scattering cross sections and the generation of partonic event samples in high-energy physics, originally presented in [1]. It is based on an importance sampling algorithm which includes the use of neural networks in order to overcome typical shortcomings of conventional approaches. At the same time, a potential pitfall of neural networks in the context of phase-space sampling, namely mappings that are non-bijective after trainings with finite data sets, is avoided by employing the technique of Neural Importance Sampling. With this, full phase-space coverage and the correct reproduction of the target distribution is guaranteed even for limited training statistics. We study the performance gains of our implementation for a prototypical high-energy physics example, namely gluon scattering into three- and four-gluon final states.

*The Eighth Annual Conference on Large Hadron Collider Physics-LHCP2020*
*25-30 May, 2020*
*online*

[*] Speaker

PoS(LHCP2020)056

## 1. Introduction

Matrix element generators in HEP are often limited by the performance of their phase space (PS) sampler. Insufficient mappings of the target distribution result in large fluctuations of the event weights and correspondingly a large number of target-function evaluations when generating unweighted events. Typically the sampling performance deteriorates significantly with the PS dimensionality/particle multiplicity [2], and the complexity of the integrand arising from intermediate resonances, regularised singularities, quantum-interference effects, or non-trivial PS cuts that the integrator can not adapt to. Recently, there has been growing interest to employ modern machine-learning techniques to this problem of PS sampling in particle physics, cf. [3–7]. In this talk, we discuss potential pitfalls when extending or replacing sampling methods using neural networks (NN), and we present and benchmark a novel sampler based on Neural Importance Sampling (NIS) [8–10], which was initially proposed in the context of rendering 3D scenes. This talk is based on our work in [1], see therein for additional details. An independent study of applying NIS to high-dimensional (HEP) integration problems has been simultaneously presented in [11, 12].

## 2. Neural Importance Sampling with Coupling Layers for High-Energy Physics

A sampling method used for event generation should fulfil three requirements. First, the MC sample should converge to the true target distribution in the entire PS in the limit of infinite sampling statistics. This should be guaranteed, even if training samples used to optimise the sampler have finite statistics and thus can not provide full coverage. Secondly, the method should be general and self-adaptive, minimising the need for manual intervention. Finally, the produced samples should consist of uncorrelated events. Standard sampling methods such as VEGAS [13] fulfil these conditions by construction. For NN-assisted sampling methods however, it can be non-trivial to fulfil them all, in particular the first one. In Ref. [14] it has been cautioned against the use of Generative Adversarial Networks (GANs) to extrapolate from finite training statistics to large event samples for physics analysis. While GANs (and NNs in general) can be similarly effective to add statistical power to a finite training data set compared to a low-parameter fit [15], and a recent study indicates that a careful NN training procedure can give a map that is bijective to a good approximation [16], caution must be applied in particular in the tails of the generated distributions. In [1], we analyse some previous network-based sampling proposals [3, 4] along these lines, finding that full PS coverage and hence a reproduction of the true distribution is indeed not always guaranteed.

We therefore choose to use NIS, which meets all of the above requirements. Here, the PS mapping is optimised through bijective maps called coupling layers, which get adjusted by NNs, as initially proposed in [8]. We employ a piecewise quadratic coupling layer structure, first suggested in [9]. Note that the usage of coupling layers for importance sampling has also been studied in [10]. The way coupling layers compose mappings gives a computationally very cheap Jacobian determinant, only growing linearly with the number of dimensions, while at the same time allowing arbitrarily complex mappings that are highly non-linear, e.g. given by deep NNs. Since the invertibility property of NIS allows its use within multi-channel sampling, we can use it as a drop-in replacement of the standard VEGAS optimisation often used within multi-channel Monte Carlo integration in HEP. The input space is a $d$-dimensional hypercube of uniformly distributed

| | 3 jets | | | 4 jets | | |
|---|---|---|---|---|---|---|
| Sample | $\epsilon_{\mathrm{uw}}$ | $E_N$ [pb] | $P_{\mathrm{acc}}$ | $\epsilon_{\mathrm{uw}}$ | $E_N$ [pb] | $P_{\mathrm{acc}}$ |
| Uniform | 3.0 % | 24806(55) | 89 % | 2.7 % | 9869(20) | 57 % |
| Vegas | 27.7 % | 24813(23) | 32 % | 31.8 % | 9868(10) | 17 % |
| NN | 64.3 % | 24847(21) | 34 % | 33.6 % | 9859(10) | 16 % |

**Table 1:** Results for sampling the total cross section for gluonic jet production at $\sqrt{s} = 1\,\mathrm{TeV}$, i.e. the MC integral $E_N$ is an estimator for $\sigma_{\mathrm{gg}\to n\,\mathrm{jets}}$. Besides $E_N$ and its MC error, we also show the unweighting efficiency $\epsilon_{\mathrm{uw}}$ and the acceptance rate $P_{\mathrm{acc}}$, comparing unoptimised sampling (“Uniform”) to Vegas- and NN-based sampling. All samples consist of $N = 10^6$ non-zero weight events.

random numbers, which is then warped by the coupling layers to better match the integrand. Note that we do not sample points directly in momentum space, which would require the NNs to exactly adhere to four-momentum conservation and on-shell conditions. Instead, the individual channels in the multi-channel algorithm employ subsequent (channel-dependent) transforms to translate the warped random numbers into momenta. This guarantees that physical constraints are automatically fulfilled. The NN training minimises the Pearson $\chi^2$-divergence between the target function and the sample distribution via gradient descent, using minibatches of sampling points for each iteration. For more details on the training procedure, please refer to [1].

## 3. Results

We compare our new implementation based on NIS with a reference implementation using Vegas. While [1] also discusses top-quark decays and lepton-induced top-quark pair production and decay, we focus here on gluon-induced multi-jet production, i.e. gg → $n$ gluons with $n = 3, 4$. The channels of the multi-channel are built according to the QCD antenna radiation pattern realised by the Haag algorithm [17, 18]. For $n = 3$, Haag constructs 24 channels. However, most are related by a permutation of the momenta, and hence we employ mappings to reduce the number of (truly independent) channels to 2. For $n = 4$, Haag constructs 120 channels, which we reduce to 3 independent ones. Note that one neural importance sampler (with embedded NNs) or alternatively one Vegas sampler is used per channel. However, during the training, we optimise all neural networks simultaneously, see [1] for further details. Each neural importance sampler consists of 5 coupling layers with 32 bins used for the piecewise quadratic mapping.

Jet cross sections are defined using the anti-$k_t$ algorithm [19] with $R = 0.4$. The renormalisation scale is set to $\mu_R = \sqrt{s}$. For the cross section regularisation, Haag uses a cut-off parameter which we set to $s_0 = 900\,\mathrm{GeV}^2$. On the final state we instigate a cut on the invariant masses of all gluon pairs of $m_{ij} > 30\,\mathrm{GeV}$, and the transverse momenta of all particles, $p_{\perp,i} > 30\,\mathrm{GeV}$. Accordingly, the optimisers have to deal with “dead” regions in PS and therefore non-continuous integrands.

In Tab. 1, we show the results of sampling the cross section without optimisation (“Uniform”), with Vegas optimisation and with our NIS-based “NN” optimisation. The unweighting efficiencies $\epsilon_{\mathrm{uw}} = w_{\mathrm{max}}/\langle w \rangle$ for $n = 3, 4$ are about 3 % for the unoptimised sampling, and increase to about 30 % by Vegas optimisation. The NN optimisation surpasses Vegas for $n = 3$ by a factor of two, whereas for $n = 4$ we find no significant improvement over Vegas. Both Vegas and NN

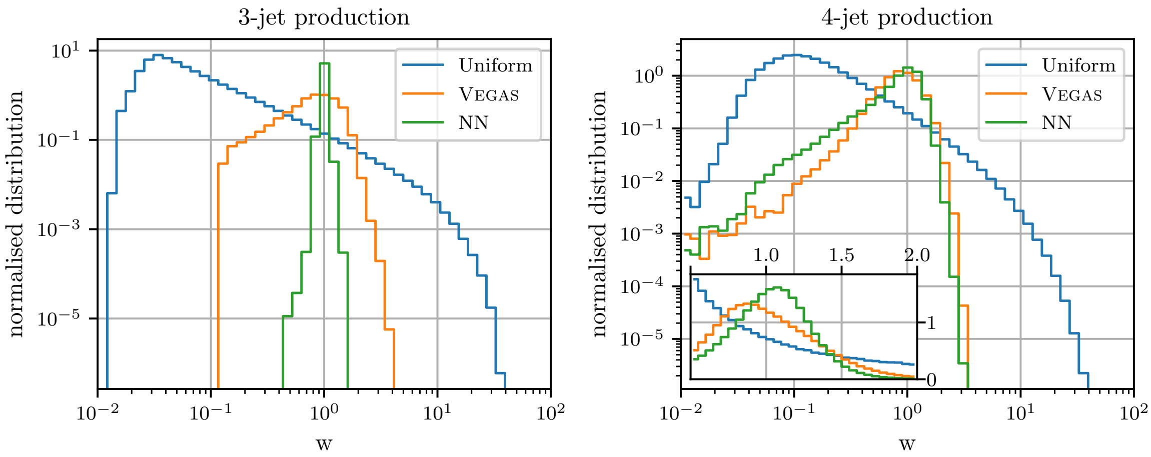

**Figure 1:** Event weight distributions for the three samples as described in Tab. 1.

optimisation give similar improvements for the estimate of the standard deviation for $n = 3$ and $n = 4$. We also quote in Tab. 1 the acceptance rate $P_{\text{acc}} = N/N_{\text{trials}}$, i.e. the probability that a proposed point passes the PS cuts and hence contributes to the integral. In our production setup, the cuts regularise the matrix elements, the integrand will therefore be large close to these boundaries. Consequently, both optimisations lead to a decrease in $P_{\text{acc}}$, as they enhance the sampling rate close to the cuts, with the side effect of proposing points also outside of the fiducial volume.

The resulting event weight distributions for the three samples are presented in Fig. 1. For 3-jet production, we find that NN optimisation gives the most strongly peaked distribution. The situation is more ambiguous for 4-jet production. Both Vegas and NN optimisation significantly sharpen the distribution, in fact providing quite similar outcomes. However, while NN optimisation results in a slightly more pronounced peak compared to Vegas and a slightly faster fall-off towards large weights, it depletes less quickly towards small weights. In particular for the 3-jet case it might be surprising that we find a comparable estimate for the standard deviation for NN and Vegas optimisation, although the distribution is narrower in the NN case. This apparent discrepancy originates from the higher fraction of zero-weight events for the optimised samples, i.e. events that fall outside the physical PS volume and are thus not accepted. The standard deviation of the integral estimate is in such a case largely determined by the corresponding acceptance rate, since the weight distribution will then actually contain two peaks: the one at a finite value and one at $w = 0$. A further improvement in the sampling accuracy would therefore require a modification of the optimisation to reduce the number of discarded points. The unweighting efficiency is not affected by $P_{\text{acc}} < 1$, since it takes into account non-zero weights only.

In this talk we have presented selected results from our recent study [1], applying NIS to optimise PS sampling in MC integration problems in high-energy physics. The method easily surpasses Vegas in all performance measures for 3-jet production, while we find them to be on par for the more complex 4-jet production. These findings are consistent with those in a parallel study [11] for V + jets production. The results for the more basic examples are very promising and indicate that conventional optimisers such as Vegas could potentially be outperformed by NN-based approaches also for more complex problems in the future.

## References

[1] E. Bothmann, T. Janßen, M. Knobbe, T. Schmale and S. Schumann, *Exploring phase space with Neural Importance Sampling*, *SciPost Phys.* **8** (2020) 069 [2001.05478].

[2] S. Höche, S. Prestel and H. Schulz, *Simulation of Vector Boson Plus Many Jet Final States at the High Luminosity LHC*, *Phys. Rev.* **D100** (2019) 014024 [1905.05120].

[3] J. Bendavid, *Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks*, 1707.00028.

[4] M.D. Klimek and M. Perelstein, *Neural Network-Based Approach to Phase Space Integration*, 1810.11509.

[5] S. Otten, S. Caron, W. de Swart, M. van Beekveld, L. Hendriks, C. van Leeuwen et al., *Event Generation and Statistical Sampling with Deep Generative Models*, 1901.00875.

[6] R. Di Sipio, M. Faucci Giannelli, S. Ketabchi Haghighat and S. Palazzo, *DijetGAN: A Generative-Adversarial Network Approach for the Simulation of QCD Dijet Events at the LHC*, *JHEP* **08** (2020) 110 [1903.02433].

[7] A. Butter, T. Plehn and R. Winterhalder, *How to GAN LHC Events*, *SciPost Phys.* **7** (2019) 075 [1907.03764].

[8] L. Dinh, D. Krueger and Y. Bengio, *NICE: non-linear independent components estimation*, in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Workshop Track Proceedings*, Y. Bengio and Y. LeCun, eds., 2015 [1410.8516].

[9] T. Müller, B. McWilliams, F. Rousselle, M. Gross and J. Novák, *Neural importance sampling*, *ACM Transactions on Graphics (Proc. SIGGRAPH)* **38** (2019) [1808.03856].

[10] Q. Zheng and M. Zwicker, *Learning to importance sample in primary sample space*, *Computer Graphics Forum* **38** (2019) 169–179 [1808.07840].

[11] C. Gao, S. Höche, J. Isaacson, C. Krause and H. Schulz, *Event Generation with Normalizing Flows*, *Phys. Rev. D* **101** (2020) 076002 [2001.10028].

[12] C. Gao, J. Isaacson and C.G. Krause, *i-flow : High-dimensional integration and sampling with normalizing flows*, *Machine Learning: Science and Technology* **1** (2020) [2001.05486].

[13] G.P. Lepage, *A new algorithm for adaptive multidimensional integration*, *Journal of Computational Physics* **27** (1978) 192 .

[14] K.T. Matchev and P. Shyamsundar, *Uncertainties associated with GAN-generated datasets in high energy physics*, 2002.06307.

[15] A. Butter, S. Diefenbacher, G. Kasieczka, B. Nachman and T. Plehn, *GANplifying Event Samples*, 2008.06545.

[16] I.-K. Chen, M.D. Klimek and M. Perelstein, *Improved Neural Network Monte Carlo Simulation*, 2009.07819.

[17] A. van Hameren and R. Kleiss, *Generating QCD antennas*, *Eur. Phys. J.* **C17** (2000) 611 [hep-ph/0008068].

[18] A. van Hameren and C.G. Papadopoulos, *A Hierarchical phase space generator for QCD antenna structures*, *Eur. Phys. J.* **C25** (2002) 563 [hep-ph/0204055].

[19] M. Cacciari, G.P. Salam and G. Soyez, *The anti-$k_t$ jet clustering algorithm*, *JHEP* **04** (2008) 063 [0802.1189].