# Optimising loop amplitude evaluations

## Simon Badger (University of Turin)

### Event generators' and N(n)LO codes' acceleration **CERN** 14th November 2023





## well known problem:

multi-scale amplitudes difficult Commit ateCommit Commit

CPU cost scales badly with higher orders

(especially when IR singular)



## some solutions already exsist

automated and optimised MCs

multi-core processing re-weighting lower orders …as well as other methods discussed at this meeting **this talk:** testing amplitude neural network for loop induced processes

 $gg \rightarrow Yyg$  and  $gg \rightarrow YVgg$ 

- Realistic hadron collider setup with SHERPA  $\mathbf{1}$ 1*/* d*/*d*pT,j*2 [GeV1]
- Precision evaluations: determining NN errors 10<sup>5</sup>



Aylett-Bullock, SB, Moodie JHEP 08 (2021) 066 1*/* d*/*d12

SB, Butter, Luchmann, Pitz, Plehn SciPost Phys.Core 6 (2023) 034 1*.*0 Ratio

 $\mathbf{A} \cup \mathbf{B}$ ration, *R*-separation is defined in Section 3.2.1, and *m*1*,*<sup>2</sup> and ⌘1*,*<sup>2</sup> are the mass and • Reflections on experience at NNLO

plitudes. The intervention of and the isophermed How expensive are the loop amplitudes... (*a*1*,...,an*2) = tr (*t a*1 *t <sup>a</sup>*<sup>2</sup> *...tan*<sup>2</sup> )+(1)*n*tr (*<sup>t</sup> a*1 *t <sup>a</sup>n*<sup>2</sup> *...ta*<sup>2</sup> )*.* (2.2) This yields (*n* 3)!*/*2 primitive amplitudes *A* for *n* 5. For example, for *n* = 4 there is a



## Precision is Important

standard one-loop matrix elements provide high precision evaluations:

OpenLoops, GoSam, Madloop, BlackHat, Recola, Helac-NLO…



use the ability to switch numerical precision (e.g. qd for 32 or 64 digits)

**Figure 5**. Accuracy for 5-jet amplitudes: (a) shows the seven gluon process and (b) the  $d\overline{d} \rightarrow \overline{d}d + 3g$  process. The thicker histograms show computations in double precision whereas the thinner curves show the distribution in quadruple precision for points which did not pass the relative accuracy of  $10^{-4}$  when calculated in double precision. Red histograms show the  $\frac{1}{\epsilon^2}$  poles, green histograms the  $\frac{1}{\epsilon}$  and blue histograms the finite part of the amplitudes.

dimension scaling tests, gauge invariance checks,…

# Wishlist

an amplitude approximation which is:

- simpler to train/fit than generating events using traditional methods
- simplest model which fits a generic process
- reliable error estimates
- robust against changes in phase-space (cuts, jet algorithms, scale variations, etc.)
- simple to distribute

#### first attempt: e+e- > jets regions resulting from each partition contain only a specific subset of singularities. In order

#### SB, (Aylett-)Bullock [2002.07516] to achieve this, a set of ordered pairs, known as FKS pairs, are introduced. In our case of

- Single NN does badly
- Splitting IR sectors via FKS sectors improves reliablity
- Error estimates by varying model initialisation (ensemble of networks) *PHILING IN SECTED 5 VIGHTS* 
	- K-factors work better than tree-level (no 1/s poles)
	- Various tests suggest single run speed improvements at least x10

$$
\mathcal{S}_{i,j} = \frac{1}{D_1 s_{ij}}, \quad D_1 = \sum_{i,j \in \mathcal{P}_{\text{FKS}}} \frac{1}{s_{ij}},
$$

$$
\mathrm{d}\sigma^{(X)} = \sum_{i,j} S_{i,j} \,\mathrm{d}\sigma^{(X)},
$$



divergent regions according to the behaviour of *Si,j* for each pair. The first pair corresponds Figure 2: Behaviour of the *Sq,g* FKS partition function

### first attempt:  $e+e- \rightarrow jets$



trained models or ensembles.

with many scales and complex infrared singularity structures. Moreover, we find that much  $\mathbb{R}$  can be learnt from the performance of the models here can be applied to the models here can be applied to the NLO Normalised  $\mathrm{d}\sigma/$   $\mathrm{d}y$  $\sim 10^{-2}$ an ensemble of networks each trained on *N*max+1 partitions of phase-space. In determining  $\mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\right]\right] & \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\right]\right] & \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\right]\right] & \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\right]\right] & \mathbb{E}\left[\begin{array}{cc} \mathbb{E}\left[\mathbb{E}\left[\mathbb{E}\right]\$  $\mathcal{L}$  matrice of  $\mathcal{L}$  of the ratio of the network of the network of the  $\mathcal{L}$  and  $\mathcal{L}$  and as well as the network's ability to approximate the cross-section. The cross-section  $\mathbf{H}$  $F_{\text{max}}$  shows the distribution of the neural network errors by calculating the ratio of t  $\begin{bmatrix} 1 & 1 & 1 \ 1 & 1 & 20 \end{bmatrix}$ the ensemble of the ensemble of networks and more gives much narrower and more Gaussian shaped distributions of the ensemble o  $t_{\rm max} = 20$  $\begin{array}{cccccccc} 0.1 & 0.2 & 0.3 & 0.4 & 0.5 & 0.6 & 0.7 & 0.8 \end{array}$   $\begin{array}{cccc} 0.2 & 0.1 & 0.2 & 0.3 & 0.4 \end{array}$ 



tree-level factorisation aware approach performs better



Maitre, Truong [2107.06625]

second attempt: gg > YY+gluons

Aylett-Bullock, SB, Moodie [2106.09474]



 $7-9\%$  of the total phase space

### third attempt:  $gg \rightarrow YY+gluons$  with **Bayesian networks** evaluate the likelihood over a mini-batch rather than the full training dataset, we rescale the

SB, Butter, Luchmann, Pitz, Plehn [2206.14831] *p*(*!|T*) The same heteroscedastic loss [19] can be used in deterministic networks, if we introduce model as a second trained quantity in addition to the amplitude values. The Bayesian net-

- weights and biases are replaced with (Gaussian) distributions and predictions
- optimised training times vs. pure ensemble approach  $t_{\rm{inter}}$  about a Gaussian  $\mu$  and  $\mu$  and  $\mu$  more general BNN. even though it might well be possible to use a deterministic network for similar appearance of similar applications.
- better defined error estimates *p*(*!*)*p*(*T|!*) e
	- improved training via loss and performance boosting networks the prior *performance* the corresponding networks the prior *performance* to prior *performance* the prior *performance* the prior *performance* the prior *performance* the prior *performance* = KL[*q*(*!*), *p*(*!*)] *d! q*(*!*) log *p*(*T|!*) + log *p*(*T*) model parameters before training. The model evidence *p*(*T*) guarantees the correct normal-



$$
A_j \to \log\left(1 + \frac{A_j}{\sigma_A}\right)
$$

- $\sim$  6k params for 2  $\rightarrow$  3
	- ~600k params for  $2 \rightarrow 4$



#### third attempt:  $gg \rightarrow \gamma \gamma +$ gluons with Bayesian networks  $S$  fining attempt of  $\rightarrow VV \pm$  of the training and  $T$ best available performance from the literature compact of the literature compact of the literature of the literatu

SB, Butter, Luchmann, Pitz, Plehn [2206.14831]

- weights and biases are replaced with (Gaussian) distributions
- optimised training times vs. pure ensemble approach
- better defined error estimates
- improved training via loss and performance boosting



### third attempt:  $gg \rightarrow \gamma \gamma +$ gluons with Bayesian networks

SB, Butter, Luchmann, Pitz, Plehn [2206.14831]



performance boosting reduces uncertainties in tails

#### Higher Order Applications external legs by evaluating the IBP solution at permuted points. For all families and identities are generated in the second integrals with  $\sim$ to  $\mathbb{R}^3$  with respect to the amplitude to those appearing in the amplitude. We obtained the amplitude them by using the amplitude. We obtained the amplitude them by using the amplitude them by using the amplitude the

precision frontier has moved to  $NNLO$  (QCD) and beyond  $2 + 200$ -100p,  $cosu$ od to



paracision requirements more subtle precision requirements more subtle 7 Conclusions In this paper we have presented a complete, full colour, five-point amplitude at two loops instabilities can be in both rational coefficients and special functions

havond  $2 \div 2$  @3-loop,  $2 \rightarrow 3@$  2-loop

 $\overline{f_{\text{max}}[f_{\text{max}}]}$  reasonable performance for 2 $\rightarrow$ 3 with analytic finite remainders (extracted transcendental constants (e.g. inte Field sampling)  $\frac{N}{\sqrt{N}}$  = 0 up to the required order. The required order order order. The Laurent expanding order. The Laurent expanding order. The Laurent expanding order. The Laurent expanding order. The Laurent expansion of the L  $\mathbb{E}_{\text{max}$  analytic finite refridingers (extracted periodic pental pentagon functions and pentagon functions and  $\mathbb{E}_{\text{max}}$ 

$$
A_{i;j}^{(L),k} = \sum_{s=-2L}^{o(L)} \sum_{r} \epsilon^s c_{r,s}(\vec{x}) \operatorname{mon}_r(f, \mathfrak{c})
$$

pentagon functions [Chicherin, Sotnikov, Zoia]

All Two-Loop Feynman Integrals for Five-Point oth rational Che-Mass Scattering, Abreu et al. [2306.15431]

## Single Photon plus Two-Jets

SB, Czakon, Hartanto, Moodie, Peraro, Poncelet, Zoia [2304.06682]

### first full colour  $2 \rightarrow 3$  differential cross section



### IR subtraction with STRIPPER approach

analytic 2L finite

RV and RR using ME from OpenLoops and AvH libraries

#### comparison with ATLAS [1912.09866] remainders which we comp

- *A*  $\alpha$  *C*  $\beta$  *C*  $\$ 1. We require at least two jets defined with the anti- $k_T$  algorithm [106] for jet radius  $R = 0.4$  that  $f$   $p_T(j) > 100$  GeV and maximal rapid A<br>
Aave minimal transverse momentum of  $p_T(j) > 100$  GeV and maximal rapidity  $|\eta(j)| < 2.5$ .
- *f*<sub>*k*</sub> *identified jets* U∑ *l de* separated from  $\mathbf{z}$ photon by  $\Delta R(\gamma, j)$  $\ddot{x}$ 2. The identified jets must be separated from the photon by  $\Delta R(\gamma, j) > 0.8$ .
- 3. One isolated photon **m** ast be present in the final state  $\cdot$  $\frac{1}{2}$ *nf*  $\frac{101}{2}$ 3. One isolated photon must be present in the final state with  $E_{\perp}(\gamma) \ge 150$  GeV,  $|\eta(\gamma)| \le 2.37$ excluding  $1.37 \leq |\eta(\gamma)| \leq 1.56$ .

# Single Photon plus Two-Jets

SB, Czakon, Hartanto, Moodie, Peraro, Poncelet, Zoia [2304.06682]



**Excellent overall agreement witles** excellent overall agreement with data

 $\overline{P}^{\rm 50}$ 





# Single Photon plus Two-Jets

SB, Czakon, Hartanto, Moodie, Peraro, Poncelet, Zoia [2304.06682]



total cost ~550 kCPUh

VV:RV:RR ~ 1: 10: 40

Note: the 1st time is always a challenge - many improvements possible

double virtual call average ~16s per point The evaluation of perturbative second-order corrections to the differential cross sections for photon [NB: 1 double precision call O(1s)]

# Future Applications

speeding up amplitude calls with NN looks like a viable option for leading order codes

I would still be interested in better control for the number of correct digits

speeding up amplitude virtual amplitudes can help, but need to make impact on real radiation

two-loop  $2 \rightarrow 3$  not dominated by virtuals, largely thanks to well studied special function basis

This is (probably) not going to continue:  $pp \rightarrow ttj$ ,  $pp \rightarrow ttH$ ,  $\rightarrow WWj$ 

at some stage analytic formula unfeasible - must take numerical route

special function basis in 'pentagon function' form will not be possible (eg. if there are elliptic structures)

expect dramatic change in evaluation time

# Conclusions I still have some questions…

phase-space : how can we use amplitude values to minimize the required number of training points

squared amplitudes vs (ordered) helicity amplitudes

analytic vs. numeric: can we improve the architecture to better satisfy amplitude properties? (c.f. bootstrap techniques)