A Machine Learning approach for DVCS identification without proton detection

Juan Sebastian Alvarado
IJCLab - Orsay

CLAS collaboration meeting
13/03/2024

Introduction

In principle, the measurement of only an electron and a photon is enough to reconstruct a DVCS event. We aim for DVCS event reconstruction without requiring final proton information.

Advantages (with respect to $ep\gamma$ detection):

- Improves GPD studies at small $-t$.
- Higher statistics, hence more precise BSA measurements or smaller bins.
- Helpful for experiments that do not consider proton detection.

Difficulties:

- The $ep\gamma$ final state includes background contributions from the whole Deep Inelastic Scattering (DIS) spectra.
- Reduced options for cuts:
    - Only one exclusivity variable: Missing mass of $ep \to e\gamma$.

Therefore, we need a method that ensures DVCS identification: Machine Learning

We test the ML approach on experimental data:

1. Validation of the method when we include the proton information.
2. Application to the case without proton information.

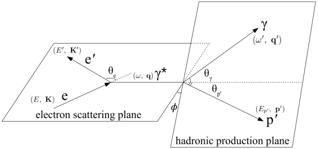
$ep \to e\gamma p$: Data selection

Analyzed data set

- fall2018 RG-A data.
- Inbending and outbending torus configuration

Kinematic window:

- $W > 2$ GeV,
- $Q^2 > 1$ GeV^2 ,
- $\mathbf{q}' > 2$ GeV (photon),
- $\mathbf{k}' > 1$ GeV (electron),
- $\mathbf{p}' > 0.3$ GeV (nucleon).



Exclusivity cuts:

We reconstruct ϕ and t in two ways:

1. Using $\gamma*$ and the outgoing photon γ: $\Rightarrow \phi(\gamma)$
2. Using $\gamma*$ and the recoil proton p: $\Rightarrow \phi(p')$

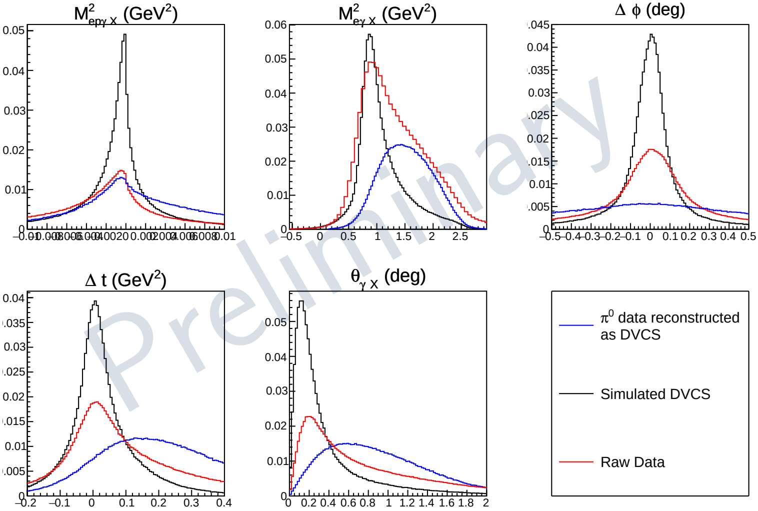
- $\Delta\phi = |\phi(p') - \phi(\gamma)| < 2°$,
- $\Delta t = |t(p') - t(\gamma)| < 2$ GeV^2,
- $\mathbf{P}_{miss} < 1$ GeV.

Event selection:

- No restriction on the number of particles in the event or detection topology.
- If multiple e, γ or p detections, we select the set (e, γ, p) that minimizes the missing mass of the process $ep \to ep\gamma$

$ep \to e\gamma p$: Model training - Inbending torus

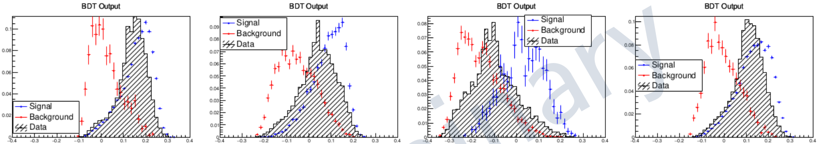
The main contamination channel is $ep \to ep\pi^0 \to ep\gamma(\gamma)$.



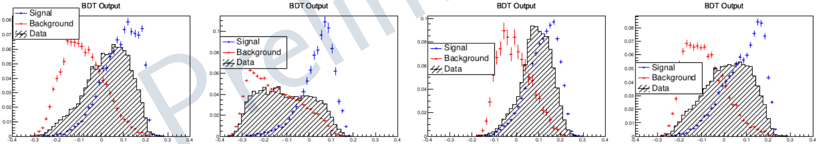
$ep \to e\gamma p$: BDT - Inbending torus

A Boosted Decision Tree (BDT) was trained to classify the events.

- Discriminating variables: $\{M^2_{ep\gamma}, M^2_{e\gamma}, \Delta\phi, \Delta t, \theta_{\gamma X}\}$.
- Simulated DVCS as signal.
- Simulated π^0 events, reconstructed as DVCS, as background.
- **Training is done on each (Q^2, x_B, t) bin.**



(i) Bin 25 (j) Bin 26 (k) Bin 27 (l) Bin 28

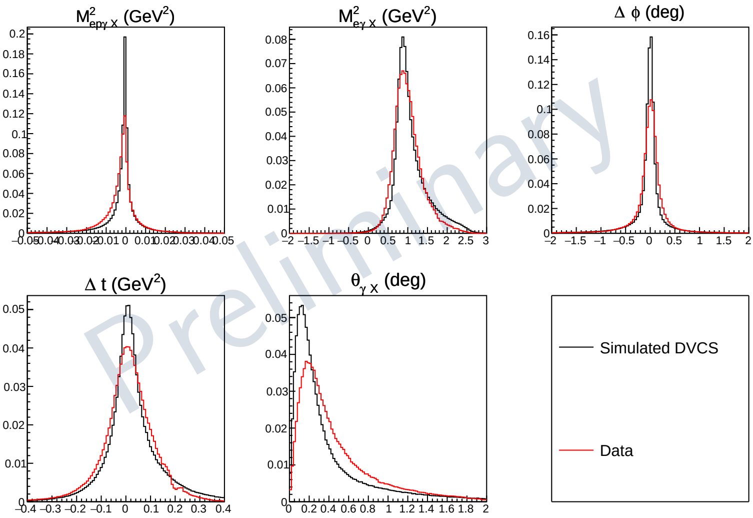


(m) Bin 29 (n) Bin 30 (o) Bin 31 (p) Bin 32

Figure: Some BDT responses for the inbending dataset.

$ep \to e\gamma p$: BDT - Inbending torus

We extract a dataset with DVCS$\sim$ 94.5% and DVMP$\sim$ 5.5%.



Histograms are normalized to 1.

$ep \to e\gamma p$: Background subtraction

To estimate and remove the residual background on each (t, Q^2, x_B, ϕ) bin and helicity state we use two methods:

Method 1:

Let us define:

- $n^{1\gamma}_{MC/Data}$ = Number of simulated π^0 events that pass the DVCS analysis.
- $n^{2\gamma}_{MC/Data}$ = Number of simulated π^0 events that are reconstructed.

The contamination is then:

$$n^{1\gamma}_{Data} = \left(\frac{n^{1\gamma}_{MC}}{n^{2\gamma}_{MC}}\right) n^{2\gamma}_{Data}.$$

Method 2:

1. Reconstruct π^0 events.
2. For each π^0, generate 1500 decays.
3. If the event pass the DVCS analysis with any photon, fill histograms.
4. If the event pass the DVMP analysis, increment $n^{2\gamma}_{MC}$ by the reconstruction efficiency.
5. At the end of the decays, DVCS events are normalized by $1/n^{2\gamma}_{MC}$.

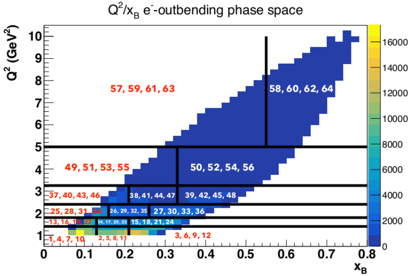
$ep \to e\gamma p$: Background subtraction

About the background subtraction:

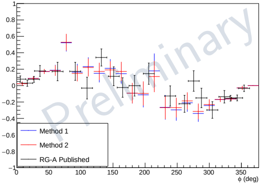
- The final estimation is given by the average of the two.
- Error on the estimation is given by the difference of each method from the average.

About the BSA measurements:

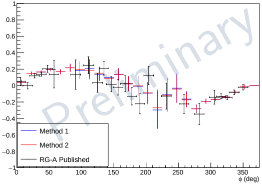
- Identical (t, Q^2, x_B) binning of the RG-A analysis note (64 in total) used for this analysis.
- Systematic uncertainties have been estimated.



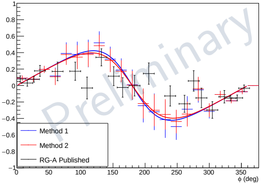
$ep \to e\gamma p$: BSA: benchmark measurements



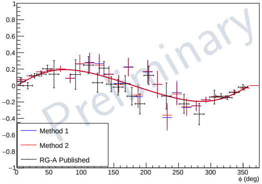
(a) Bin 53 inbending



(a) Bin 28 inbending



(a) Bin 53 outbending



(a) Bin 28 outbending

Bin 53: $3.25 < Q^2(\mathrm{GeV}^2) < 5.0, \quad x_B < 0.33, \quad 0.4 < -t(\mathrm{GeV}^2) < 0.8$

Bin 28: $1.8 < Q^2(\mathrm{GeV}^2) < 2.4, \quad x_B < 0.16, \quad 0.2 < -t(\mathrm{GeV}^2) < 0.4$

$ep \to e\gamma p$: BSA

All in all, we saw that:

- The inbending dataset can still provide an important contribution in the $Q^2 < 1.8$ GeV^2 region.
- Inbending and outbending configuration measurements are compatible.
- The background contamination after BDT is small.
- Both background subtraction methods give similar results.
- The BDT classification boost the statistics importantly.

$ep \to e\gamma(p)$: Data selection

Kinematic window:
We apply the same kinematic restrictions:

- $W > 2$ GeV,
- $Q^2 > 1$ GeV^2 ,
- $\mathbf{q}' > 2$ GeV (photon),
- $\mathbf{k}' > 1$ GeV (electron).
- $-\frac{t}{Q^2} < 1$,

Exclusivity cuts:
However, our exclusivity cuts are no longer useful.

- $\cancel{\Delta\phi = |\phi(p) - \phi(\gamma)| mod(180) < 2^\circ}$,
- $\cancel{\Delta t = |t(p) - t(\gamma)| < 2 \text{ GeV}^2}$,
- $\cancel{\mathbf{P}_{miss} < 1 \text{ GeV}}$.

Event selection

- Only analyze events with 1 or 2 photons.
- The event is selected by taking the most energetic photon and electron.

BDT training:

- Training using experimental data:
  - (Background) signal are the events that (do not) pass the analysis with proton information.
- Discriminating variables: $\{M^2_{e\gamma X}, M^2_{eX}, \text{t} \}$.

$ep \to e\gamma(p)$: Model training - Inbending

The following variables are used for training.

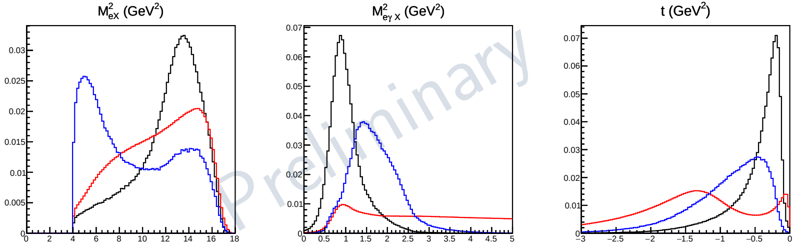


Figure: Missing masses $M^2_{e\gamma X}$, M^2_{eX} and t, normalized to 1, for raw data (red), training DVCS dataset (black) and training π^0 dataset (blue).

Histograms are normalized to 1.

$ep \to e\gamma(p)$: Background subtraction

Without proton detection, the $e\gamma$ final state receives contributions from a large set of processes. However:

1. Photon emission comes mainly neutral meson decays, being π^0 the dominant one.

$ep \to e\gamma(p)$: Background subtraction

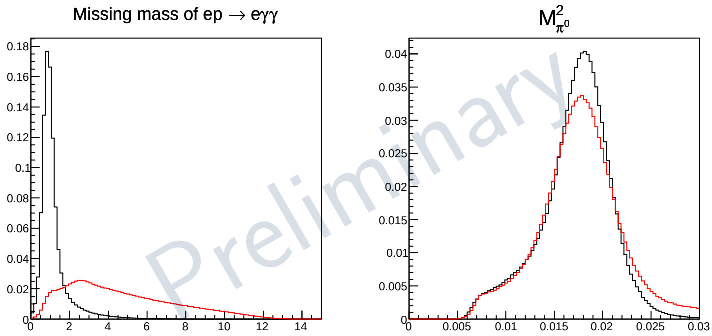
Without proton detection, the $e\gamma$ final state receives contributions from a large set of processes. However:

1. Photon emission comes mainly neutral meson decays, being π^0 the dominant one.
2. The contamination channel is now **inclusive** π^0 production.

$ep \to e\gamma(p)$: Background subtraction

Without proton detection, the $e\gamma$ final state receives contributions from a large set of processes. However:

1. Photon emission comes mainly neutral meson decays, being π^0 the dominant one.
2. The contamination channel is now **inclusive** π^0 production.
3. Both background subtraction methods are valid for such case, and it only depends on a good π^0 reconstruction.



$ep \to e\gamma(p)$: Comparison with $e\gamma p$ detection

After BDT cut and background subtraction, there is an important increase on statistics

Inbending **Outbending**

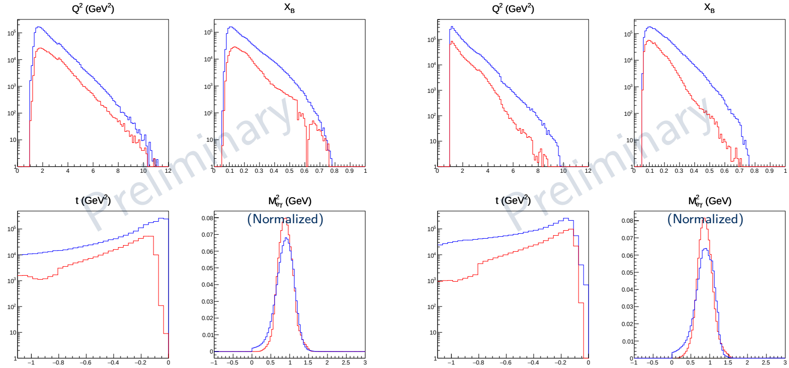
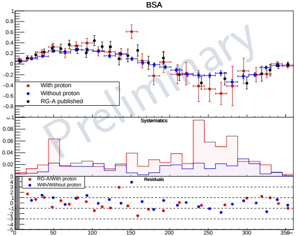


Figure: Kinematic variables for the analysis with proton (red) and without proton (blue) information.

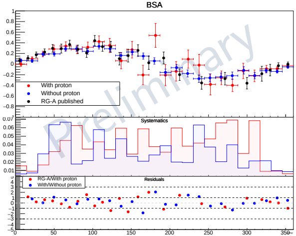
$ep \to e\gamma(p)$: BSA - Benchmark measurements

Bin 26: $1.8 < Q^2(\mathrm{GeV}^2) < 2.4$, $0.16 < x_B < 0.26$, $-t(\mathrm{GeV}^2) < 0.2$

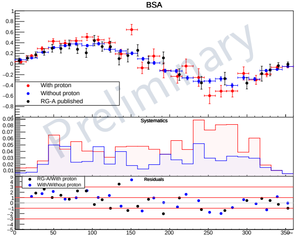
Inbending



Outbending



Merged



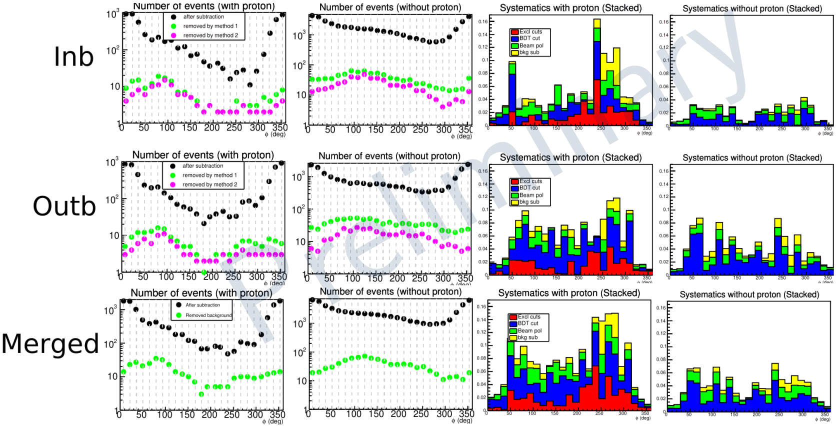
Systematic Errors

It will be done by comparing the BSA amplitude when the analysis is done with some modifications.

- Due to the exclusivity cuts:
  - Re-do the full analysis using slightly tighter selection cuts
  - $|\Delta t|$, $|\Delta\phi|$, P_{miss}.
- Due to the beam polarization uncertainty
  - Estimated to be $\sim 5\%$ of the BSA.
- Due to the choice of BDT cut.
  - Re-do the analysis using a different BDT cut.
- Due to the background subtraction.
  - $\delta A = \frac{A^{raw} - A_{\pi^0}}{(1-f)^2}\delta f$
  - f is the contamination before subtraction, $A_{\pi^0} \approx 0.05$ and δf is the estimation difference of both methods.
- Total error as the quadratic sum of components.

Systematic Errors

Bin 26: $1.8 < Q^2(\mathrm{GeV}^2) < 2.4,\ 0.16 < x_B < 0.26,\ -t(\mathrm{GeV}^2) < 0.2$



- Background subtraction methods agree.
- Systematics have decreased.

Conclusions

- Boosted decision trees presents an alternative for channel selection on an event-by-event basis.
- When the final proton is included:
  - DVCS exclusivity variables have enough separation power to allow DVCS and Deep Exclusive π^0 Production identification in an efficient way.
- When the final proton information is ignored:
  - There is a wider phase space towards the small t region.
  - There is a boost on statistics leading to more precise BSA measurements.
- Without any restriction on the detection topology we extract datasets of $\sim 95\%$ DVCS events.
- In general, results are compatible with the published RG-A results.

Outlook

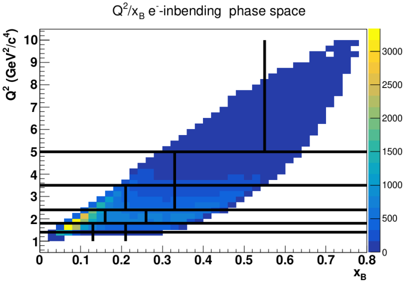
- An analysis note will be submitted to review soon.
- An analysis on pass2 data is planned as well.
- The next step is to test this method on RG-B data for nDVCS BSA measurements.

Thanks

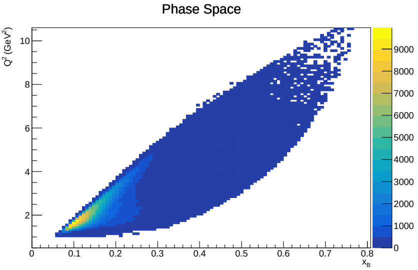
Backup

Phase space with proton information - Inbending

Let's compare the (Q^2, x_B) phase space.



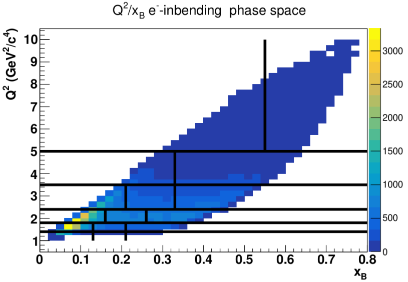
(a) RG-A analysis note.



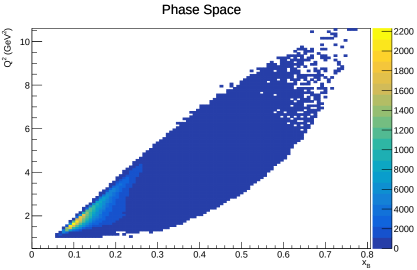
(b) Extracted phase space.

Phase space without proton information - Inbending

Let's compare the (Q^2, x_B) phase space.



(a) RG-A analysis note.



(b) Extracted phase space.

Kinematics with proton information

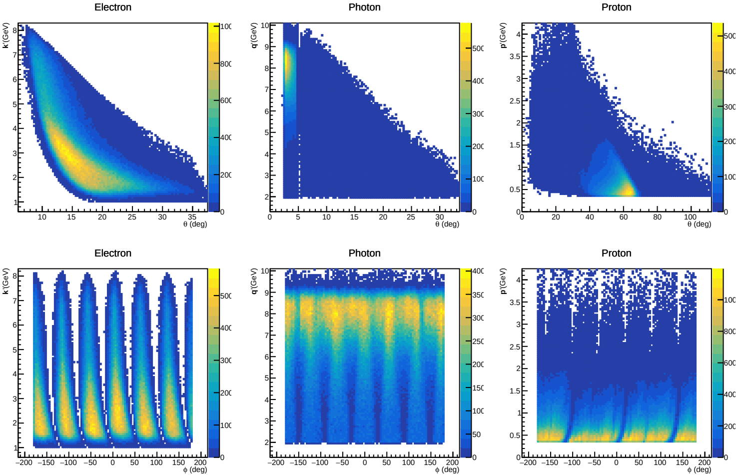


Figure: Momentum of the final particles as a function of the polar angle (first row) and detection polar vs azimuthal angle for each final state.

Kinematics without proton information

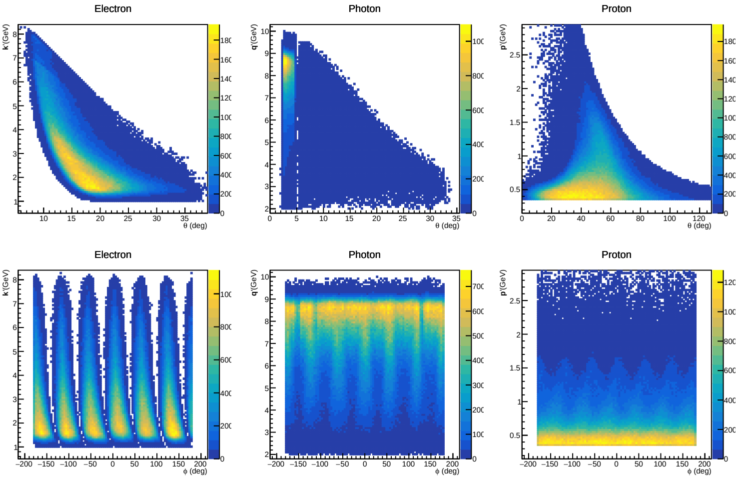
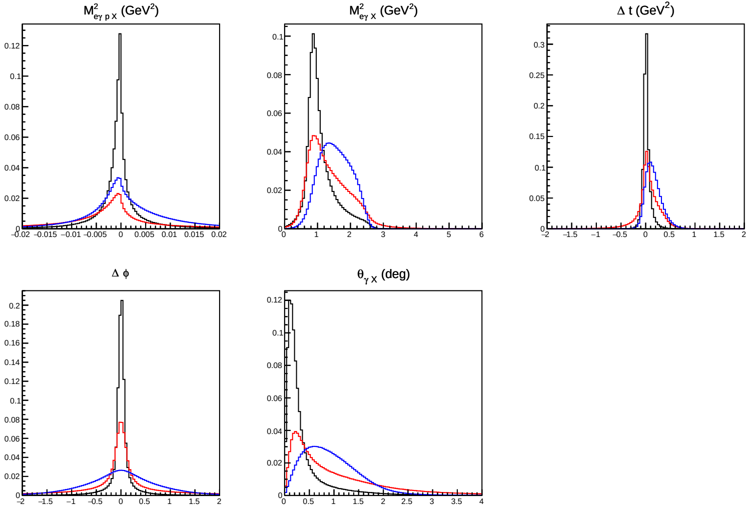


Figure: Momentum of the final particles as a function of the polar angle (first row) and detection polar vs azimuthal angle for each final state .

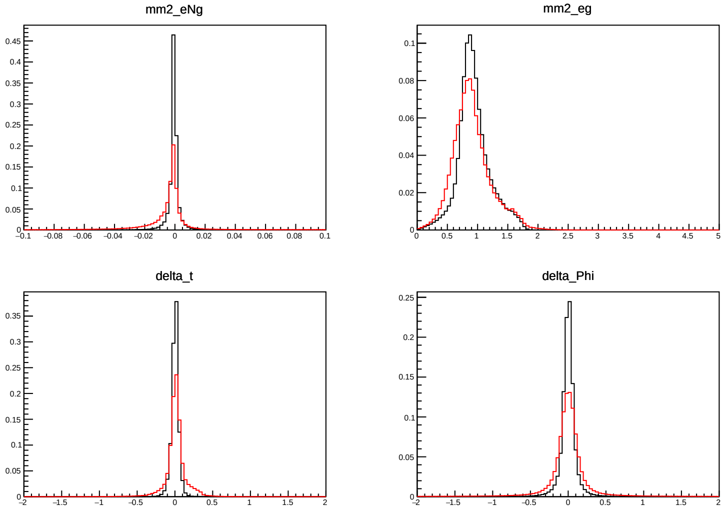
$ep \to e\gamma p$: Model training - Outbending torus

The main contamination channel is $ep \to ep\pi^0 \to ep\gamma(\gamma)$.



$ep \to e\gamma p$: BDT - Outbending torus

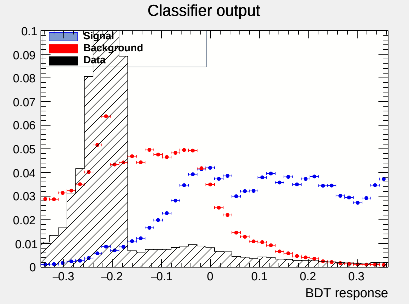
We extract a dataset with DVCS$\sim$ 96.6% and DVMP$\sim$ 3.4%.



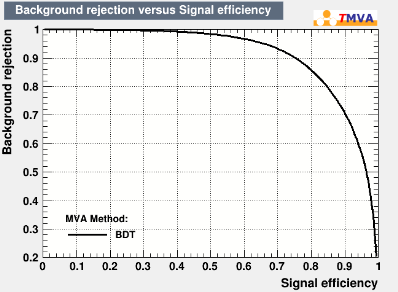
$ep \to e\gamma(p)$: NP BDT - Inbending

To optimize the DVCS event selection, a Boosted Decision Tree (BDT) is trained to classify the events.

- Discriminating variables: $\{M^2_{e\gamma X}, M^2_{eX}, t\}$.
- Simulated DVCS as signal.
- π^0 production data, reconstructed as DVCS, as background.



(a) BDT output distributions for different datasets.



(b) ROC curve of the model and applied cut.

$ep \to e\gamma(p)$: Model training NP- Outbending

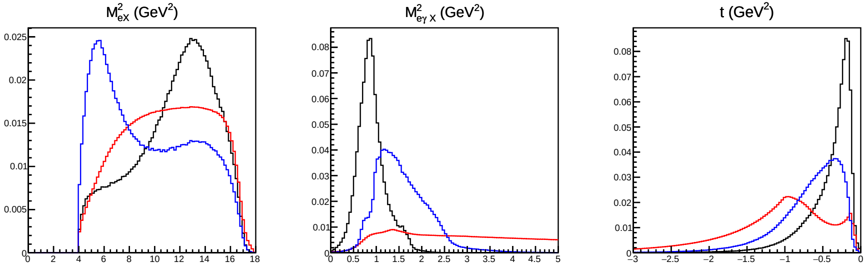
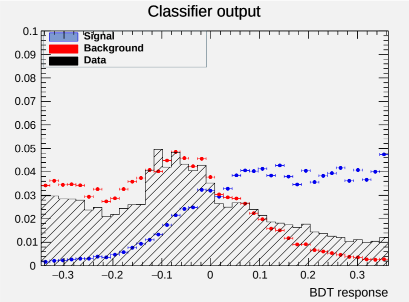


Figure: Missing masses $M^2_{e\gamma X}$, M^2_{eX} and t, normalized to 1, for data (red), training DVCS dataset (black) and training π^0 dataset (blue).

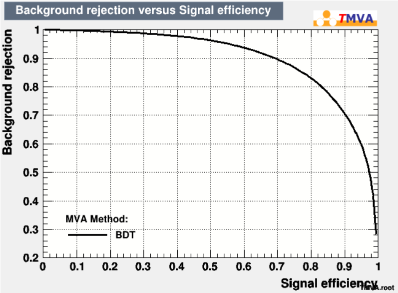
$ep \to e\gamma(p)$: NP BDT - Outbending

To optimize the DVCS event selection, a Boosted Decision Tree (BDT) is trained to classify the events.

- Discriminating variables: $\{M^2_{e\gamma X}, M^2_{eX}, t\}$.
- Simulated DVCS as signal.
- π^0 production data, reconstructed as DVCS, as background.

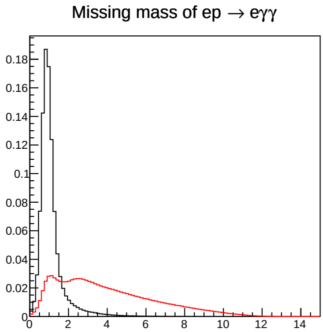


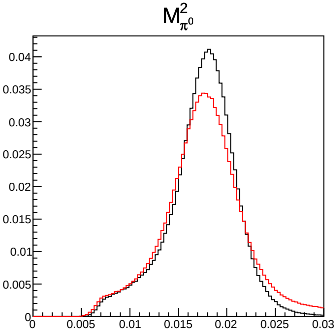
(a) BDT output distributions for different datasets.



(b) ROC curve of the model and applied cut.

eppi0 NP outbending



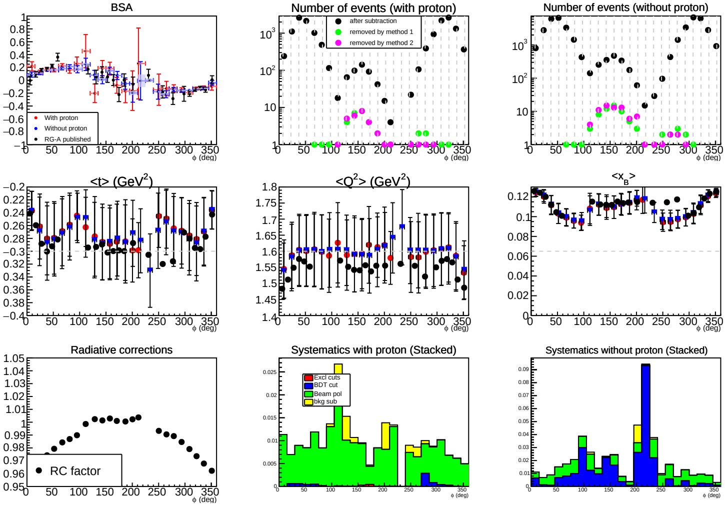


RGA bins

Bin no.	Q^2 (GeV^2)	x_B	$\|t'\|$ (GeV^2)
1		< 0.13	
2	< 1.4	0.13 – 0.21	< 0.2
3		> 0.21	
4		< 0.13	
5		0.13 – 0.21	0.2 – 0.4
6		> 0.21	
7		< 0.13	
8		0.13 – 0.21	0.4 – 0.8
9		> 0.21	
10		< 0.13	
11		0.13 – 0.21	> 0.8
12		> 0.21	
13		< 0.13	
14	1.4 – 1.8	0.13 – 0.21	< 0.2
15		> 0.21	
16		< 0.13	
17		0.13 – 0.21	0.2 – 0.4
18		> 0.21	
19		< 0.13	
20		0.13 – 0.21	0.4 – 0.8
21		> 0.21	
22		< 0.13	
23		0.13 – 0.21	> 0.8
24		> 0.21	
25		< 0.16	
26	1.8 – 2.4	0.16 – 0.26	< 0.2
27		> 0.26	
28		< 0.16	
29		0.16 – 0.26	0.2 – 0.4
30		> 0.26	
31		< 0.16	
32		0.16 – 0.26	0.4 – 0.8
33		> 0.26	
34		< 0.16	
35		0.16 – 0.26	> 0.8
36		> 0.26	
37		< 0.21	
38	2.4 – 3.25	0.21 – 0.33	< 0.2
39		> 0.33	
40		< 0.21	
41		0.21 – 0.33	0.2 – 0.4
42		> 0.33	
43		< 0.21	
44		0.21 – 0.33	0.4 – 0.8
45		> 0.33	
46		< 0.21	
47		0.21 – 0.33	> 0.8
48		> 0.33	
49	3.25 – 5.0	< 0.33	< 0.2
50		> 0.33	
51		< 0.33	0.2 – 0.4
52		> 0.33	
53		< 0.33	0.4 – 0.8
54		> 0.33	
55		< 0.33	> 0.8
56		> 0.33	
57	> 5.0	< 0.55	< 0.2
58		> 0.55	
59		< 0.55	0.2 – 0.4
60		> 0.55	
61		< 0.55	0.4 – 0.8
62		> 0.55	
63		< 0.55	> 0.8
64		> 0.55	

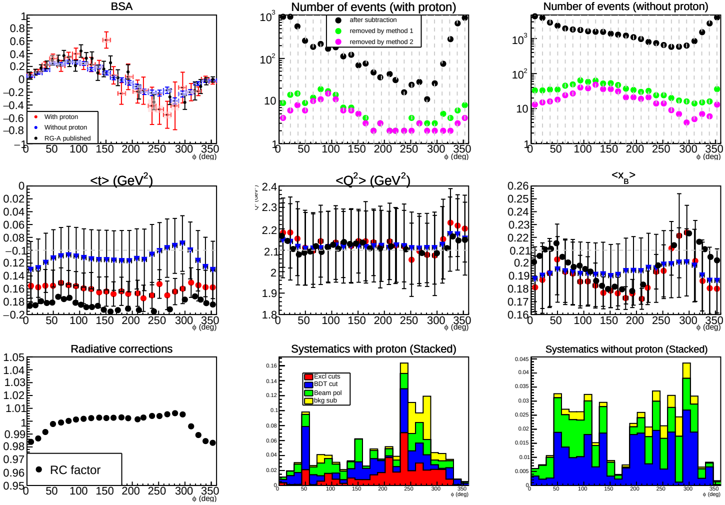
$ep \to e\gamma(p)$: BSA - Benchmark measurements

Bin 16 - Inbending: $1.4 < Q^2(\mathrm{GeV}^2) < 1.8$, $x_B < 0.13$, $0.2 < -t(\mathrm{GeV}^2) < 0.4$



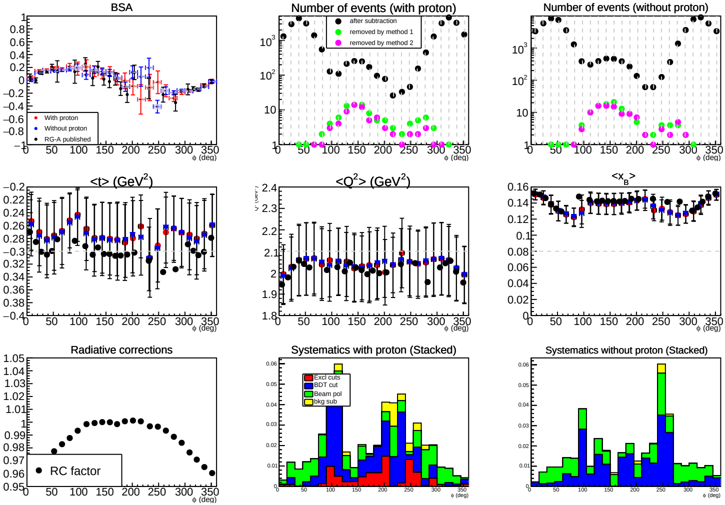
$ep \to e\gamma(p)$: BSA - Benchmark measurements

Bin 26 - Inbending: $1.8 < Q^2(\text{GeV}^2) < 2.4$, $0.16 < x_B < 0.26$, $-t(\text{GeV}^2) < 0.2$



$ep \to e\gamma(p)$: BSA - Benchmark measurements

Bin 28 - Inbending: $1.8 < Q^2(\mathrm{GeV}^2) < 2.4$, $x_B < 0.16$, $0.2 < -t(\mathrm{GeV}^2) < 0.4$



The update:

Now with the correct t definition!

k'>1 GeV?

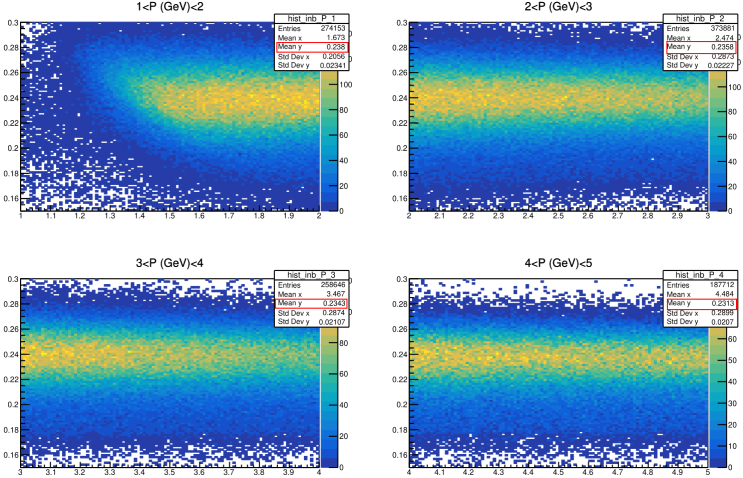
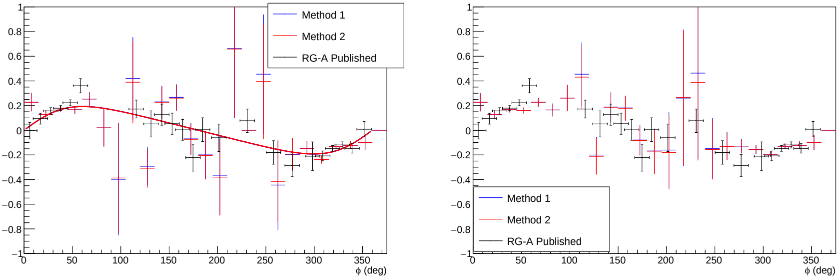


Figure: Sampling fraction vs electron momentum with proton detection on the inbending torus configuration.

SF is 0.23 **in all momentum ranges.**

k'>1 GeV?



(a) $k' > 2$ GeV **(b)** $k' > 1$ GeV

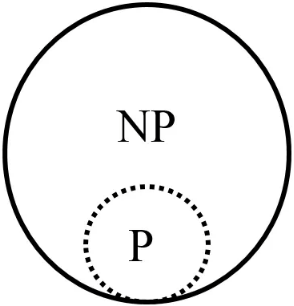
Figure: BSA at bin 16.

It may affect importantly the BSA.

BDT score per bin

About the performance...

- BDT classification without proton information keeps 80% of the events classified with proton information
- That represents 30% (40%) of the in(out)bending datasets.



η contamination

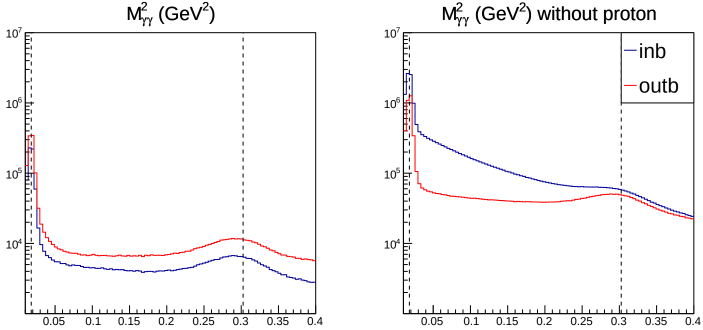


Figure: 2-photon invariant mass.

η contamination is at least 10 times smaller than π^0.

η contamination

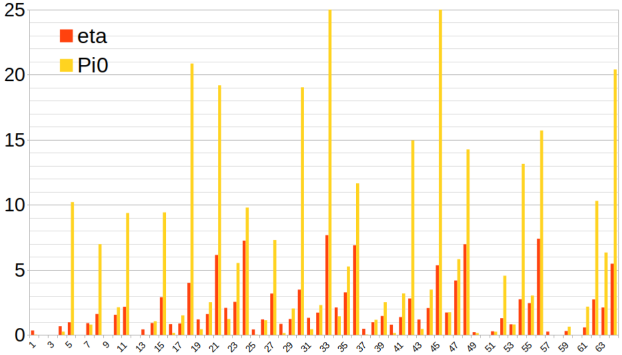


Figure: π^0 and η contamination (%) per bin after BDT for the inbending dataset without proton information.

- If proton information is included: contamination is less than 1% on all bins.
- If proton information is ignored: contamination is less than 2% on most bins. Maximum is 7%.

η contamination

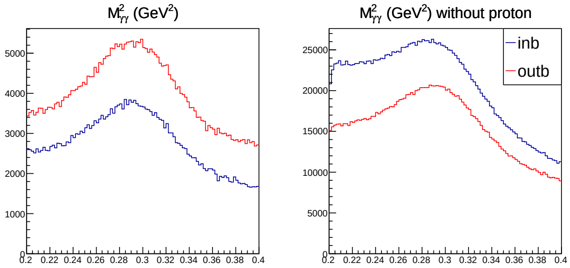


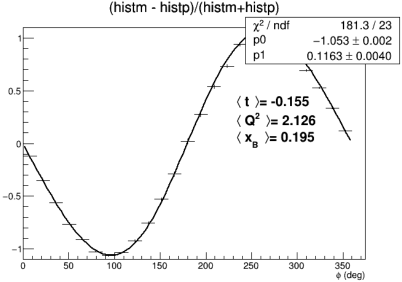
Figure: 2-photon invariant mass.

- If proton information is included: contamination is less than 1% on all bins.
- **If proton information is ignored: contamination is less than** 2% **on most bins. Maximum is 7%.**
    - However, more than half the events are from combinatorial background.
    - No subtraction was implemented then.

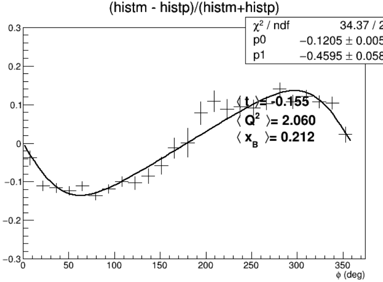
Fixed BSA cross-check

Using dvcsgen, a DVCS and π^0 asymmetry was generated.

- 1000 jobs with 10k events on each one for DVCS and π^0.
- –scale 2 , for getting a custom BSA



(a) DVCS



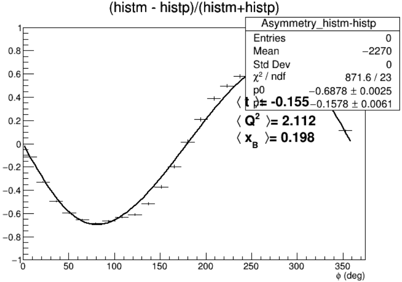
(b) π^0

Figure: Generated BSA at bin 26

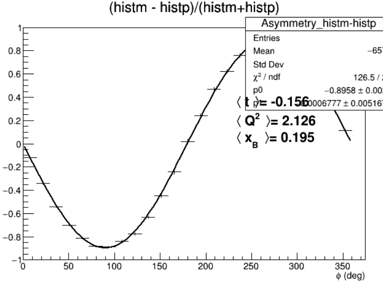
Goal is to recover unit BSA amplitude.

Fixed BSA cross-check

BSA on the combined dataset.



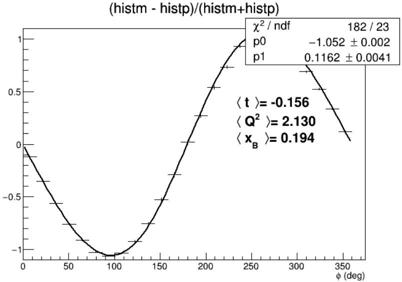
(a) Before BDT.



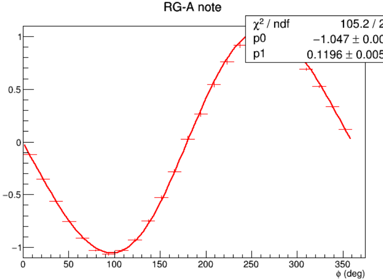
(b) After BDT.

BDT removes a big part of the contamination.

Fixed BSA cross-check



(a) DVCS after BDT.

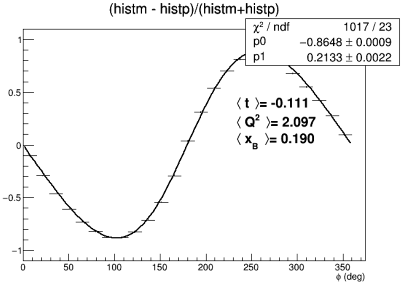


(b) After BDT analysis.

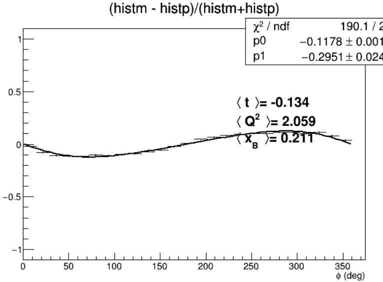
After BDT and background subtraction we recover the full amplitude.

Fixed BSA cross-check: No proton

Now ignoring the proton information:



(a) DVCS



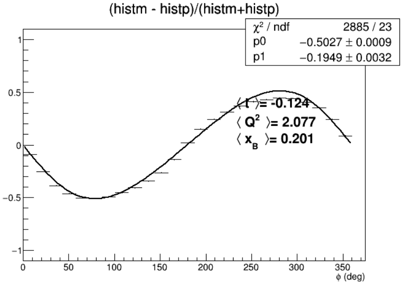
(b) π^0

Figure: Generated BSA at bin 26

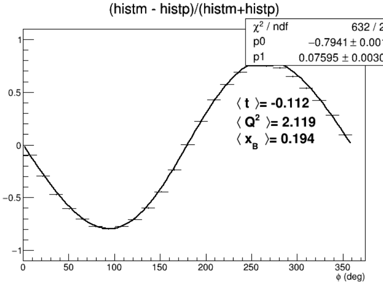
Goal is to recover unit BSA amplitude.

Fixed BSA cross-check: No proton

BSA on the combined dataset.



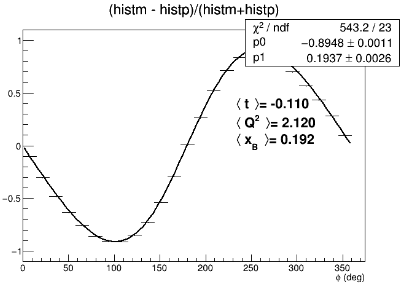
(a) Before BDT.



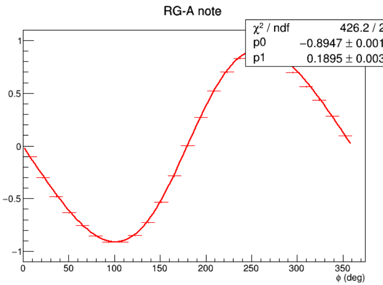
(b) After BDT.

BDT removes a big part of the contamination.

Fixed BSA cross-check: No proton



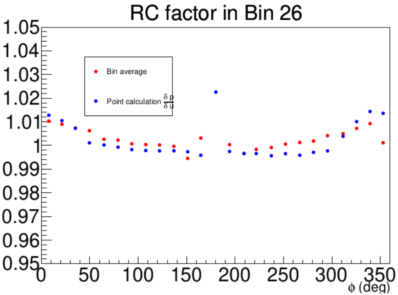
(a) DVCS after BDT.



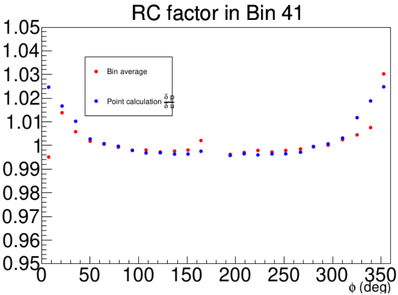
(b) After BDT analysis.

After BDT and background subtraction we recover the full amplitude.

RC factor



(a) RC factor on bin 26.



(b) RC factor on bin 41.

Computing systematics

Merging BSA $$A = \frac{\frac{A_{inb}}{\sigma(A_{inb})} + \frac{A_{outb}}{\sigma(A_{outb})}}{\sqrt{\frac{1}{\sigma(A_{inb})^2} + \frac{1}{\sigma(A_{outb})^2}}}$$ $$\sigma(A) = \frac{1}{\sqrt{\frac{1}{\sigma(A_{inb})^2} + \frac{1}{\sigma(A_{outb})^2}}}$$

Merging kin $$Q^2 = \frac{Q^2_{inb} n_{inb} + Q^2_{outb} n_{outb}}{n_{inb} + n_{outb}}$$ $$\sigma(Q^2) = \frac{\sigma(Q^2_{inb}) n_{inb} + \sigma(Q^2_{outb}) n_{outb}}{n_{inb} + n_{outb}}$$

Merging sys $$A_{\pm} = \frac{\frac{A_{inb} \pm \sigma^{cut}_{inb}}{\sigma(A_{inb})} + \frac{A_{outb} \pm \sigma^{cut}_{outb}}{\sigma(A_{outb})}}{\sqrt{\frac{1}{\sigma(A_{inb})^2} + \frac{1}{\sigma(A_{outb})^2}}}$$ $$\sigma^{cut} = \sqrt{\frac{(A_+ - A_0)^2 + (A_- - A_0)^2}{2}}$$

Bkg sub sys err $$\sigma^{bkg} = \frac{A^{raw} - A^{\pi^0}}{(1-f)^2} \delta f$$