An alternative PID and channel identification for **KaonLT** analysis

Ioana & Gabriel Niculescu,
James Madison University

Introduction

As advertised in previous updates...

- in the larger context of analyzing the 10.6 GeV **KaonLT** data & extracting Σ^0/Λ ratios (possibly excited states too!)...
- (for awhile now) the JMU group worked on improving the hyperon identification (PID & channel identification)
- we are happy(-ish) to report that the latter work is done (10.6 GeV only)
- furthermore, in the process, we developed (we adapted/improved, not **invent**!) a new PID approach, which can be used in other Hall C (and A) experiments.
- ...and this is what I would like to talk about today.

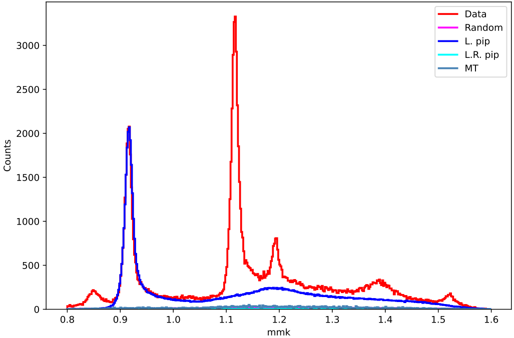
But why?

- std. ID procedure breaks down in statistically–limited bins
- can lead to biases, syst. errors unaccounted for
- so we sought a different/better/more consistent approach.

Motivation

"Standard" MM fit approach

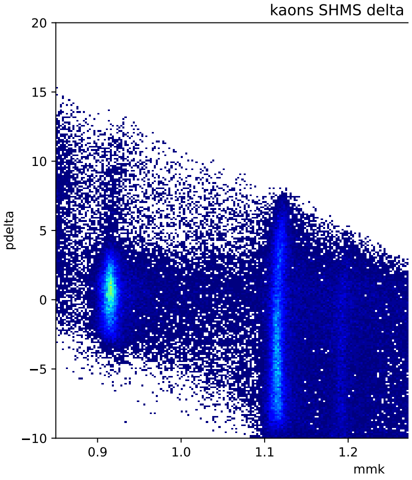
- The typical Hall C experiment (usually) involves a **2 into 3** reaction
- With only two particles detected in the final state, the third particle is **identified** using E and $\vec{p}$ conservation constraints (i.e. missing mass plot)
- After applying cuts, this might look like this:

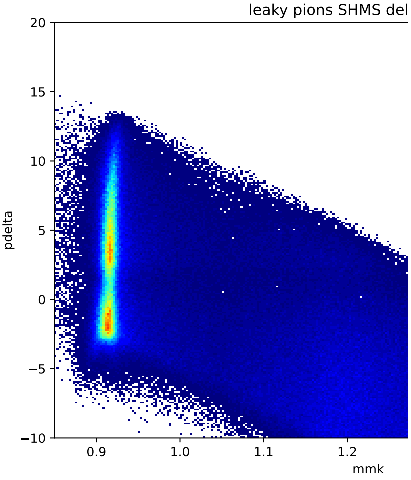


then...

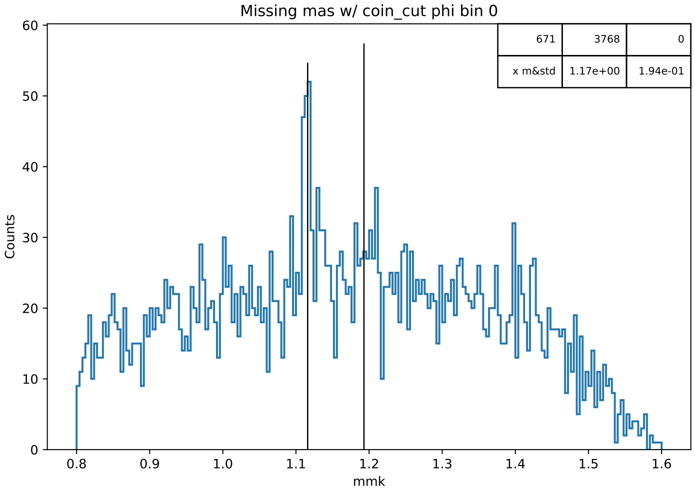
- subtract or not **leaky** π^+
- also randoms, MT, etc.
- still **irreducible** bkg.
- fit peaks+bkg.
- ...for **EACH** physics bin!

(as a side note...)





What if we make, 12 bin in ϕ (for example)?

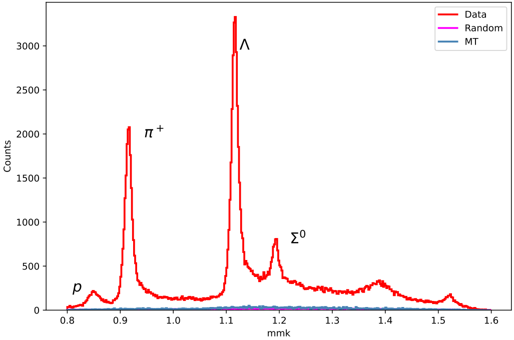


Closely & Critically examining these slides...

There a few undeniable (if uncomfortable) truths:

- leaky π^+s have completely diff. distribution, so subtracting might not be ideal
- even if subtracting..., there is still background under the peaks (**irreducible**!)
- background itself has a shape right under the peaks
- background size & shape differs from bin to bin, as do peaks!
  (physics a/o detector peculiarities...)
- inferring bkg. **things** for the whole **mm** does not help for individual bins!
- the peaks+bkg. fit need to be carried out for each bin
- ...and dwindling statistics do not help.
- Leaving aside the latter problem, let's take a look at some fits.

Two peaks and a background...



We shall:

- pick peak shape
- pick bkg. form
- set limits for the parameters of the above
- **adhere** to the choices!!!
- do a **mm** fit for the whole setting (no bins) (in the Λ & Σ^0 region)
- ...repeatedly

Two peaks and a background (II)

Model selection

- there are many diff. choices for both **peak** and **background** shapes
- all of them come w/ +-es and −-es...
- which do not go away only because we happen to like a particular choice!
- on the next couple of pages...
- we list our choice & possible alternatives
- *Cave cadas!*: (choices that seem linear but aren't!)

Peak

Peak model:

- **skewed Voigt profile** - the convolution between a Cauchy–Lorentz distribution and a (skewed) Gaussian. Centered **vp**, peaking at zero:

$$V(x;\sigma,\gamma) = \int_{-\infty}^{\infty} G(x';\sigma)L(x-x';\gamma)dx'$$

$$G(x;\sigma) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} \qquad L(x;\gamma) = \frac{\gamma}{\pi(\gamma^2+x^2)}$$ see Wikipedia for extra details

- ours uses a **skewed Gaussian** so:

$$f(x;A,\mu,\sigma,\gamma,skew) = V(x;A,\mu,\sigma,\gamma)\left(1 + erf(\frac{skew(x-\mu)}{\sqrt{2}\sigma})\right)$$

- **NOTE:** we keep $\gamma = \sigma$ (default option)

Peak (II)

Alternative peak choices:

- *Vanilla* **Gaussian** - only the tip of the distribution
- **Lorentzian** - might be too narrow (not account for resolution effects)
- **skewed** versions of the above (better but w/ same reservations)
- simc–based **template** fit. tempting, but beware of the *ouroboros!*

Background

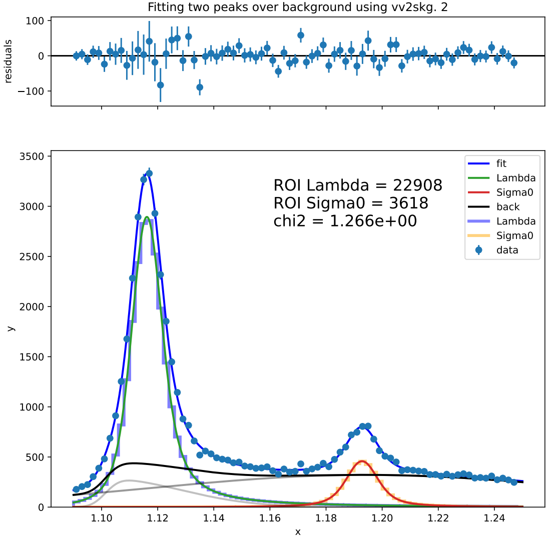
Background model. Huge array of choices:

- polynomial (degree? + -ity constraint?)
- **Chebyshev** polynomial. better. truncated where?
- **cubic splines**. # of nodes? location??
- **template** - based on a better/more complete **simc** for bkg. proc. errors?
- **Gaussian** representation. i.e. sum of Gaussians. (how many?)
- **skewed Gaussian** sum. Our choice.

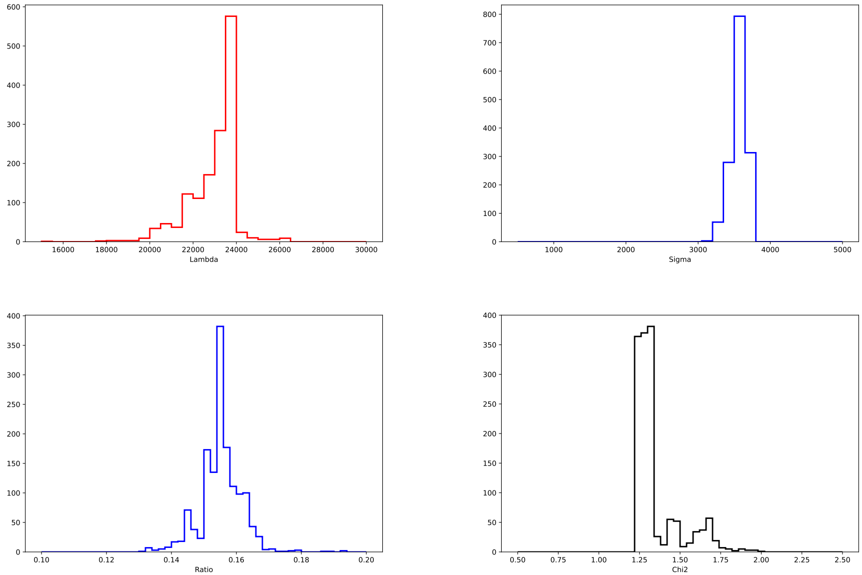
NOTE 1: Regardless of peak & background choice, the fit is **NOT** linear! Therefore the outcome will depend on the initial *ansatz* for the parameters.

NOTE 2: Look for yield variation within a given choice.

w/ our choices: 10,000 fits!



10k 2p+bkg. Summary plots



Therefore...

With given choice of peak and background:

- one gets a 10–15% variation in the yield
- ...at full statistics (could get worse for smaller bins!!!)
- this is a systematic, point–to–point effect

But wait...

- Inspired by these...

[2] M. Williams, M. Bellis, and C. A. Meyer, "Multivariate side–band subtraction using probabilistic event weights", JINST **4**, P10003, 2009. (see also arXiv:0809.2548v3 [nucl-ex])

[3] P. Roy *et al* CLAS collaboration, "Measurement of single- and double- polarization observables in the photoproduction of $\pi^+\pi^-$ meson pair off the proton using CLAS at Jefferson Laboratory", arXiv:2504.21119v1, April 2025.

[4] S. Adhikari *et al*, "Measurement of beam–recoil observables C_x and C_z for $K^+\Lambda$ photoproduction", Phys.Rev.C, 2025

[5] Z. Baldwin, "A multidimensional, event–by–event, statistical weighting procedure for signal to background separation", EPJ web of conferences **295**, 06002, 2024.

[6] M. Albrecht *et al*, "Coupled channel analysis of $\bar{p}p \to \pi^0\pi^0\eta$, $\pi^0\eta\eta$, and $K^+K^-\pi^0$ at 900 MeV/c and of $\pi\pi$–scattering data", arXiv:1909.07091v2 [hepex], 2020.

[7] Ablikim *et al*, "Evidence of doubly OZI–suppressed decay $\eta_c \to \omega\phi$ in the radiative decay $J/\Psi \to \gamma\eta_c$", arXiv:2504.01823v1, 2025.

- We think we found a better way!

Introducing the Quality Factor Fitting (Q–fit)

Projecting on the missing mass dimension...

- Looses all the additional information associated with the events
- The **mm** lumps together events from different portions of the acceptance
- also lumps togethere^-K^+ pairs w/ potentially diff. $\vec{p}$
- which we know w/ good precision!
- Furthermore, **lumping** in 1D also exacerbates problems related to physics & detector–specific effects

In contrast

Q–fitting...

- Leverages the additional (high precision) information of the measurement
- It is **bin–less** (bin agnostic?)
- Provides a clear way of evaluation the uncertainties associated with the fit
- (and the systematic unc. is bin–independent!)
- Wide(-ish) spread usage in the field

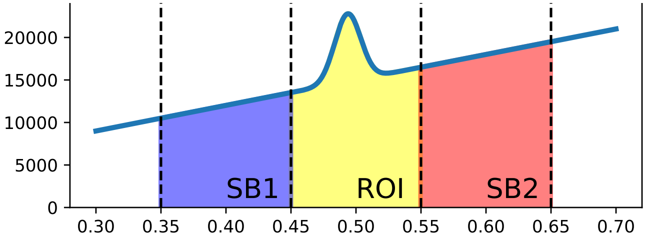
...

- Yes, it is CPU–intensive.
- But: “One cannot/doesn’t have to be stuck in the VAX/VMS era”! (IN)
- Especially given the issues of w/ the “standard” yield extraction procedure.
- **Hopefully you are interested/intrigued enough to find out more...**

Enter the Q–fit

Q–fit method

- aims to address the peak–over–background issue...
- by assigning a **quality factor** (Q–factor) for each event
- once found, the Q–factor can be used as a weight in further analysis
- it is (as it will become apparent later) a generalization of the **side–band** subtraction method (see below)



Q–fit (IV)

How does it work?

- After applying all cuts data contains **signal** and **irreducible** background
- Data is multidimensional: $\xi_1, \xi_2, ..., \xi_n$
- There is a dimension in which the **s** & **b** can more easily be distinguished: The reference variable, ξ_r (**mm**, **im**, **m**?). So: $\xi_1, \xi_2, ..., \xi_n = (\xi_r, \xi_{nr})$
- Assume* that **s** & **b** can be parameterized, up to unknown constants, in ξ_r
- No binning in ξ_{nr} is needed!
- Instead, and this is the **crux** of the method, the **N–nearest neighbors** (based on hyper–distances measured in the non–reference sub–space) of event i are used for the fit!!
- Provides statistical & systematic uncertainty associated with Q_is

Q–fit (V)

How does is work? (cont...)

- Manifestly unbinned, completely circumvents the low stat. bin issues
- ...assuming that a suitable number of neighbors (N_n) can be found
- Data–centered, minimal assumptions on **s** or **b**
- Generally selects a **tiny** portion of phase space (simpler background shape!)
- **GN's $0.02:** "Each event is judged (in the **s** or **b** sense) by a jury of its peers!"
- So an event w/, say, the electron @ $\Delta = -5\%$ is compared w/ events whose electrons are also in the same range, not at $\Delta = 5\%$!!
- *ditto* for all other dimensions of ξ_{nr}
- **Warning:** There is math on the next slide!

Q–fit (VI). the math.

Some notations. Some equations.

$$d^2 = \sum_{k=1}^{n} \left(\frac{\xi_k^i - \xi_k^j}{r_k} \right)^2 \quad (1)$$

- hyper–distance in the non–ref. subspace, r_k - **norm** for dimension k

$$R_k^i = \frac{1}{N} \sqrt{\sum_{j=0}^{N} \left(\frac{\xi_k^i - \xi_k^j}{r_k} \right)^2} \quad (2)$$

- **reduced hyper–sphere radius** in the k coordinate
- average R_k^i for all events in the sample to get:

$$R_k = \frac{1}{M} \sum_{i=0}^{M} R_k^i \quad (3)$$

- the average radius of the hyper–sphere in the k coord.

more math.

Defining Q_i

- With this notation, the data distribution in ξ_r becomes:

$$f(\xi_r) = f_s \cdot S(\xi_r) + (1 - f_s) \cdot B(\xi_r) \tag{4}$$

- f_s: fitted signal fraction. S : **signal**, B: **background** (duh!)
- Fit eq. 4 to the sample of N_n, obtaining the parameters for S and B **for each event** in the ROI!
- Calculate the Q_i for the event as:

$$Q_i = \frac{f_s^i \cdot S(\xi_r^i)}{f(\xi_r^i)} \tag{5}$$

- **N.B.:** Eq. 5 needs to be evaluated at the ξ_r of the current event i!!
- That is it!
- **Q: Does it work? A: Thank you for the question! Let's take a look.**

Q–fit implementation details

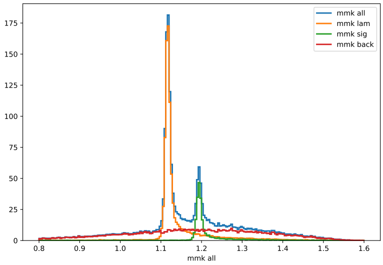
...in the JMU KaonLT software framework

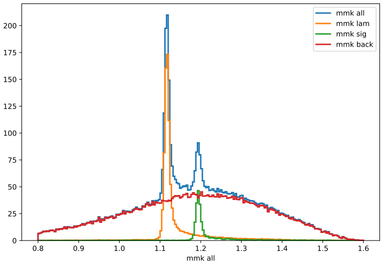
- ξ_r: **mmk**. ξ_{nr} choice: could have done Q^2, W, t, ϕ, etc...
- opted* for 7D ξ_{nr}: E_z^{beam}, $2 \times (\Delta,\ xptar,\ yptar)$ for HMS/SHMS
- tightly integrated w/ the framework
  (working on making the framework ready for release to the public!)
- most fit options controllable via config file (ROI lims, N_n, pdf output freq...)
- adds extra columns to the DST (Q#, R#e, Q#e, etc.)
- uses **lmfit** (which implements Voigt, skg, etc. already!)
- looking to add **iminuit2** option
- fairly fast (day–ish at most for the largest statistics settings)

MC validation

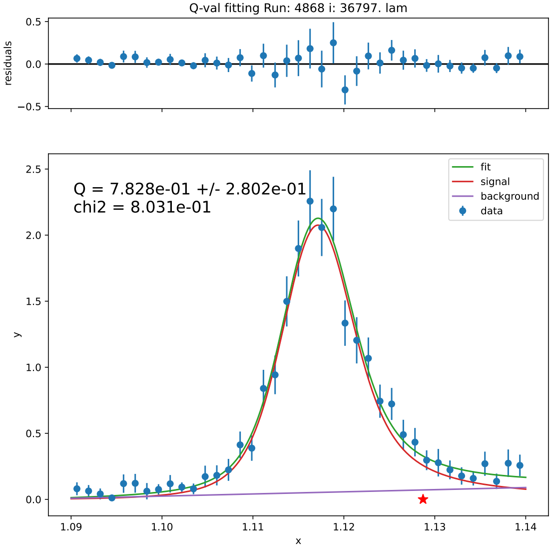
To test the procedure:

- prepare a sample of Λ, Σ^0 and background **simc** events
- the latter obtained by scrambling e and K^+ between events
- as we know the GT (ground truth) yield, we should be able to compare it w/ the eventual Q–fit yield.

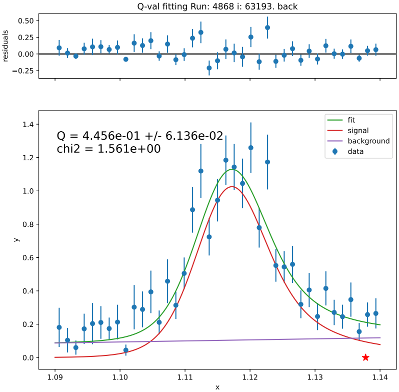


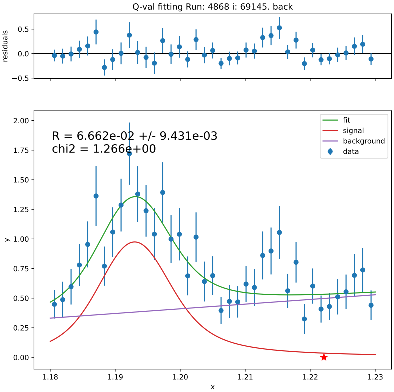


Sample Q–fit

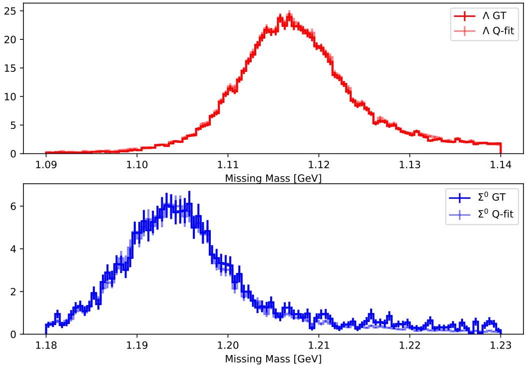

Q-val fitting Run: 4868 i: 36797. lam
residuals
0.5
0.0
−0.5
Q = 7.828e-01 +/- 2.802e-01
chi2 = 8.031e-01
fit
signal
background
data
y
2.5
2.0
1.5
1.0
0.5
0.0
1.09
1.10
1.11
1.12
1.13
1.14
x

Λ and Σ^0 fits.





simc Q–fit yields



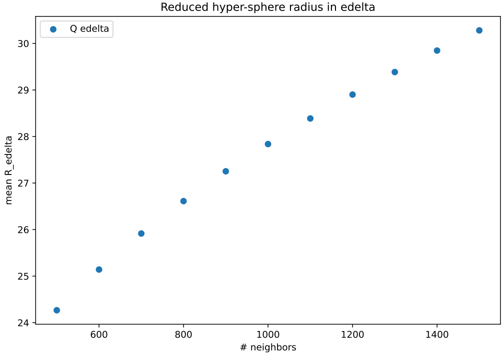
GT & Q–fit overlay

- The Q–fitted distribution matches well the GT for both hyperons.
- Several kin. & bkg. levels: Q–fit matches GT w/ **3–5%** precision

N_n study

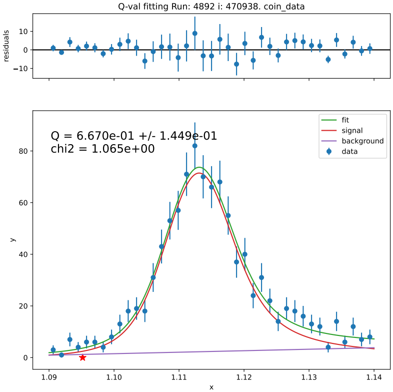
- redo Q–fit for several N_n choices
- examine the average hyper–sphere radius

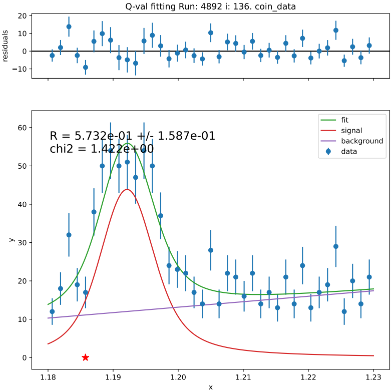
```
my_ref_var = mmk
my_var = edelta eph eth pdelta pph pth ebeamz
my_res = 5.e-2 2.e-4 2.e-4 5.e-2 2.e-4 2.e-4 5.e-5
```



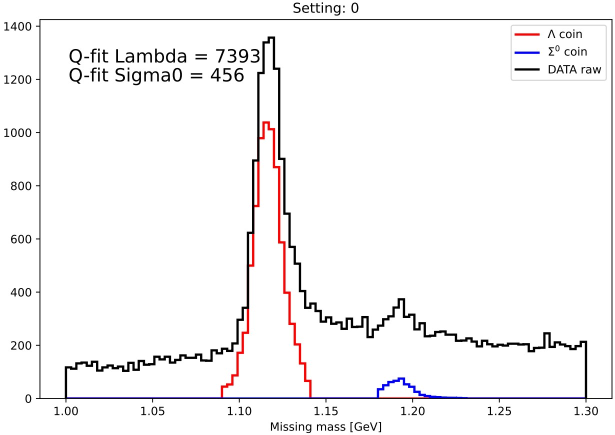
So our choice of N_n, 1000, selects 1/5–1/10th of the range!

DATA Λ and Σ^0 fits



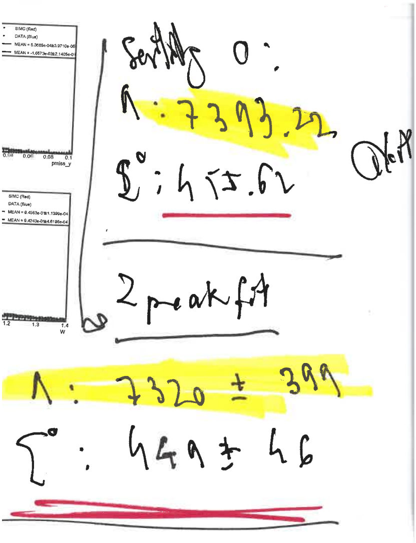


Full setting signal–background separation *à la* Q–fit



- looks not too bad/plausible
- restricted to the respective ROIs (see discussion a few weeks back about data/simc tails...)

Comparing 2–peak and Q–fit yields



Translating in a more readable form...

- Setting 0:
- Q–fit yields:
- $\Lambda = 7383.22$
- $\Sigma^0 = 455.62$
- 2-peak fit:
- $\Lambda = 7329 \pm 399$
- $\Sigma^0 = 449 \pm 46$
- *q.e.d.!*

Summary

I hope I showed/convinced you that:

- an in–depth analysis of the “classic” yield extraction method
- reveals substantial & often neglected systematic unc.
- introduced an adapted form of the **Q–fit** method for Hall C work: unbinned, event–by–event, good syst. control
- validated **Q–fit** with **simc** events
- compares well w/ “classic” method on **KaonLT** data

Quo Vadis?

- adapt method for excited hyperon identification
- explore **iminuit2** option
- paper (journal TBD) on the method

THANK YOU!