Seeding in the Tracker Subsystem: Updates on the Seeding Algorithm depending on the Signal Distribution

LUXE Simulation and Analysis Meeting

Arka Santra, December 30, 2020
Weizmann Institute of Science,
Rehovot, Israel

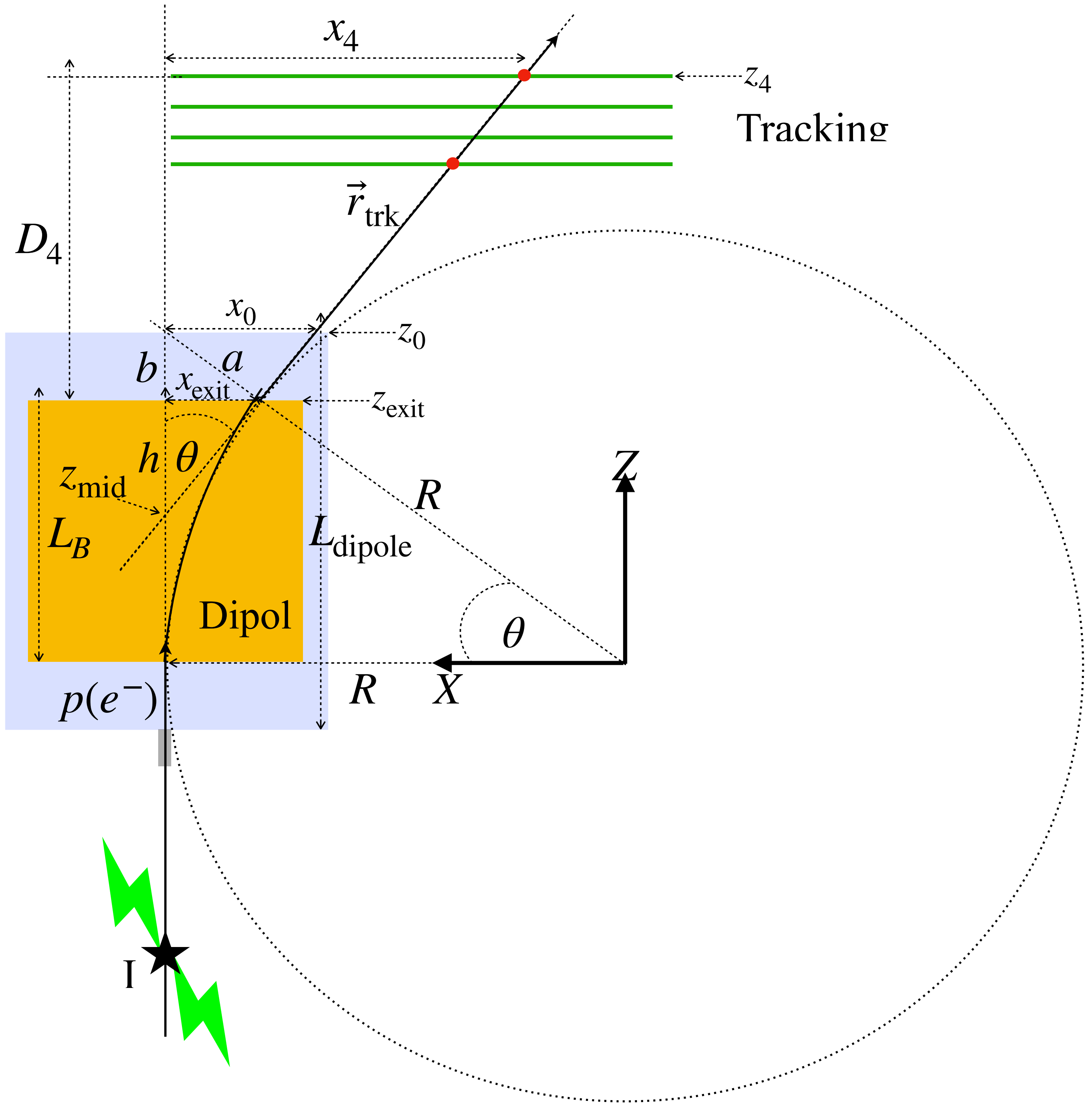
Setting up the stage:

- Motion of electron in a uniform magnetic field
  - $qBv = mv^2/R \longrightarrow p = qBR \longrightarrow p[\mathrm{GeV}] = 0.3 \cdot B[\mathrm{T}] \cdot R[\mathrm{m}]$
- Getting the momentum from two measurements in the first layer ($\vec{r}_1$) and the last layer ($\vec{r}_4$) of the tracker
- Getting the predicted position in the exit plane and in the last layer
  - Depending on the predicted position from the signal, coming up with a new distance cut (details later).
- The track fit is done with the Single Value Decomposition (SVD).
  - The minimum of singular value corresponds to the solution of the linear least square.
  - A flat cut on the fit parameter in many signal tracks does not perform well.
    - Good for high track multiplicity but bad for low track multiplicity bunch crossing
    - Solution is to come up with different cuts on the fit parameter for different track multiplicity.
      - Within 5% is the relative difference between actual signal tracks and the seed tracks found from the algorithm — irrespective of the signal track multiplicity.

Positions prediction, Using More Realistic Dipole Setup from GEANT4

$$p[\mathrm{GeV}] = 0.3 \cdot B[\mathrm{T}] \cdot R[\mathrm{m}]$$

$$R = \frac{p}{0.3B}$$



From Noam

Circle equation wrt the origin at the centre of the circle defined by the track: $X^2 + Z^2 = R^2$

$$Z_{\mathrm{exit}} = L_B \longrightarrow X_{\mathrm{exit}} = \sqrt{R^2 - L_B^2}$$

$$x_{\mathrm{exit}} = R - X_{\mathrm{exit}} = R - \sqrt{R^2 - L_B^2}$$

Tangent equation: $Z = m \cdot X + c$. The tangent gradient, m, is -1 over the gradient of the radius at the point where the tangent is defined, i.e.: $m = -1/(\Delta Z/\Delta X)$ at the point $\left(Z_{\mathrm{exit}}, X_{\mathrm{exit}}\right)$

$$m = -1/(\Delta Z/\Delta X) = -\left(X_{\mathrm{exit}} - 0\right)/\left(Z_{\mathrm{exit}} - 0\right) = -\left(\sqrt{R^2 - L_B^2}\right)/L_B = -\sqrt{\frac{R^2}{L_B^2} - 1}$$

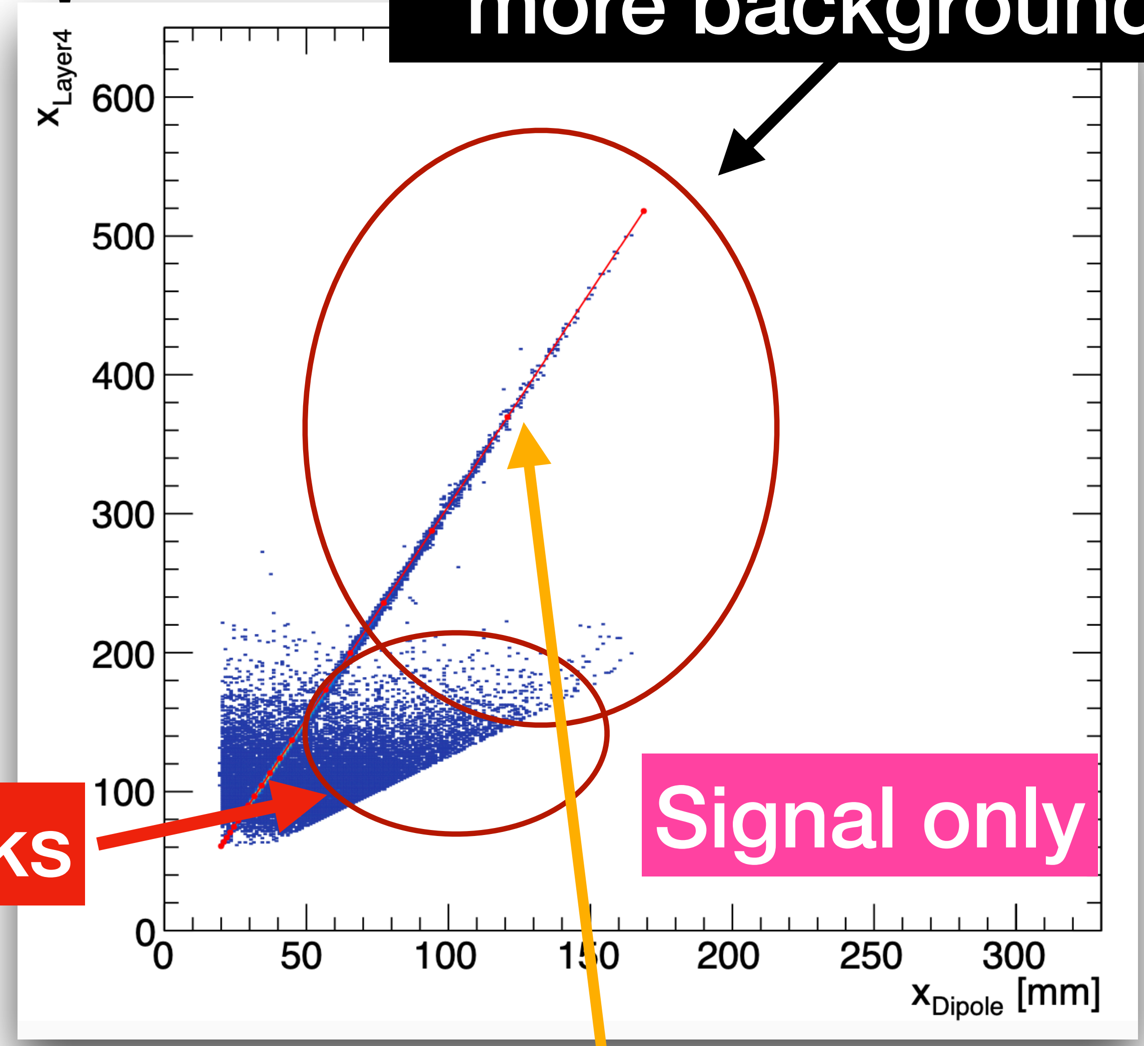
Using the point $\left(Z_{\mathrm{exit}}, X_{\mathrm{exit}}\right)$ we get the intersection: $c = Z - m \cdot X$

$$c = L_B - \left(-\sqrt{\frac{R^2}{L_B^2} - 1}\right) \cdot \sqrt{R^2 - L_B^2} = L_B + \frac{R^2 - L_B^2}{L_B} = \frac{R^2}{L_B}$$

$$Z_{\mathrm{tangent}} = -\sqrt{\frac{R^2}{L_B^2} - 1} \cdot X_{\mathrm{tangent}} + \frac{R^2}{L_B}$$

$$X_{\mathrm{tangent}} = \left(\frac{R^2}{L_B} - Z_{\mathrm{tangent}}\right) \frac{L_B}{\sqrt{R^2 - L_B^2}}$$

Want to exploit this correlation to reduce more background

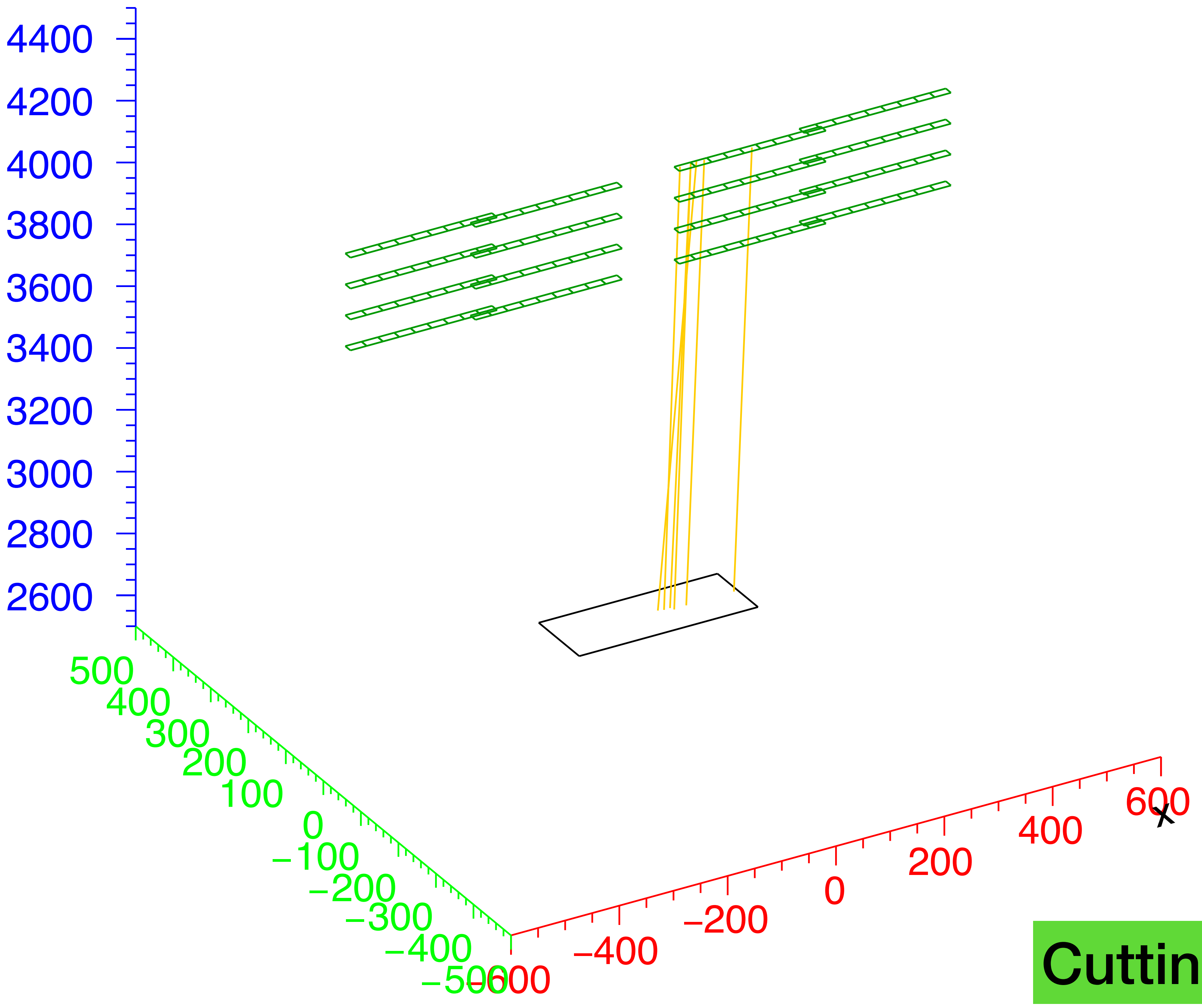


Wrong combination of tracks

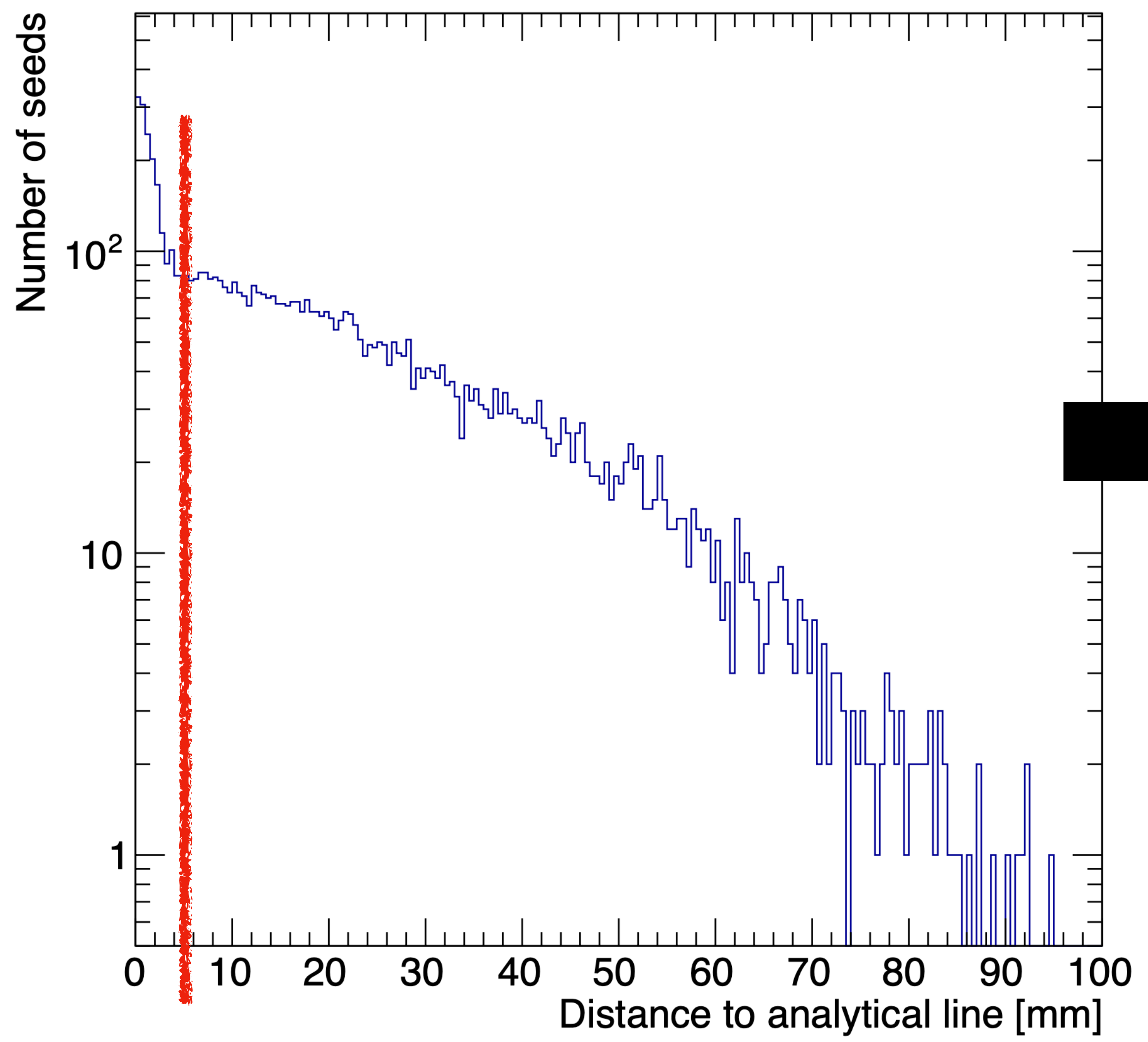
Putting $Z_4 = L_B + D_4$ or $Z_0 = L_B + \frac{L_{\mathrm{dipole}} - L_B}{2} = \frac{L_B + L_{\mathrm{dipole}}}{2}$ or $Z_{\mathrm{exit}} = L_B$ in the X_{tangent} expression, it is possible to get the prediction for the distance x from the beam axis ($z = 0$), recalling that: $x = R - X_{\mathrm{tangent}}$

Any point far away from this line is not interesting

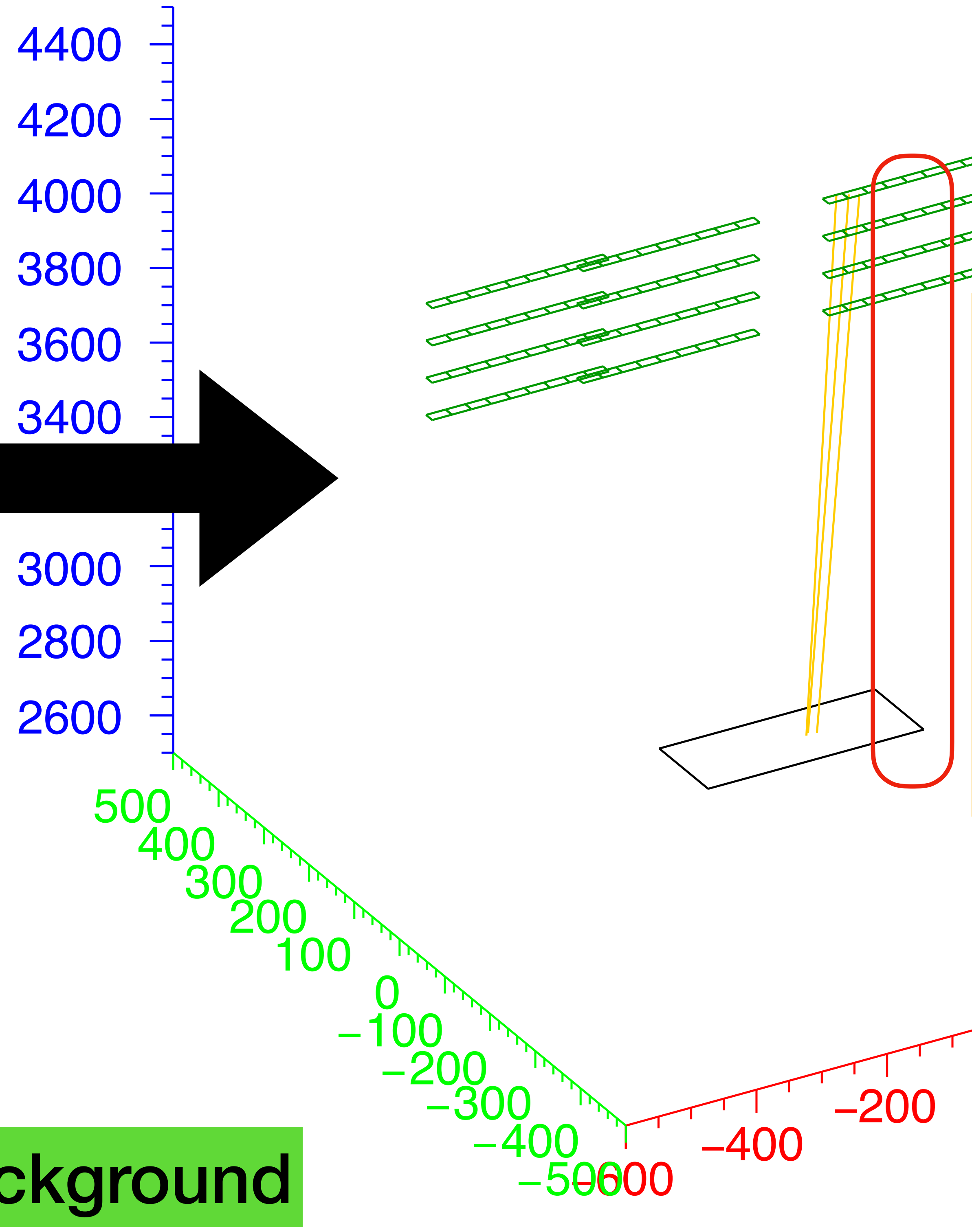
Plot from Electron Beam only Setup: Effect of Distance Cut on the Background



Electron beam only case without any cut on the distance from x_Layer4:x_Dipole straight line



Cutting around d < ~5 mm reduces the background



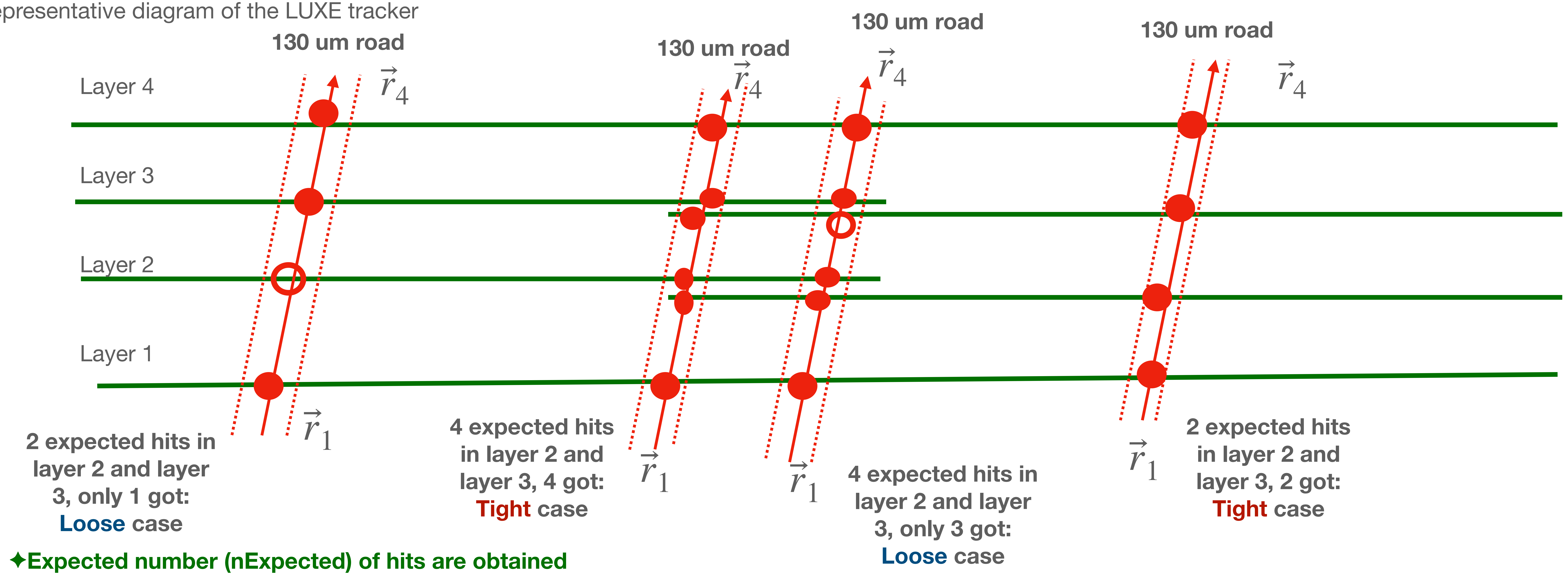
Removes track from edges, which are not possible to occur in a magnetic field

Electron beam only case with a cut on the distance from x_Layer4:x_Dipole straight line: d < 5 mm

Tight Tracks and Loose Tracks Scenario Depending on the Hits in the Tracker Layers

- The seeding algorithm always starts from one hit in layer 4 (outer layer) and one hit in layer 1 (inner layer).

Representative diagram of the LUXE tracker



✦Expected number (nExpected) of hits are obtained
by the projection of line joining $\vec{r}_1$ and $\vec{r}_4$ and see how many hits expected in layer 2 and layer 3 given the direction.
if the projection road goes through the inactive material between two chips, then expected hit in that layer is 0.

✦Tight case if expected hit equals to obtained hit (nObtained == nExpected)

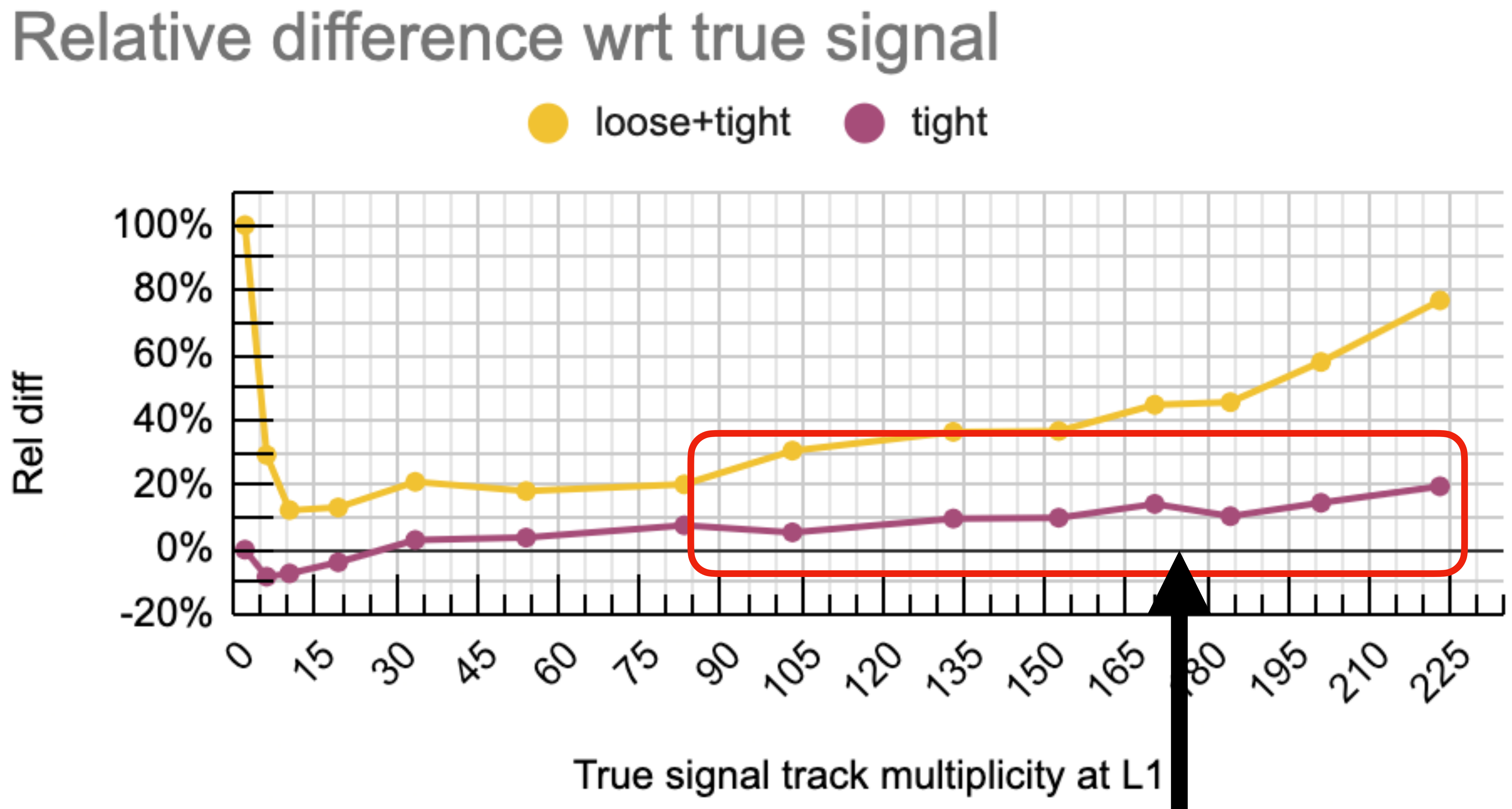
✦Loose case if obtained hit is one less than expected hit (nObtained == nExpected - 1)

✦Any other scenarios are rejected from further processing.

The First Run with the Updated Algorithm

- **Usual Seeding Algorithm cuts (discussed before, in backup)**
  + SVD parameter cut
  Second fit parameter < 0.1 and Third fit parameter < 0.05
  + distance from the expected analytical line in $x_{dipole} : x_{layer4}$ line
  The distance d < 5 mm

Approximate multiplicity of true signal tracks	--- Summary --- Sig, # of Tru	S+B, # of seeds w/o fit	loose+tight S+B	loose+tight Rel diff loose+tight	tight S+B	tight Rel diff tight	loose S+B	loose Rel diff loose
1	2.0	22.0	4.0	100%	2.0	0.0%	2.0	0.0%
5	6.0	25.5	7.8	29%	5.5	-8.3%	2.3	-62.5%
10	10.3	30.0	11.5	12%	9.5	-7.3%	2.0	-80.5%
20	19.3	42.8	21.8	13%	18.5	-3.9%	3.3	-83.1%
30	33.5	73.0	40.5	21%	34.5	3.0%	6.0	-82.1%
50	54.0	104.5	63.8	18%	56.0	3.7%	7.8	-85.6%
80	83.3	165.0	100.0	20%	89.5	7.5%	10.5	-87.4%
100	103.3	224.8	134.8	31%	108.8	5.3%	26.0	-74.8%
130	133.0	316.0	181.3	36%	145.8	9.6%	35.5	-73.3%
150	152.5	373.3	208.3	37%	167.5	9.8%	40.8	-73.3%
170	170.3	466.3	246.3	45%	194.3	14.1%	52.0	-69.5%
185	184.3	516.3	268.0	45%	203.3	10.3%	64.8	-64.9%
200	201.0	632.5	317.3	58%	230.0	14.4%	87.3	-56.6%
220	223.0	819.3	394.3	77%	266.5	19.5%	127.8	-42.7%



Constant rise in relative difference for high track multiplicity

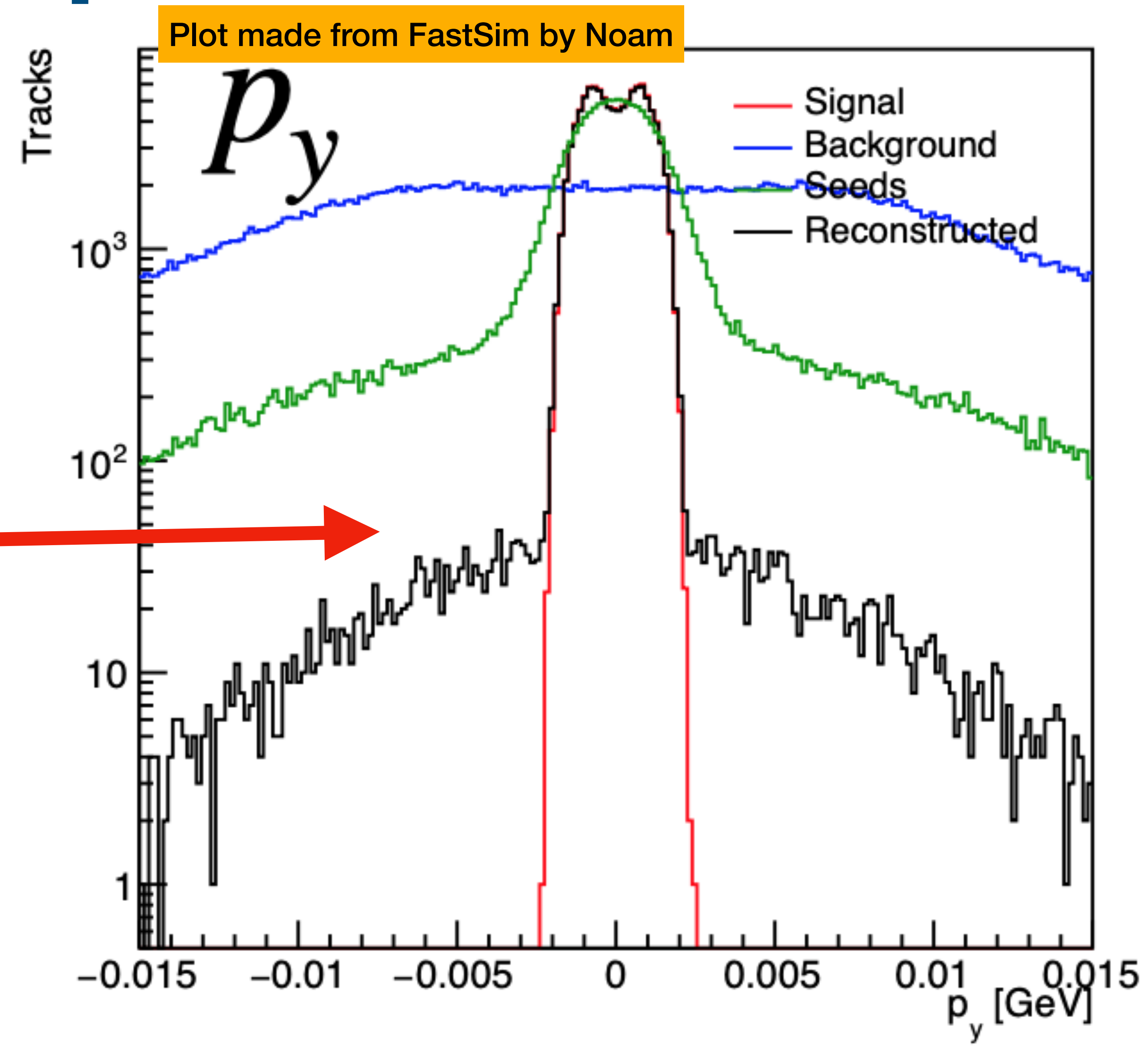
Need to tighten the SVD cuts and distance cuts

The Run with the Tighter Cuts on SVD parameters and Distance d (and p_Y)

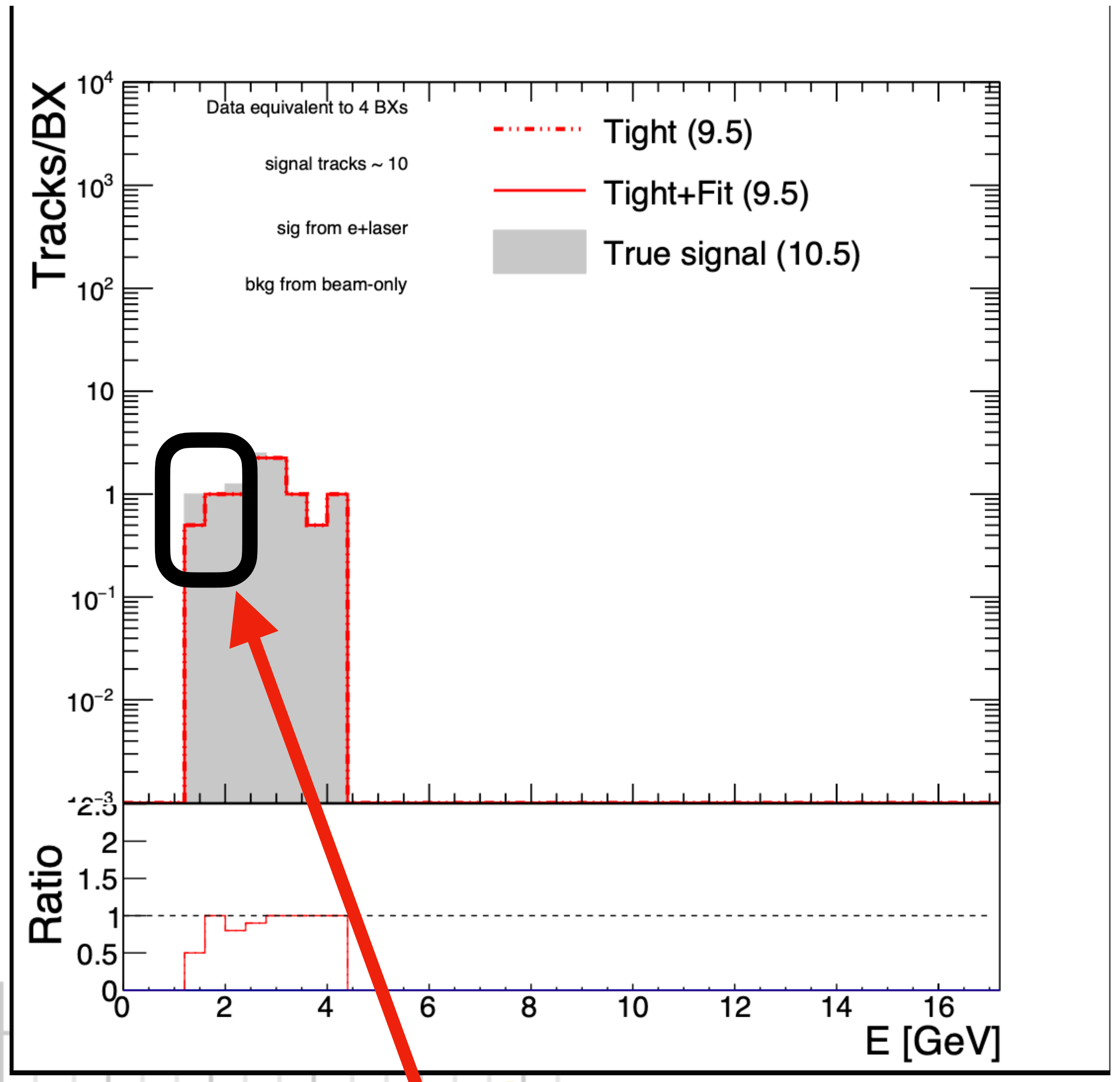
- **Usual Seeding Algorithm cuts (discussed before, in backup)**
  + SVD parameter cut
  Second fit parameter < 0.05 and Third fit parameter < 0.01
  + distance from the expected analytical line in x_{dipole} : x_{layer4} line
  The distance d < 2 mm
  + $|p_Y|$ < 0.005 GeV

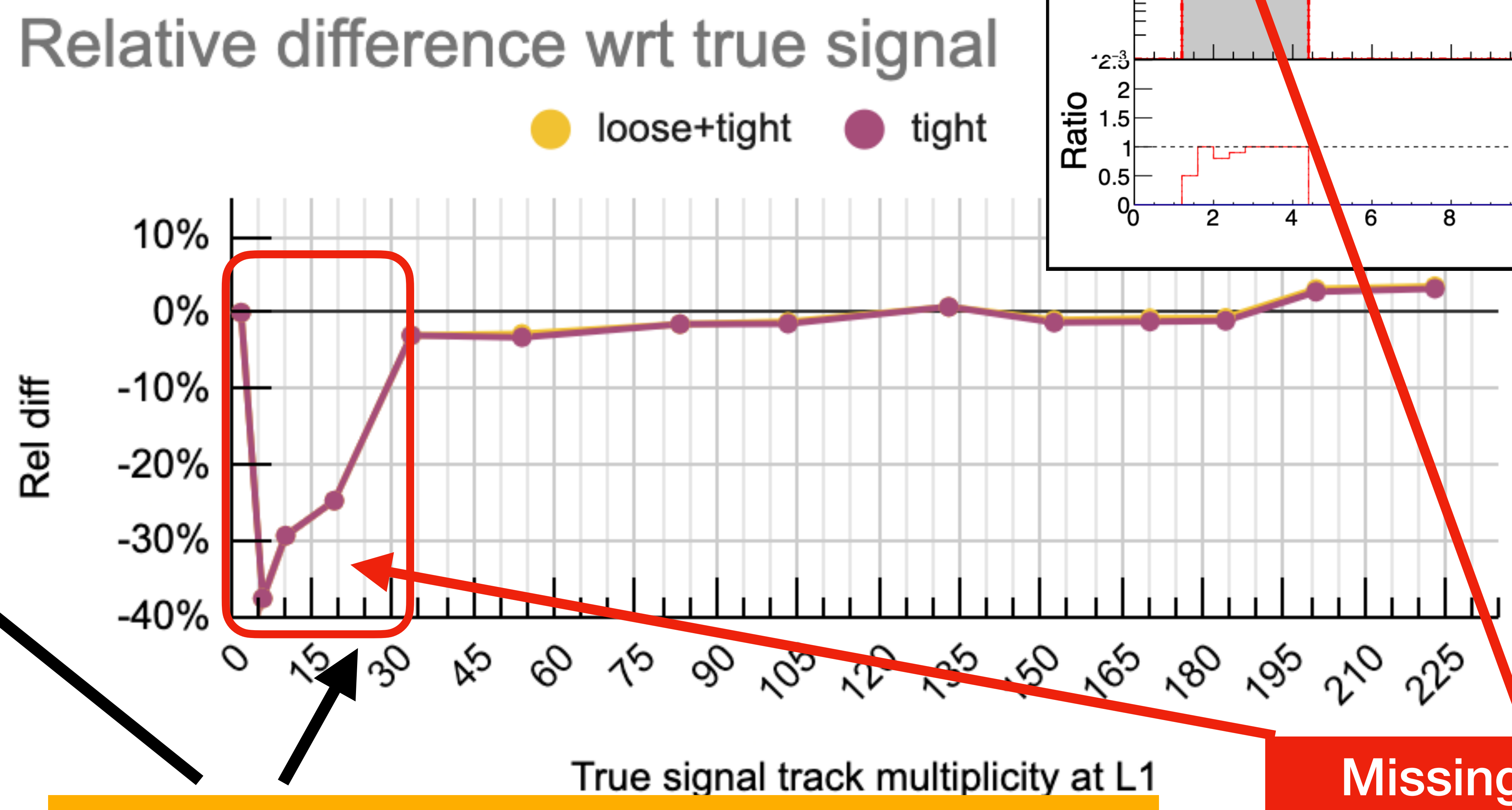
Approximate multiplicity of true signal tracks	Sig, # of Tru	S+B, # of seeds w/o fit	loose+tight S+B	loose+tight Rel diff loose+tight	tight S+B	tight Rel diff tight	loose S+B	loose Rel diff loose
			--- Summary ---					
1	2.0	22.0	2.0	0%	2.0	0.0%	0.0	-100.0%
5	6.0	25.5	3.8	-38%	3.8	-37.5%	0.0	-100.0%
10	10.3	30.0	7.3	-29%	7.3	-29.3%	0.0	-100.0%
20	19.3	42.8	14.5	-25%	14.5	-24.7%	0.0	-100.0%
30	33.5	73.0	32.5	-3%	32.5	-3.0%	0.0	-100.0%
50	54.0	104.5	52.5	-3%	52.3	-3.2%	0.3	-99.5%
80	83.3	165.0	82.0	-2%	82.0	-1.5%	0.0	-100.0%
100	103.3	224.8	102.0	-1%	101.8	-1.5%	0.3	-99.8%
130	133.0	316.0	134.0	1%	134.0	0.8%	0.0	-100.0%
150	152.5	373.3	151.0	-1%	150.5	-1.3%	0.5	-99.7%
170	170.3	466.3	169.0	-1%	168.3	-1.2%	0.8	-99.6%
185	184.3	516.3	183.0	-1%	182.3	-1.1%	0.8	-99.6%
200	201.0	632.5	207.3	3%	206.5	2.7%	0.8	-99.6%
220	223.0	819.3	230.8	3%	230.0	3.1%	0.8	-99.7%

Plot made from FastSim by Noam



★Another interesting distribution is the seed p_Y
★Cutting on $|p_Y|$ < 0.005 GeV can isolate the signals from the background even more.





Now lower track multiplicity is worse!

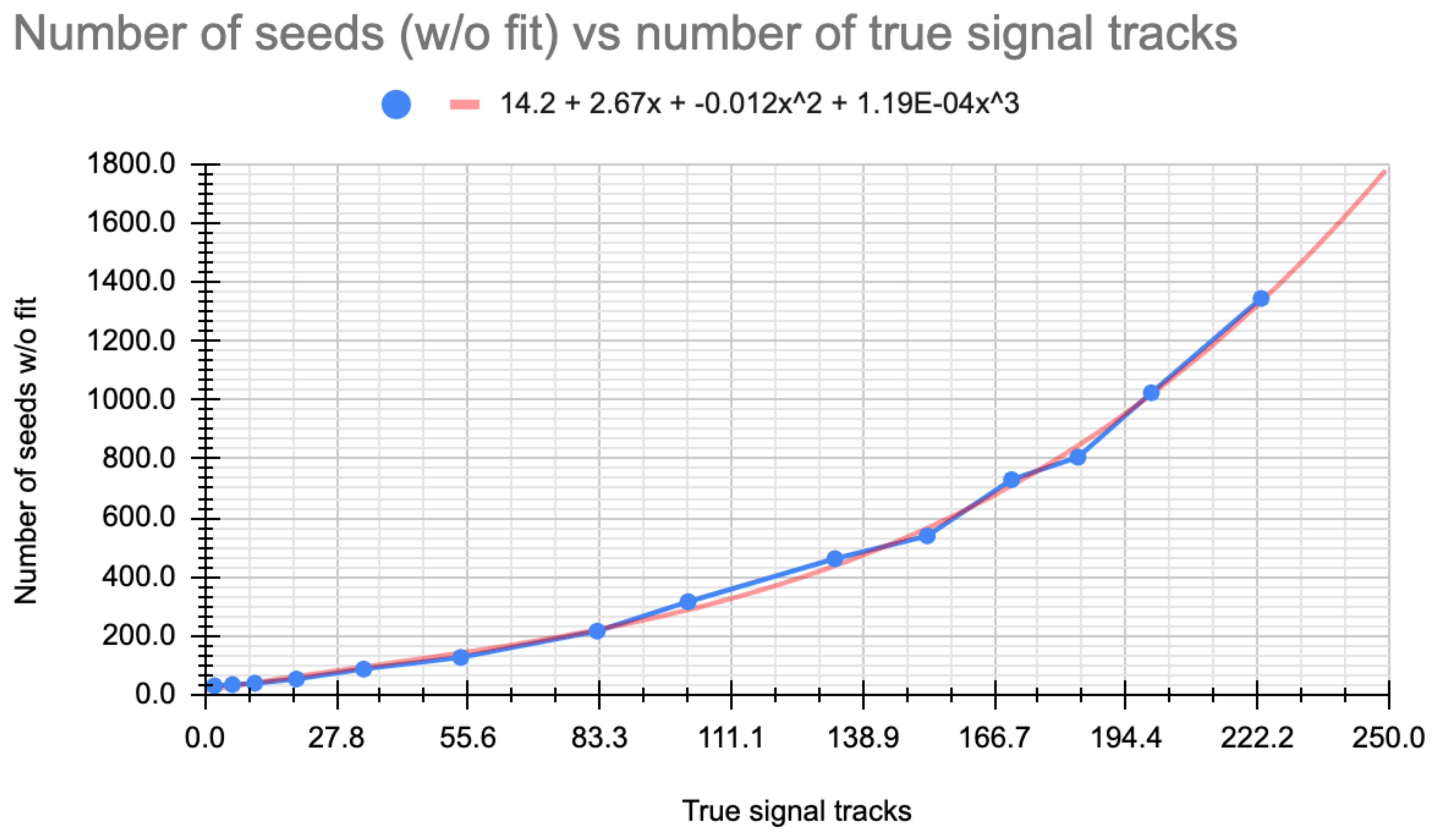
Missing the low end of energy

Way Out: “No One Cut Fits All” (unfortunately)

- ✦Put different cuts on steps!
  - ✦Steps are determined by the crude estimation of the signal+background multiplicity from the tracker
    - ✦Crude because no fit cut done to find out the proper seed multiplicity
  - ✦A polynomial function fits the number of seeds without fit with the true signal tracks.
- ✦High multiplicity case: > 960 possible crude seeds
- ✦Medium multiplicity case: 65 < and <= 960
- ✦Low multiplicity case: < 65

A crude estimate of number of seeds possible from the tracks available in tracker layer 1 and layer 4

(Before applying any fit parameter cut)



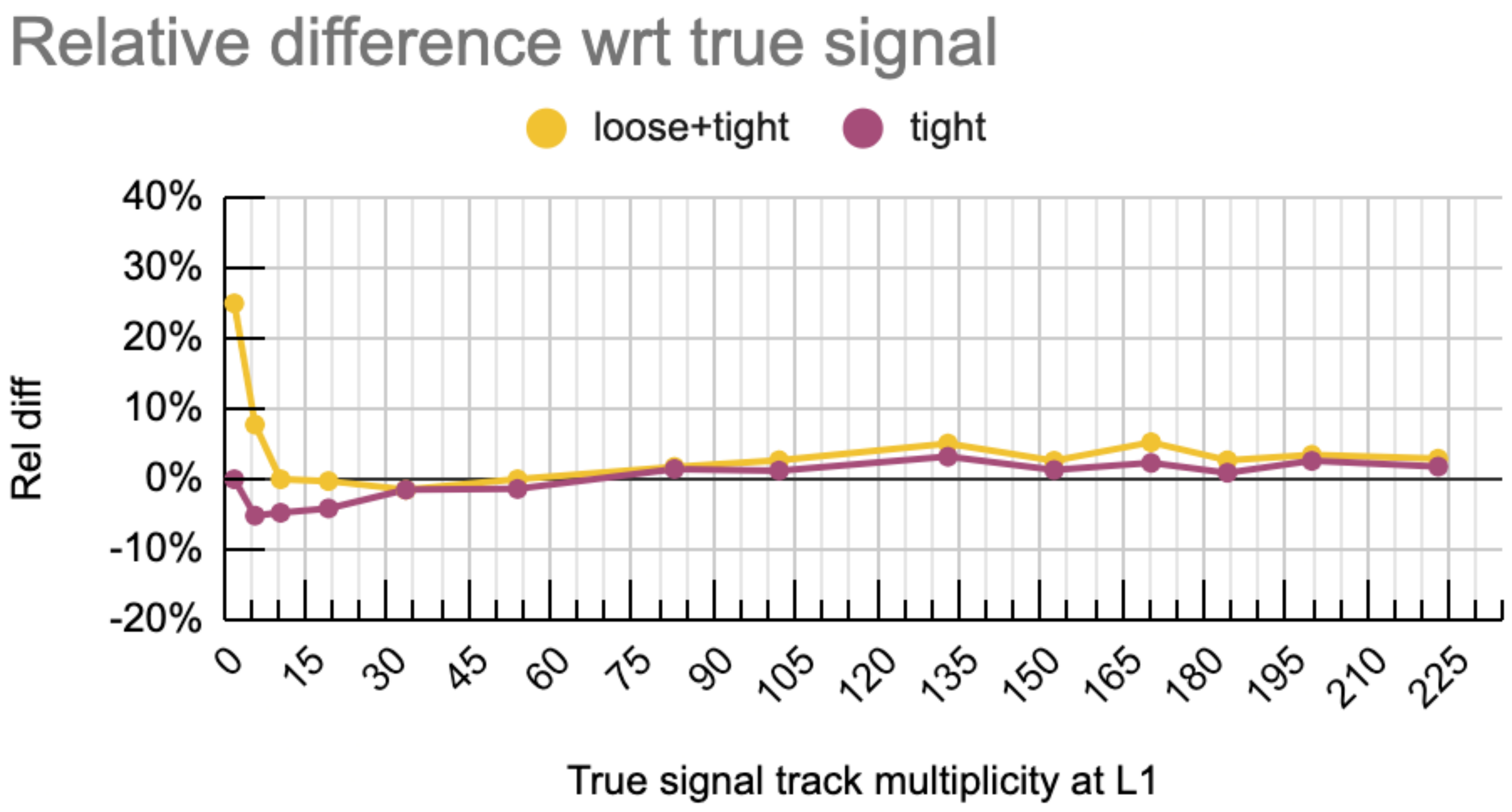
	High	Medium	Low	
	All tracks	All tracks	Tight tracks	Loose tracks
Second SVD Parameter	0.05	0.06	0.1	0.07
Third SVD Parameter	0.01	0.06	0.1	0.05
pY [GeV]	0.005	0.005	0.01	0.0075
distance d [m]	0.003	0.003	0.006	0.004

N of true sig trks	15.0	20.0	30.0	195
N of seeds	52.0	63.8	86.7	960.9

Result with the Hybrid Step Cut

Approximate multiplicity of true signal tracks	--- Summary --- Sig, # of Tru	S+B, # of seeds w/o fit	loose+tight S+B	loose+tight Rel diff loose+tight	tight S+B	tight Rel diff tight	loose S+B	loose Rel diff loose
1	2.0	30.5	2.5	25%	2.0	0.0%	0.5	-75.0%
5	5.8	34.8	6.3	8%	5.5	-5.2%	0.8	-87.1%
10	10.5	39.3	10.5	0%	10.0	-4.8%	0.5	-95.2%
20	19.3	53.5	19.3	0%	18.5	-4.1%	0.8	-96.1%
30	33.5	87.3	33.0	-1%	33.0	-1.5%	0.0	-100.0%
50	54.0	126.8	54.0	0%	53.3	-1.4%	0.8	-98.6%
80	82.8	216.0	84.3	2%	84.0	1.4%	0.3	-99.7%
100	102.0	315.8	104.8	3%	103.3	1.2%	1.5	-98.5%
130	133.0	461.8	139.8	5%	137.3	3.2%	2.5	-98.1%
150	152.5	539.0	156.5	3%	154.5	1.3%	2.0	-98.7%
170	170.3	729.8	179.3	5%	174.3	2.3%	5.0	-97.1%
185	184.3	806.0	189.3	3%	186.0	0.9%	3.3	-98.2%
200	199.8	1024.3	206.8	3%	205.0	2.6%	1.8	-99.1%
220	223.0	1344.3	229.5	3%	227.0	1.8%	2.5	-98.9%

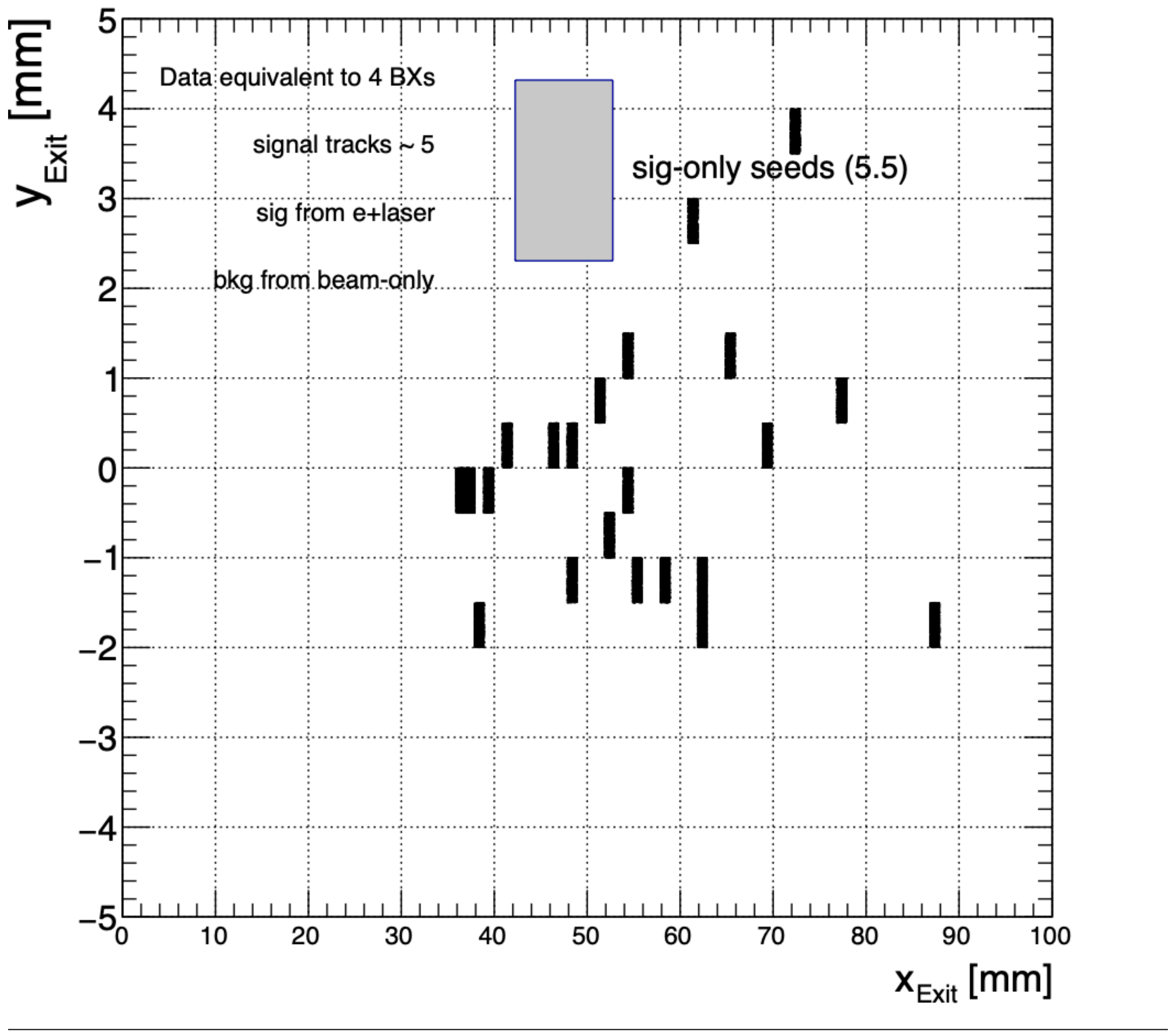
Within 5%

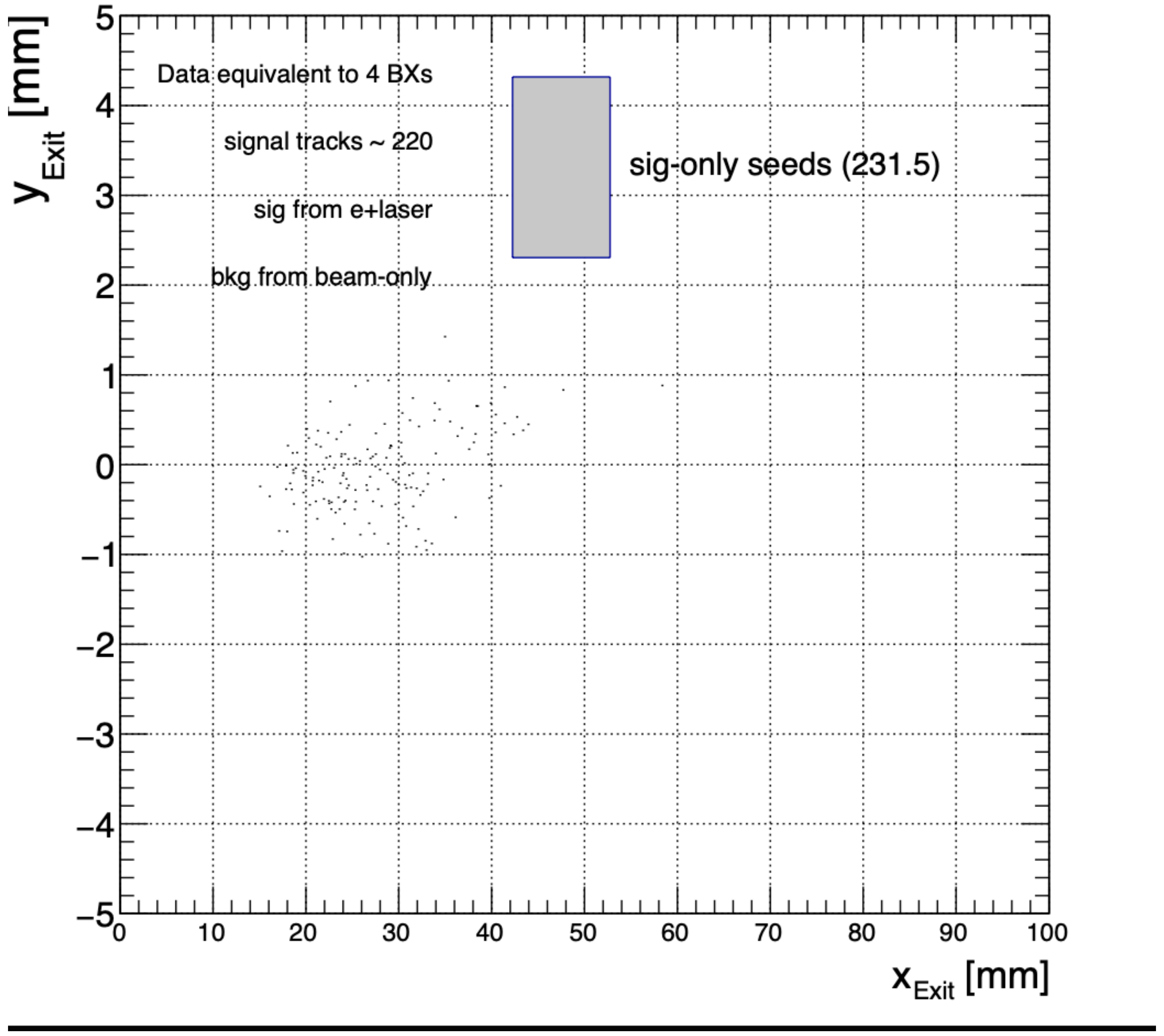


- ✦**Things are better in the low end.**
- ✦**High end can be optimised even more.**
- ✦**The message is clear: playing with different cuts improves the situation.**
- ✦**Sasha's updated samples now can find actual signal tracks on the layers — this will help us identify more suitable cuts to improve the situation even more!**

Summary

- Playing with the seeding algorithm cuts
  - The SVD fit parameter cut and the distance from the $x_{dipole} : x_{layer4}$ really improves the situation.
  - Also the seed momentum in the Y direction
- Some may cost us in efficiency in some ranges but still we cut tight as that can reduce the bkg and the wrong signal combinatorics.
  - Bring us as close as possible to the true-signal tracks
- When we have the proper signal-tagging branch in the trees we will re-derive and optimise the cuts properly.
  - Sasha produced them already, need some time to reprocess.
- Some kinematic distribution plots are attached to the agenda page: Link
- Work with the photon+laser to evaluate the algorithm there as well.

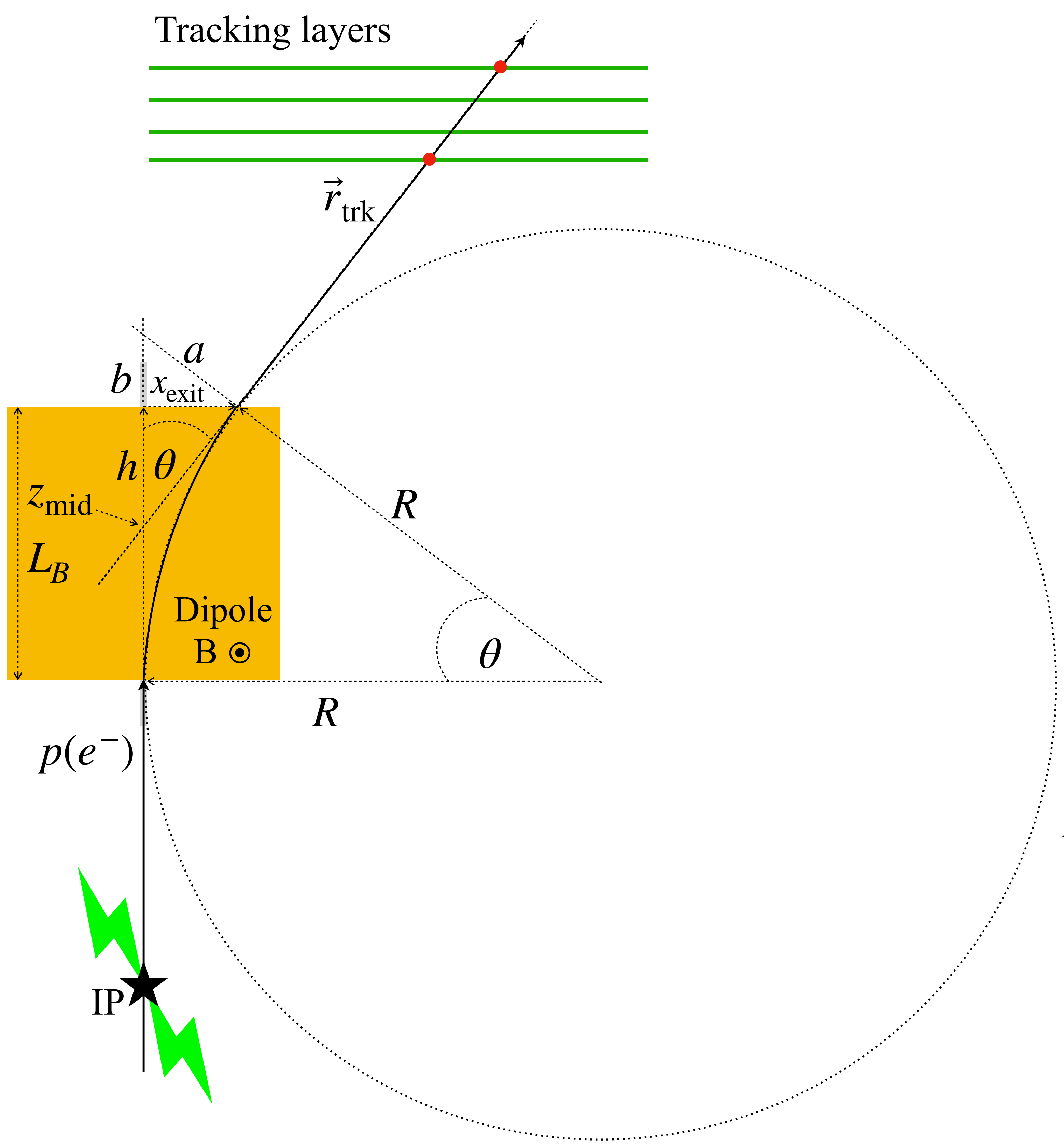




x_Exit and y_Exit at the dipole exit from the best fit line

Backup

Momentum from 2 measurements: Simple Magnet Dipole setup



To get R, need to extrapolate the track backwards to the dipole exit plane and obtain the $x_{\rm exit}$ coordinate then intercept with the $x = 0$ line to find $z_{\rm mid} \rightarrow h$. Extrapolation is done using 2 points at layer 4 and 1

$$\tan\theta = \frac{x_{\rm exit}}{h} = \frac{L_B + b}{R} \longrightarrow R = h\frac{L_B + b}{x_{\rm exit}} \text{ and}$$

$$\tan\theta = \frac{b}{x_{\rm exit}} \longrightarrow R = h\frac{L_B + x_{\rm exit}\tan\theta}{x_{\rm exit}}$$

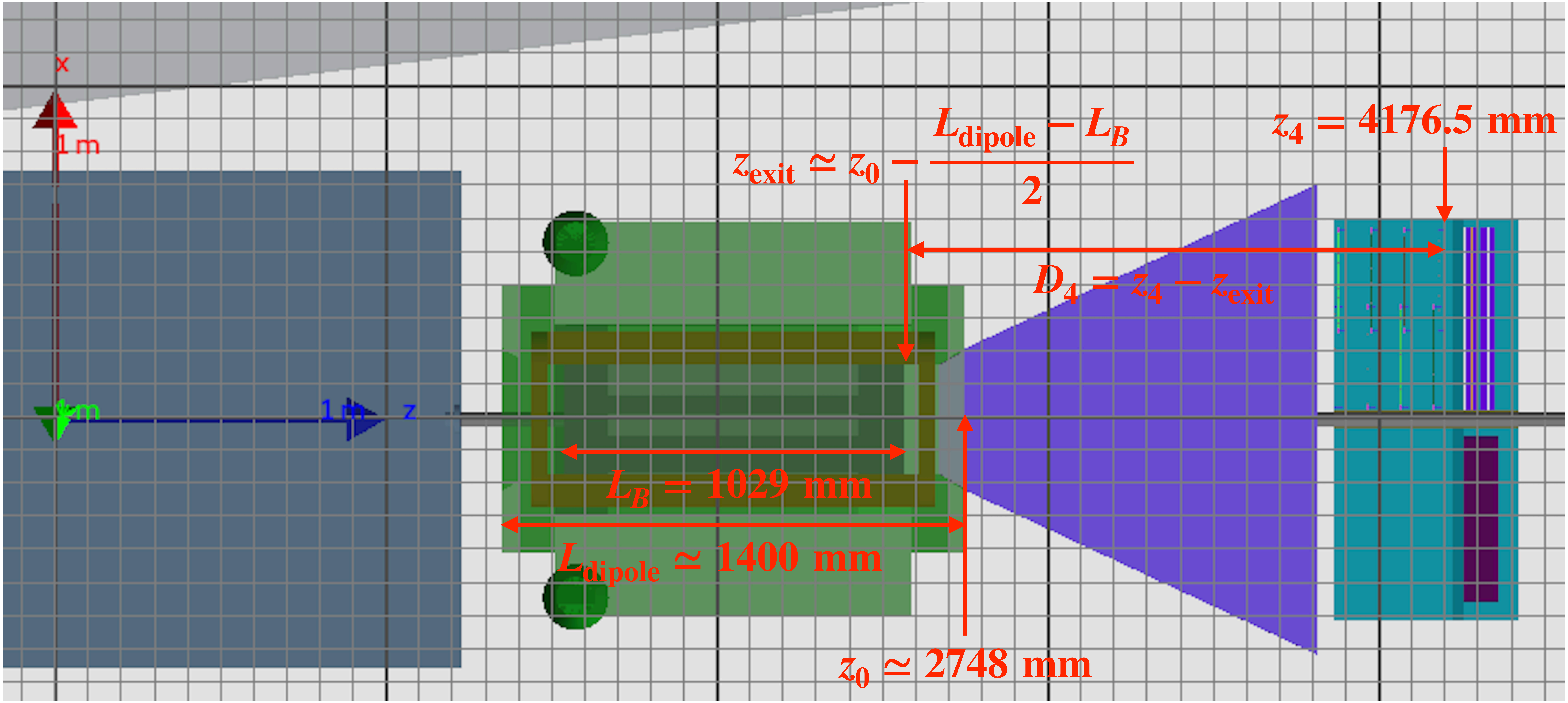
$$R = h\frac{L_B + x_{\rm exit}^2/h}{x_{\rm exit}} = \frac{hL_B}{x_{\rm exit}} + x_{\rm exit}$$

$$p_{\rm seed} = 0.3B\cdot\left(\frac{hL_B}{x_{\rm exit}} + x_{\rm exit}\right) \text{ and } \vec{p}_{\rm seed} = \left(\sim 0,\ p_{\rm seed}\frac{y_{\rm trk}}{r_{\rm trk}},\ p_{\rm seed}\frac{z_{\rm trk}}{r_{\rm trk}}\right)$$

where p_x can be a random number drawn from the truth distribution at the vertex

Numbers

z4: 4176.5 [mm], z0: 2748 [mm], zExit: 2562.5 [mm], D4: 1614 [mm]

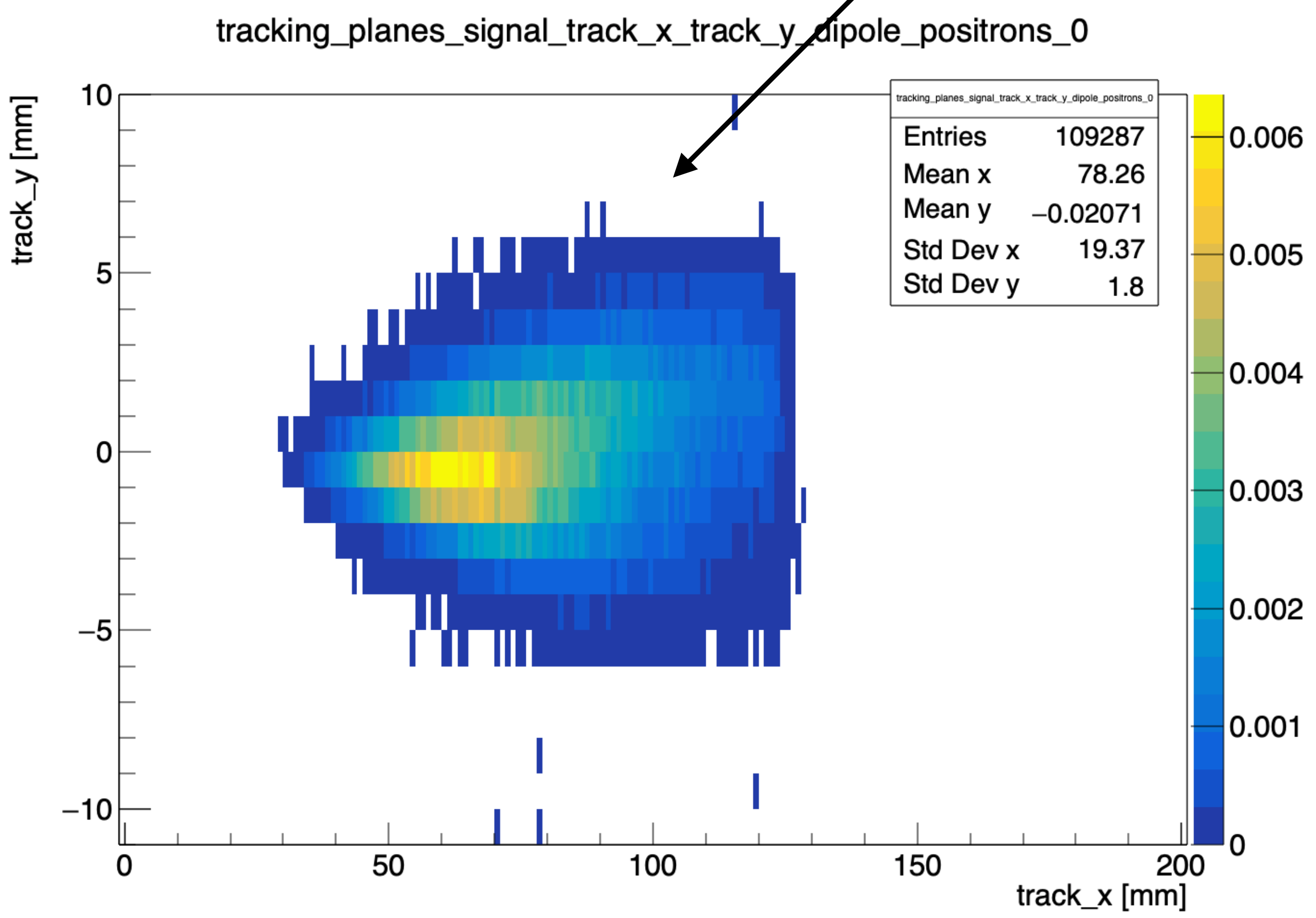


From Noam

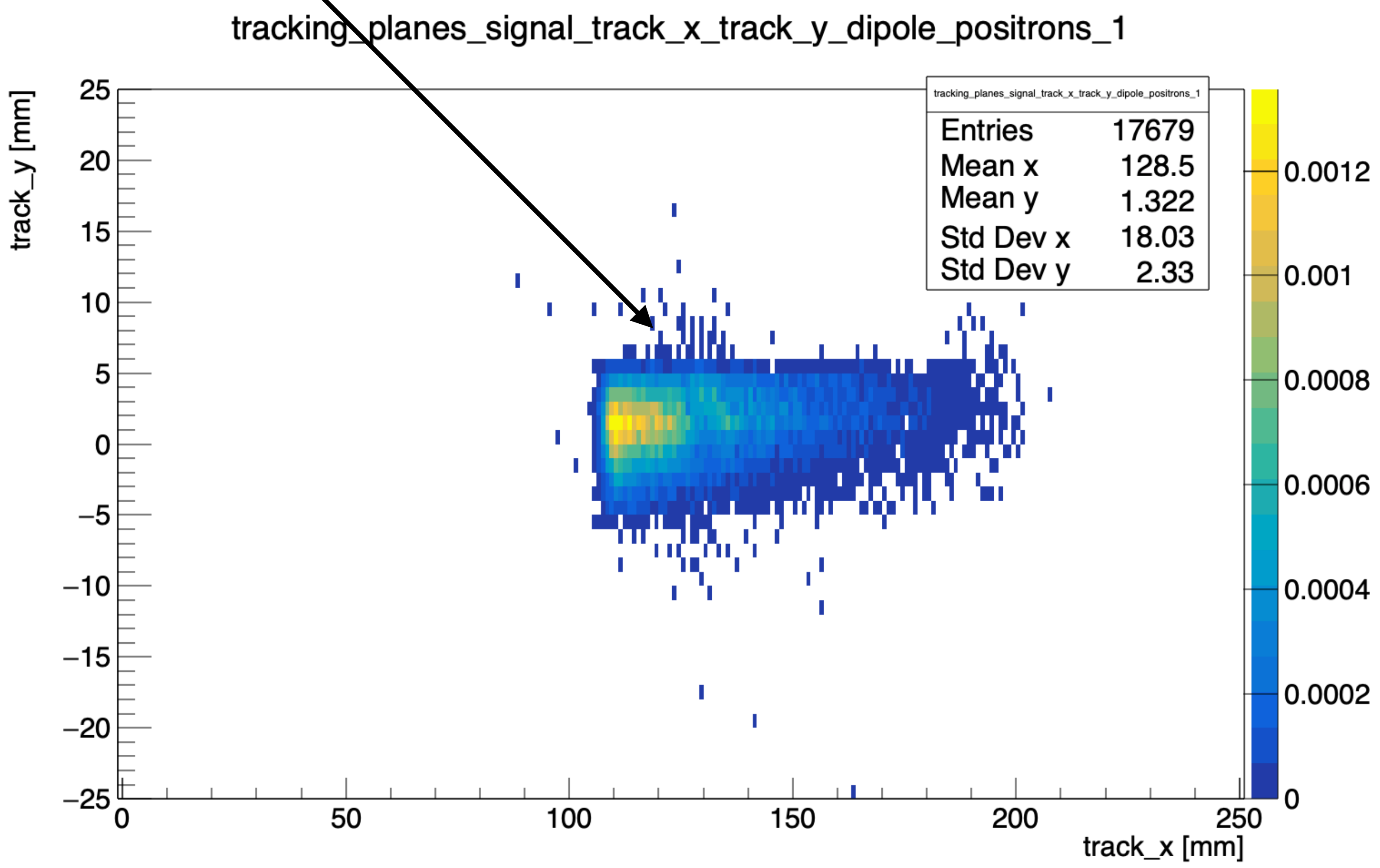
The seeding algorithm

- Keep unique set of tracks from first layer and last layer of tracker.
    - Overlap region is removed (by cutting on x value of the tracks) from
      the outer stave of first layer (innermost) and from the inner stave of
      the last layer (outermost).
- Loop over all pairs in layers 4 and 1, now only positron side (x>0).
- Reject pair of clusters if
    - |x1| > |x4| or they have different sign
    - |z1| == |z4|
    - |y_exit| > 5.4 mm
    - |x_exit| < 20 mm and |x_exit| > 165 mm
    - If not one cluster in the road of 130 um connecting vector r1 and r4 in both layer 2 and layer 3.
    - The seed energy is greater than 17.5 GeV or less than 0.5 GeV.
        - The seed energy is calculated from the track.
    - Apply the SVD fit parameter cut and distance cut

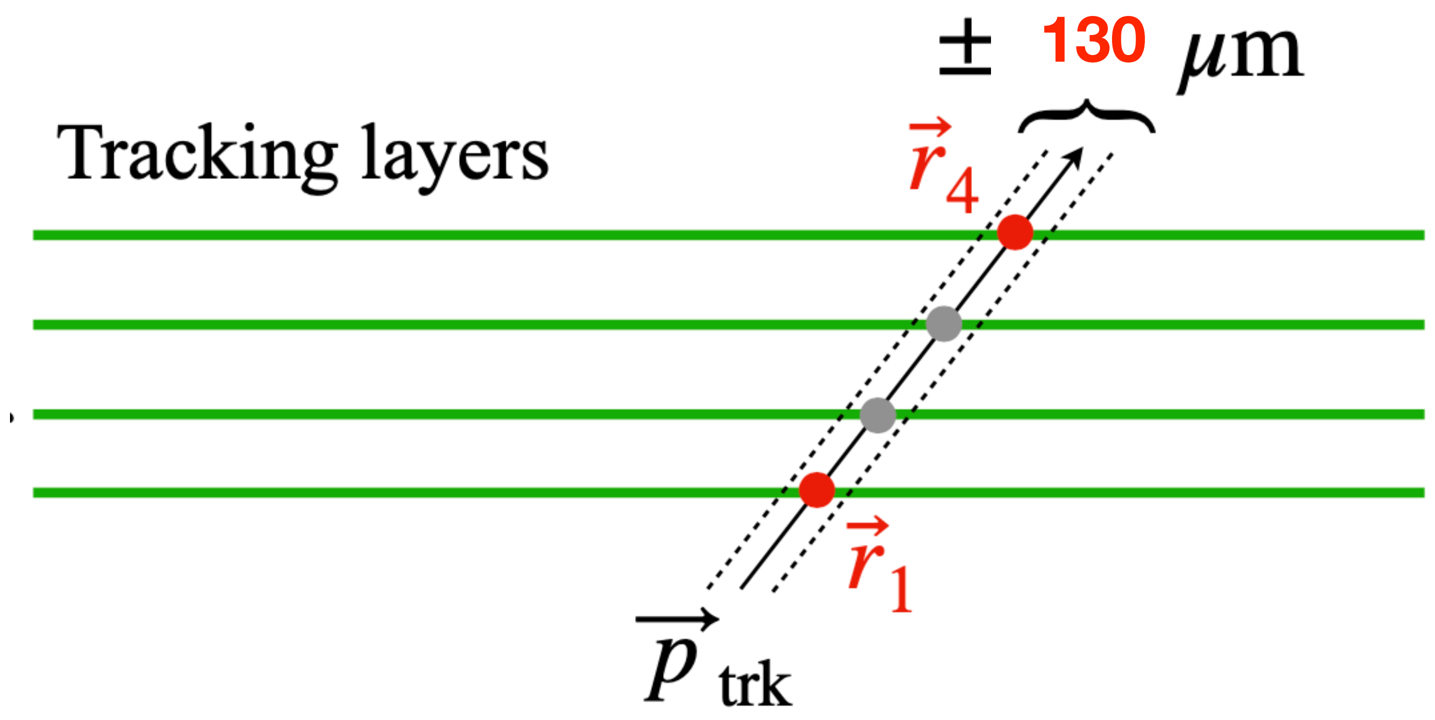
P side Signal at the Dipole exit (New signal)



Stave 0



Stave 1



SVD track fitting

- A common application of the singular value decomposition (SVD) is in fitting solutions to linear equations.
- Suppose we collect (x_i, y_i) data, which can be fit to some linear homogeneous equation $ax_i + by_i - c = 0$. If we have our data in a matrix A, and the coefficients in some vector v, we can write this problem as $Av = 0$, where we'd like to figure out what v is given A
- We usually have more data than coefficients, in which case we can't solve this
- However, we can fit v to A in order to minimise the value of Av via SVD
- Suppose we take the singular value decomposition of $A \rightarrow A = U\Sigma V^{\mathrm{T}}$
- If one of the singular values is 0 then we have an exact solution
- If none of the singular values are zero, we have no exact solution
- However, it can be shown that the smallest singular value corresponds to the solution of the linear least squares fitting problem
- Namely, if we want to find the least squares fit to the data, we need to look at the smallest (usually the last) singular value and read the respective column of V to get the best fit coefficients (a, b, c)

From Noam