

Table of Contents

Search for new physics with disappearing tracks.....	1
Presentations.....	1
ARC review.....	1
ARC action items from July 25 meeting.....	1
On-going 2018 estimate updates (last updated July 9).....	2
Questions from ARC.....	8
Email questions from Kevin Stenson August 27.....	8
Comments from Giacomo Sguazzoni HN August 13.....	9
Questions from Joe Pastika HN August 7.....	11
"Some followup questions" from Kevin Stenson HN July 24.....	13
First set of comments from Kevin Stenson HN July 14.....	15
Questions from Juan Alcaraz (July 5).....	26
Pre-approval.....	27
Additional pre-approval followup Ivan Mikulec HN June 21.....	27
Questions from pre-approval EXO June 1.....	28
Comments on the paper.....	29
Juliette Alimena (v0) May 22.....	29
Object reviews.....	30
Anshul Kapoor (Electron object review).....	32
Benjamin Radburn-Smith (Muon object review).....	32
Klaas Padeken (Tau object review).....	33
Chad Freer (MET object review).....	33
Eirini Tziaferi (Jet object review).....	34
Questions from subgroup conveners.....	34
Juliette Alimena email comments on AN v2, 19 Feb 2019.....	34
Steven Lowette email comments on AN v2, 23 Feb 2019.....	35

Search for new physics with disappearing tracks

- CADI line: EXO-19-010 [↗](#)
- Analysis note: AN2018_311_v7.pdf iCMS [↗](#)
- Hypernews forum: link [↗](#)
- Analysis contacts: Brian Francis, Yuri Gershtein, Andrew Hart, Chris Hill, Kai Wei

Presentations

Only notable roundtable presentations are listed here.

Occasion	Content	Links
19 Oct 2018 LL EXO WG	Initial lepton background estimates for 2017, optimizing search region for Phase I pixel upgrade	Agenda slides ↗
30 Nov 2018 roundtable	Settled on signal region binning	Agenda slides ↗
14 Dec 2018 LL EXO WG	Initial fake background estimates for 2017, initial limits	Agenda slides ↗
18 Jan 2019 LL EXO WG	Background closure, updates	Agenda slides ↗
1 Mar 2019 LL EXO WG	Fake background method update, signal corrections/systematics	Agenda slides ↗
22 Mar 2019 roundtable	Fake background method update, clarification of technical needs	Agenda slides ↗
29 Mar 2019 LL EXO WG	Full status update, extension of signal to 1000, 1100 GeV charginos	Agenda slides ↗
21 May 2019 EXO general	Extrapolation from 2017 to high pileup (postponed)	Agenda slides ↗
24 May 2019 LL EXO WG	First 2018 plots	Agenda slides ↗
31 May 2019 LL EXO WG	Pre-approval	Agenda slides ↗
7 June 2019 LL EXO WG	Followup to pre-approval	Agenda slides ↗
19 July 2019 LL EXO WG	2018 ABC background estimates	Agenda slides ↗

Expand all Collapse all

NOTE: Questions are in **Red** (Unanswered), or **Green** (Answered), or **Purple** (In Progress) while answers are in **Blue** .

ARC review

ARC action items from July 25 meeting

Show Details Hide Details

Combine the nine dxy sideband regions in the fake estimate into one larger sideband.

Done. The fake estimates change slightly but are all within 0.5sigma (statistical) of the previous estimates. The AN is updated.

Compare the pileup distributions in ZtoMuMu, ZtoEE, and BasicSelection events. If there is a big difference, try reweighting and see how much it changes the estimate.

nPV	ratios to BasicSelection
-----	--------------------------

See the above plots. Using the ratios as weights to the fake selection (ZtoMuMu/EE + DisTrk (no d0 cut)), the overall weights applied to P_fake^raw would be:

	ZtoMuMu	ZtoEE
nLayers = 4	0.994 +- 0.064	1.01 +- 0.34
nLayers = 5	1.013 +- 0.088	1.0 +- 1.4
nLayers >= 6	1.0 +- 0.21	1.02 +- 0.83

Despite the plots above these average weights are very consistent with one, e.g. the estimate does not depend on this.

If possible, find the justification for using dxy with respect to the origin for the track isolation pileup subtraction.

We could not find a justification and have concluded that regardless of what inspired it, calculating the track isolation with respect to the origin was/is a mistake. However, as noted in the ARC meeting, this mistake is confined to the selection of tracks to be included in track isolation sum. The effect of which is to reduce the efficacy of the track isolation requirement. Redoing this sum would require prohibitively large reprocessing but fortunately, due to the redundancy of this cut with the calorimeter isolation requirement for charged hadrons and electrons, and the muon delta R requirement for muons, the effect on the analysis is not significant.

Suggestion: compare the nominal Gaussian fit to a flat line for NLayers5, as there's a concern the bias towards the PV changes as nLayers increases.

With a flat line the transfer factor is purely a normalization issue and has no uncertainty; it is always $0.02 / 0.45 = 0.0444$. The table below is added to the AN:

□

ZtoMuMu NLayers5 (gaussian fit from NLayers4)	ZtoMuMu NLayers5 (pol0 fit, finer binning)
---	--

The flat assumption reduces the estimates by ~2/3 and the agreement is worse, especially as there would be no 40-50% fit uncertainty.

On-going 2018 estimate updates (last updated July 9)

Show Details Hide Details

( - todo  - doing  - done)  Reset

- Click me to change my state from 'todo' to 'doing'
t: 2019-11-17 - 11:07 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'todo'
t: 2019-11-17 - 11:07 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'doing'
t: 2019-11-17 - 11:07 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'done'
t: 2019-11-17 - 11:07 - Main.TWikiGuest, last state: tobedone
[X]   
-  Produce skimmed ntuples with CRAB
 - ◆ Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 MET

- ◆ Click me to change my state from 'doing' to 'done'
t: 2019-12-23 - 07:29 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'todo'
t: 2019-12-23 - 07:29 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'doing'
t: 2019-12-23 - 07:29 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'done'
t: 2019-12-23 - 07:29 - Main.TWikiGuest, last state: tobedone
[X]   

 EGamma

- ◆ Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 SingleMuon

- ◆ Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 Tau

- Click me to change my state from 'done' to 'todo'
t: 2019-11-10 - 23:59 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'todo'
t: 2019-11-10 - 23:59 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'doing'
t: 2019-11-10 - 23:59 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'done'
t: 2019-11-10 - 23:59 - Main.TWikiGuest, last state: tobedone
[X]   

 Create fiducial maps *

- Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 Muon *

Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 Electron (D in progress)

- Click me to change my state from 'todo' to 'doing'
t: 2019-12-23 - 09:14 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'todo'
t: 2019-12-23 - 09:14 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'doing'
t: 2019-12-23 - 09:14 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'todo' to 'done'
t: 2019-12-23 - 09:14 - Main.TWikiGuest, last state: tobedone
[X]   

 Run channels without fiducial track selections

- ◆ Click me to change my state from 'done' to 'todo'
t: 2019-11-11 - 05:19 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'todo'
t: 2019-11-11 - 05:19 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'doing'
t: 2019-11-11 - 05:19 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'done'
t: 2019-11-11 - 05:19 - Main.TWikiGuest, last state: tobedone
[X]   

 basicSelection

- ◆ Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   

 ZtoEE (D in progress)

- ◆ Click me to change my state from 'done' to 'todo'
t: 2019-11-17 - 12:41 - Main.TWikiGuest, last state: done
Click me to change my state from 'done' to 'todo'
t: 2019-11-17 - 12:41 - Main.TWikiGuest, last state: done
Click me to change my state from 'done' to 'doing'
t: 2019-11-17 - 12:41 - Main.TWikiGuest, last state: done
Click me to change my state from 'done' to 'done'
t: 2019-11-17 - 12:41 - Main.TWikiGuest, last state: done
[X]   

 ZtoMuMu

- ◆ Click me to change my state from 'done' to 'todo'
t: 2019-11-10 - 21:50 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'todo'
t: 2019-11-10 - 21:50 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'doing'
t: 2019-11-10 - 21:50 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'done'
t: 2019-11-10 - 21:50 - Main.TWikiGuest, last state: tobedone
[X]   
- Trigger efficiencies with muons
- Click me to change my state from 'done' to 'todo'
t: 2019-11-17 - 08:49 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'todo'
t: 2019-11-17 - 08:49 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'doing'
t: 2019-11-17 - 08:49 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'done' to 'done'
t: 2019-11-17 - 08:49 - Main.TWikiGuest, last state: tobedone
[X]   
- Background estimates and systematics (ABC complete) *
Click me to change my state from 'doing' to 'done'
t: 2019-12-22 - 03:59 - Main.TWikiGuest, last state: todo
Click me to change my state from 'doing' to 'todo'
t: 2019-12-22 - 03:59 - Main.TWikiGuest, last state: todo
Click me to change my state from 'doing' to 'doing'
t: 2019-12-22 - 03:59 - Main.TWikiGuest, last state: todo
Click me to change my state from 'doing' to 'done'
t: 2019-12-22 - 03:59 - Main.TWikiGuest, last state: todo
[X]   
- Electron *
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'todo'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'doing'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'tobedone' to 'done'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
[X]   
- Muon *
Click me to change my state from 'doing' to 'done'
t: 2019-11-11 - 09:21 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'todo'
t: 2019-11-11 - 09:21 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'doing'
t: 2019-11-11 - 09:21 - Main.TWikiGuest, last state: tobedone
Click me to change my state from 'doing' to 'done'
t: 2019-11-11 - 09:21 - Main.TWikiGuest, last state: tobedone
[X]   
- Tau *
Click me to change my state from 'todo' to 'doing'
t: 2019-11-17 - 14:29 - Main.TWikiGuest, last state: doing
Click me to change my state from 'todo' to 'todo'
t: 2019-11-17 - 14:29 - Main.TWikiGuest, last state: doing

Click me to change my state from 'todo' to 'doing'

t: 2019-11-17 - 14:29 - Main.TWikiGuest, last state: doing

Click me to change my state from 'todo' to 'done'

t: 2019-11-17 - 14:29 - Main.TWikiGuest, last state: doing

[X]   

 Fake *

Click me to change my state from 'todo' to 'doing'

t: 2019-11-17 - 09:55 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'todo'

t: 2019-11-17 - 09:55 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'doing'

t: 2019-11-17 - 09:55 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'done'

t: 2019-11-17 - 09:55 - Main.TWikiGuest, last state: tobedone

[X]   

 ZtoMuMu *

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'doing'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'done'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 ZtoEE *

Click me to change my state from 'todo' to 'doing'

t: 2019-11-11 - 08:21 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'todo'

t: 2019-11-11 - 08:21 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'doing'

t: 2019-11-11 - 08:21 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'done'

t: 2019-11-11 - 08:21 - Main.TWikiGuest, last state: tobedone

[X]   

 Fetch RAW and re-reco lepton P(veto) passing events

• Click me to change my state from 'done' to 'todo'

t: 2019-11-11 - 07:20 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'todo'

t: 2019-11-11 - 07:20 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'doing'

t: 2019-11-11 - 07:20 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'done'

t: 2019-11-11 - 07:20 - Main.TWikiGuest, last state: tobedone

[X]   

 Signal corrections *

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'doing'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'done'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 Pileup *

Click me to change my state from 'todo' to 'doing'

t: 2019-12-23 - 05:59 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'todo'

t: 2019-12-23 - 05:59 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'doing'

t: 2019-12-23 - 05:59 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'todo' to 'done'

t: 2019-12-23 - 05:59 - Main.TWikiGuest, last state: tobedone

[X]   

 ISR weights *

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'doing'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'done'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 missing middle/outer hits (requires fiducial maps) *

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'doing'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'done'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 Trigger scale factors

- Click me to change my state from 'doing' to 'done'

t: 2019-11-11 - 03:14 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'doing' to 'todo'

t: 2019-11-11 - 03:14 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'doing' to 'doing'

t: 2019-11-11 - 03:14 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'doing' to 'done'

t: 2019-11-11 - 03:14 - Main.TWikiGuest, last state: tobedone

[X]   

 Signal Systematics

- Click me to change my state from 'done' to 'todo'

t: 2019-11-10 - 20:48 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'todo'

t: 2019-11-10 - 20:48 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'doing'

t: 2019-11-10 - 20:48 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'done' to 'done'

t: 2019-11-10 - 20:48 - Main.TWikiGuest, last state: tobedone

[X]   

 Expected upper limits

- Click me to change my state from 'tobedone' to 'todo'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'todo'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'doing'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'done'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 [X]   
 Unblind observation

Questions from ARC

Email questions from Kevin Stenson August 27

Show Details Hide Details

AN Table 22 question. I do not understand the answer. It seems like you are agreeing with me. Maybe I don't understand Table 22. Let's just take the muon case for now. You want to measure the probability that a muon passes the lepton veto. You basically just measure the fraction of probe tracks that pass the lepton veto. According to Table 22, you measure the fraction of probe tracks that pass the criteria $\text{minDeltaR}_{\text{track,muon}} > 0.15$ and missing outer hits > 2 . This is P_{veto} . However, the actual veto that you apply in the signal region is the 5 requirements in Tables 20-21. So, why don't you measure P_{veto} for muons by measuring the fraction of probe tracks that pass the criteria of all 5 requirements in Tables 20-21.

We require a pure sample of each flavor, and using all five requirements would reduce that purity. The listed requirements are the most powerful at rejecting the given flavor, so they are the correct choices to study.

Figure 19 and Tables 29-31 questions. If Figure 19 and Tables 29-31 both utilize all probe tracks then I would expect them to give the same result. So I'm not sure how that affects anything. In principle, the same sign subtraction could have an effect. Hopefully once the tables I asked for are made I will be able to understand if that is the reason.

These are "all probe tracks", there has been no $Z \rightarrow T\&P$ requirement yet. This is a technical necessity in creating a pool of tags and probes in which to find all possible pairs. The table you request has been made for the coming AN update, and for this channel:

$$P(\text{veto}) := (35 - 2) / (1213660 - 1437) = 2.72e-5$$

Section 6.1.5 question. I'm not sure this makes sense. You say that you only use $t\bar{b}$ because that contributes the most. But the analysis includes a tag-and-probe assuming Z production. So, while $Z \rightarrow ll$ may not contribute to the overall background, it will certainly contribute a great deal to the measurement of P_{veto} . I think you should use all of the MC samples you have if you are trying to imitate the data.

The closure test doesn't necessarily need to mimic the data, but should demonstrate that the method closes within a given sample. The other samples just did not have the statistics available to provide a reasonable observation and estimate.

Section 8.2.2 question. You write that the chargino behaves like a muon and so muons should be used as proxy for measuring the efficiency. It is true that the chargino does not undergo hadronic interactions so it is more like a muon than a hadron. However, the chargino of interest for this analysis does NOT get reconstructed as a muon in the muon system. So it is unlike a muon in that sense. The track reconstruction explicitly uses information from the muon detectors to find all possible muon tracks and to associated as many hits as possible with the track. This is possible for muons but NOT for charginos that decay before the muon system. Therefore, using muons may overestimate the hit efficiency. If you found the same result for hadrons, then I would not be concerned. Or, if you just used muon tracks that were not found by the special muon-finding steps, then I would be happy. I know you have already used this but for reference, this shows

the overall tracking efficiency for muons with and without the muon-specific steps:
https://cds.cern.ch/record/2666648/files/DP2019_004.pdf As you can see there is a significant reduction in efficiency when the muon-specific steps are removed and a much larger disagreement between data and MC. I have no idea how this translates (or not) to the hit finding efficiency.

We do include the data/simulation difference in DP2019_004 as a systematic uncertainty however, so the overall track reconstruction efficiency for charginos is well-covered. Comparing the 0.02% hit systematic to DP2019_004's 2.1%, the hit systematic is negligible even if the extra iterations increase the hit-association efficiency several-fold.

Section 8.2.5 question. This answer misses the point of my comment. Your estimate of the systematic uncertainty is simply a measure of MC statistics. It does not address the question of the systematic uncertainty associated with the method itself.

As discussed in first ARC meeting, this is a very common procedure used by many analyses. There is an observed difference between simulation and data in the hadronic recoil of di-muon events which must be accounted for, and the only uncertainty to apply is the statistics of the observed difference.

Finally, I see in the answer for the ARC meeting that the calculation of d_{xy} for the track isolation may not have been done in the intended fashion. Can you specify what information you have on what was done. Is there any way to see the d_{xy} distribution as calculated? Is there any way at this point to change to use d_{xy} relative to the primary vertex (like $d_{xy} < 0.02\text{cm}$ as is done for the signal tracks)? I see that this cut has a pretty big effect for the short lifetime case as shown in Figure 16.

See the above answer from the ARC action items.

Comments from Giacomo Sguazzoni HN August 13

Show Details Hide Details

Comments are for the paper (v0 2019/05/16) with references to the AN (v7) when applicable.

L 15-16: "Decaying to a weakly-interacting, stable neutralino and 16 unreconstructed pion, a chargino decay often leaves a disappearing track in the AMSB model." --> it seems the disappearing track is left in the model!

Reworded slightly.

l 25: exclude --> *excludes*

Fixed.

l 31-32: the interpretation ... are --> the interpretation ... *is*

Fixed.

The CMS detector: As already noted, a more detailed description of the tracker is needed. Here the geometry of the tracker is not described at all, while this is important to clarify the concept of 'measurement layer' and their numbers. I'm wondering if a picture of the tracker layout would be appropriate in this paper.

A sentence has been added to specify the position of each layer of the Phase 1 upgrade. L 186-190 should clarify "nLayers". We likely have room for another figure, but we can discuss what would be best to include.

l 75: blank missing between χ^0_1 and 'mass'

Fixed.

l 87: these correction factors are huge (350%); with these correction factor involved, how can you trust the simulation? Moreover the description in the paper is misleading, I think. You give the impression that the 350% factor derives from the ISR mismodelling in the Z->mumu events. But, reading the AN section 7.4, I understand that the big factor derives from AN Fig. 38, i.e. Madgraph vs. Pythia for Z->mumu, that you need since your signal is simulated with Pythia. I think this point has to be better explained in the paper, cause the reader could be surprised to learn that we are 350% off in Z->mumu simulation. See below also the discussion on the systematic uncertainty associated to this correction.

A brief comment has been added, saying that most of this value is due to the ISR modeling in Pythia.

l 112: I think PF as an acronym of particle flow has never been defined

Particle flow is first mentioned on L 95 and the "(PF)" is now defined there.

l 133-155: concept is clear but the description is complicated and I think there is room for improvement; in case they are missing, 'hits' are counted as 'layers' (what about using a different nomenclature? e.g. 'layer measurement'); but what about not missing hits, the one on which you apply the cut? what about the 4 hits you require? This is relevant with respect to overlaps. Clarify (you may need to introduce a discussion on overlaps here and in the tracker description).

Until L 186-190, the actual quantity of nLayers isn't used so the distinction wasn't made clear. Before that point we make frequent reference to "layers", but as physical layers of the detector itself in which hits can exist.

l 184-185: track coordinates? Do you mean track parameters?

That is a better phrasing and we use that now.

l 253-255: The phrasing could be improved.

This section has been rewritten.

l 262-271 the d0 fit description could be improved (need to check the AN to better understand what's going on). In particular:

l 263: when you say 'we first fit the observed d0 of selected tracks to a Gaussian distribution in the range $0.1\text{cm} < |d_0| < 1.0\text{cm}$ ', you mean that you fit d0 excluding, from the fit, the range $-0.1\text{cm} < d_0 < 0.1\text{cm}$. Is that correct? If this is the case, I think the way to describe the fit is misleading.

This section requires some rewriting as in the review we've moved to a single sideband instead of the nine. This rewording makes the fit range and process more clear.

l 265: $|d_0| \implies |d_0|$

Fixed.

l 272: 'independent': is that true? the transfer factor derive from the same fit, i.e. the same gaussian. How could $N_{est}^{i, fake}$ be independent?

It would be more appropriate to say that $P_{fake}^{raw, i}$ is independent, correct. However with the use of one larger sideband this sentence is no longer relevant.

l 326: among the systematic uncertainties, the one associated to the ISR modelling is the largest; nevertheless is it sufficient? You just consider 1sigma of statistical fluctuation (according to the AN, Section 8.2.5), a

quantity that, in principle, you can reduce by increasing the statistics. Is there no systematic associated to the reweighing method itself? I think it is needed given the large correction factors you end up with (350%). A possibility (indeed extreme but for sure conservative) would be to evaluate the efficiency change with and without correction factors. Which would be the systematic in this case?

Comparing with/without the correction would be extremely large (~70% for some example points) and not relevant since differences between Pythia and Madgraph are well known. Removing the Madgraph/Pythia correction, we do not have a comparison of data/Pythia to make the correction -- an entirely Pythia-based SM background campaign would be required. In the limit of infinite statistics in data/Madgraph/Pythia samples, we would essentially have a perfect tune for Z ISR and there would be no uncertainty. Another possibility we are pursuing is to generate a sample of our signal in Madgraph, the expectation being an ISR distribution equal to the 10M DY+jets we generated to form the Madgraph/Pythia weights.

Questions from Joe Pastika HN August 7 [↗](#)

Show Details [▢](#) Hide Details [▢](#)

L254: "bad charged hadron filter" is listed as "not recommended" on the JetMET twiki. Is there a reason this is still included in your filter list?

The recommendation was changed between our analysis of 2017 and 2018 data, and it wasn't feasible to re-process 2017 for this; all of the MET filters listed remove only 2% of the /MET/ dataset, so it is a small issue. In the next AN version "(2017 only)" is added here.

L268: Could you use a difference symbol in the text/tables for when you use dx/dz referenced from 0,0,0 (maybe dz_{000} or something else reasonable) to differentiate it clearly from measurement w.r.t. the beamspot?

Now used is $d_{z}^{\{0\}}$ and $d_{xy}^{\{0\}}$ where relevant.

L311: What effect does the choice of "2 sigma" on the inefficiency of tracks have on the analysis? How is "sigma" defined here? Is it a uncertainty or is it related to the standard deviation of the distribution?

The sigma here is the standard deviation of the sample mean of the veto inefficiency for each flavor. This procedure is taken as a conservative measure so the value of 2 isn't rigorously optimized. For an example of the effect, in a sample of 900 GeV, $\tau = 100$ cm charginos in the NLayers6plus category, 745 tracks are selected with the $<2\sigma$ requirement and 730 tracks are selected with $<1\sigma$ required instead.

L322: What is the signal efficiency for your benchmark points for the fiducial cuts?

The fiducial cuts all together remove roughly 20% of signal tracks.

L334: Can you help me understand what the jet $p_T > 110$ GeV cut is achieving? Do you have the ratio of each passing this cut? Its hard to tell how effective it is from the 2D plots in figure 11.

Figure 11 will be extended to jet $p_T > 30$ GeV on the y-axis to better show the issue the text mentions. The $p_T > 110$ GeV cut is used to be consistent with the online requirement of MET from an ISR jet. The efficiency of this cut is 84.6% in data and 87.2% for the signal sample shown.

L375: Does rho include neutral or charged + neutral average energy?

Rho is from the "fixedGridRhoFastjetCentralCalo" collection which uses all PF candidates within $l_{\text{etal}} < 2.5$

Figure 16: Is the difference from one in the first bin labeled "total" from acceptance?

This was normalized incorrectly and will be corrected to one in the next AN version.

L458: Can you add plots (at least a few examples) to the AN of the Z->ll mass distributions in the OS and SS categories used for the T&P method?

We will add these plots.

L490: I don't understand how this test shows that P_veto does not depend on track pT. How significant is the KS test for distributions with so few events?

This question was asked in the pre-approval (see below in the responses). Investigating we found that the estimate as presented is statistically consistent with several other hypotheses of P_veto's pt-dependence, for example a linear dependence. In short we do not have the statistics to determine a pt-dependence or for any potential dependence to affect the estimate.

L523: Can you say a few more words about how the trigger efficiency is calculated?

This is simply the efficiency to pass the signal trigger requirement. As in Section 7.3 we require lepton pt > 55 GeV, so that the efficiency is measured on the IsoTrk50 track leg plateau. An additional sentence to this effect is added now to the AN here.

Table 29-31: Are the uncertainties statistical only?

Statistical only. A comment has been added to these captions.

L590: What level of signal track contamination in the ee/mumu CR would be required before it would affect the background estimate significantly?

To have signal contamination here, there would need to be sources of ee/mumu pairs (a Z or otherwise) in the signal which does not occur in any of our samples; we have 0% contamination now. Even if signal did contain Z->ee/mumu candidates, a track would need to have $|dx| \geq 0.05$ to be fake estimate contamination. The efficiency for the $|dx| < 0.02$ cut on one sample (900 GeV, 100 cm, NLayers6plus) is 99.8% -- so the contamination would be even less than 0.2% times the transfer factor.

L646: Do I understand correctly that the cross check here is to simply take the ration of integrals of the sideband vs. signal region instead of using the fit to determine this ratio?

Not quite. The cross check takes the normal estimate from the sideband (count/integrate the events and scale by the fitted transfer factor), and compares its estimation of the events in the peak to the actual observation in the peak.

L742: Did you compare the trigger efficiency calculated with other reference triggers than SingleMuon?

From a pre-approval question we measured the trigger efficiency using the SingleElectron dataset as well. It was very similar, although electrons introduce hit efficiency effects like conversions so we do not use it in the analysis.

Data comparison	Ratio (electrons/muons)
------------------------	--------------------------------

L837: Is there a good argument why the source of differences in the electron case really is applicable to the muon and tau cases?

The online/offline MET isn't expected to be strongly dependent on the nLayers of a mis-reconstructed lepton, since they've failed to be reconstructed, and so naively one should expect P_offline and P_trigger are all the same. But there is a small difference for electrons, and we assume muons/taus have a similar difference. The

statistical uncertainties for these muon/tau estimates are already 100% or more so even a much larger systematic would not make a difference.

"Some followup questions" from Kevin Stenson HN July 24 [↗](#)

Show Details [▢](#) Hide Details [▢](#)

Table 19: More clarification on dxy and sigma of the track is needed. If dxy is truly with respect to the origin, that is a terrible idea. The beamspot is significantly displaced from the origin (by several mm). So dxy should be measured with respect to either the beamspot or the primary vertex. Regarding sigma, I guess you are saying that it only includes the calculated uncertainty on the track parameters. Can you provide a plot of dxy and sigma for the tracks. Preferably for all tracks. These are applied in the signal selection, so why not here?

See the above answer from the ARC action items. Regarding sigma that is correct.

Table 22: I don't understand your response. Regarding Table 22, my specific questions are: for electrons, why is there no veto on $\text{minDeltaR_track,muon}>0.15$ or $\text{minDeltaR_track,had_tau}>0.15$ or $\text{DeltaR_track,jet}>0.5$ for muons, why is there no veto on $\text{minDeltaR_track,electron}>0.15$ or $E_calo\text{GeV}$ or $\text{minDeltaR_track,had_tau}>0.15$ or $\text{DeltaR_track,jet}>0.5$ for taus, why is there no veto on $\text{minDeltaR_track,electron}>0.15$ or $\text{minDeltaR_track,muon}>0.15$

See the above answer from the ARC action items.

Figure 19 and Tables 29-31: I'm still not sure I understand. Is Figure 19 the plot for all probe tracks, regardless of whether there is a matching tag? If it is just all probe tracks that survive the tag-and-probe requirements, then the fact that they are all plotted and that you use all combinations should mean we get the same answer. On the other hand, it could be the same-sign subtraction is the cause of the difference. In addition to providing the four numbers that go into Equation 6, can you also provide the integrals of the blue and red points in the three plots of Figure 19? I'm hoping that $N_{T\&P}$ and $N^{\text{veto}}_{T\&P}$ will be similar to the integrals of the blue and red points, respectively.

See: above.

L514-518 and Tables 29-31: So my understanding is that when you write "tau background" you are really intending to identify the sum of taus and single hadronic track backgrounds. I think this approach is fine but there may still be some issues with the implementation. The measurement of P_{veto} is going to be dominated by taus as it is measured from Z events and has the same-sign background subtracted. If you are trying to apply P_{veto} to the sum of taus and single hadronic track background it seems like it is necessary to show that P_{veto} is the same for taus and single hadronic tracks. For P_{offline} , the single-tau control sample will clearly be a mix of taus and single hadronic tracks as the selection is relatively loose. This is probably good. However, it will also include fake tracks as there is no same-sign subtraction in this measurement. So I think there is still the possibility that you are including fake tracks here. I guess the fact that there are basically no 4 or 5 layer tracks suggests that the fake track contribution is negligible.

We agree. The tau control region is not sufficient to study the composition of real taus versus single hadronic tracks so there is no feasible study on fake contamination here. Even if for example the contamination was 50-100% the estimates would still be statistically consistent with what we have now.

Figure 25 and Tables 29 and 31: The Figure 25 caption seems to suggest (to me) that the plots show the projection of the events in the upper-right of the red lines. In actuality, the plots show the full distribution. I would suggest changing the plots to only include the events from the upper-right of the red lines in Figures 20-22.

Instead the caption has been changed to be more accurate.

Section 6.1.5: Your response keeps mentioning that P_{veto} is small and therefore other samples are not useful. It may be true that other samples will not contribute when you select the signal region and compare to the background estimate. However, I would think the non-ttbar background will contribute to the measurement of P_{offline} and P_{trigger} , since these simply come from single-lepton samples. So, again, if you want this to mimic the data, I think you need to include all of the background MC samples.

See: above.

Section 6.2: Thanks for the info. I think you are correct that fake tracks are the main contributors to the Gaussian and flat portion of the dxy plot. I don't think there is any bias in the final fit that is done. However, the pattern recognition does have a bias for finding tracks that originate from the beamline. So that could be the reason. I am concerned about one statement. You write that you label a track as fake if it is not "matched to any hard interaction truth particle". Can you clarify what you mean? I am worried that you only check the tracks coming from the "hard interaction" rather than all the truth tracks (including from those from pileup). I think it would be wrong to classify pileup tracks as fake tracks.

Technically this means there is no "packedGenParticles" (e.g. status=1) object with $pt > 10$ GeV within $\Delta R < 0.1$ of the selected track. This designation is not used for any measurement in the analysis, and whatever the source of fake tracks one would need to treat them in data as we've done; we just do not separate them by source.

Section 8.2.2: It is good to know that the hit efficiencies seem to be accurate. However, you also write that charginos behave like muons and so using muons is the correct way to evaluate the hit efficiency. You also write that "the reconstruction is done only with the tracker information". As I wrote earlier, for real muons, there are two additional tracking iterations that use muon detector information to improve the track reconstruction. This won't be the case for your charginos of interest because they decay before reaching the muon stations. So that is why I worry that muons are not a good substitute for your charginos.

See: above.

Section 8.2.5: This response does not really address the heart of the question. Suppose you did have infinite statistics in data and MC. Would we then be comfortable quoting no systematic uncertainty? Are we 100% sure that taking the Z pT and using that to reweight the pT spectrum of the electroweak-ino pair gives the correct distribution? Has anyone looked at applying this procedure to diboson production or ttbar production to confirm that it works?

See: above.

Section 8.2.11: Are you saying that for your 700 GeV chargino signal, the increase in trigger efficiency when going from HLT_MET120 to HLT_MET120 || HLT_MET105_IsoTrk50 is only 1%? I'm not sure this is relevant to my concern though. Suppose you have a signal that produces PFMET of 125 GeV. When you measure the efficiency for nlayers=4,5,6 tracks, you will get the same efficiency because the MC includes a HLT_PFMET120_PFMHT120_IDTight trigger in it. So, you say great, there is no systematic because the efficiencies are the same. However, in some fraction of the data this trigger path is disabled and so the efficiency would be quite different for nlayers=4 tracks (which won't get triggered) and nlayers=6+ tracks which will get triggered by HLT_MET105_IsoTrk50. So, my suggestion was to do the same study but only include triggers that are never disabled or prescaled. Basically, you are currently using HLT_PFMET120_PFMHT120_IDTight || HLT_PFMETNoMu120_PFMHTNoMu120_IDTight || HLT_MET105_IsoTrack50 to make the calculation. My suggestion is to use HLT_PFMET140_PFMHT140_IDTight || HLT_PFMETNoMu140_PFMHTNoMu140_IDTight || HLT_MET105_IsoTrack50. Alternatively, you could correctly weight the MC to account for the different luminosity periods when each trigger is active.

The increase in signal acceptance*efficiency is only 1%, the increase in trigger efficiency is as shown in

Figure 44.

What we've done with trigger efficiency scale factors is to average the data efficiency over the entire data set, which includes the histories of every path. As you suggest we could alternatively measure a different scale factor for every data period, and then take a lumi-weighted average of these different periods -- but these methods are equal. The loss of efficiency in 2017 B is included in the data efficiency and thus the scale factor on signal efficiency.

First set of comments from Kevin Stenson HN July 14 [↗](#)

Show Details [▣](#) Hide Details [▣](#)

Table 5: There seems to be a third column of percentages but there is no column heading or description of what this is in the caption or text. Please clarify.

The \backslash pm was missing; fixed.

L208-209, Table 9: You mention the single tau trigger was prescaled. This appears to result in less than 6 fb⁻¹ of luminosity. Is there no other single tau trigger with less of a prescale that could be used? What is the situation for 2018 data?

Yes, Table 1 details it as 5.75/fb. The only other single tau triggers have much higher thresholds or additional requirements that are unsuitable. The prescale was higher in 2018.

L243-4: Are there eta restrictions on the jets you use? $l_{\text{eta}} < 5$ or $l_{\text{eta}} < 3$ or $l_{\text{eta}} < 2.4$ or $l_{\text{eta}} < 2.1$ or something else? In Table 17 there is a cut of $l_{\text{eta}} < 2.4$ but I'm not clear if this applies to all jets used in the analysis.

Yes, $l_{\text{eta}} < 4.5$ overall; this is added to the AN. Different requirements are as listed e.g. Table 17. The 10 GeV has also been fixed to 30 GeV; MINIAOD contains only >10 GeV jets but the analysis considers only >30 GeV.

L260: I have a general idea of how electrons and muons are reconstructed but not so much with the taus. I seem to think there is some flexibility in tau_h reconstruction. Can you add some information about how the taus are reconstructed? I seem to recall that one can select (with some ambiguity) tau+ decays to pi+ or pi+ pi0 or pi+ 2pi0 or pi+ pi- pi+. Do you consider all tau_h decays or just one-prong decays? It wouldn't hurt to add a few sentences about muon and electron reconstruction as well (or at least some references).

As stated we use the POG recommended decay mode reconstruction with light flavor rejection which does target multiple tau_h decays. This selection is only used to normalize the tau_h background estimate and must be inclusive to all tau_h decays. A brief reference has been added for the PF lepton reconstruction description.

L260-4: Have you checked the efficiency for selecting the correct primary vertex in signal events? One could imagine selecting the vertex based on the origin of the isolated track and/or ISR jet.

We use the standard PV recommendation using the highest sum-pt² vertex (L 260-264); we have no need of a specialized vertexing requirement. Figure 26 for example demonstrates that signal tracks are well-associated with the PV already.

Table 17: I'm a little confused. I get that there must be at least one jet which simultaneously has $p_T > 110$ GeV and $l_{\text{eta}} < 2.4$ and passing tight ID with lepton veto. But when you measure $\max |\Delta \phi_{\text{jet}, \text{jet}}|$, do each of the jets in the comparison need to pass those cuts as well? If so, then this cut only applies in events where there are two or more jets with $p_T > 110$ GeV. That doesn't seem right.

See above; jets are considered if $p_T > 30$ and $l_{\text{eta}} < 4.5$. So $|\Delta \phi_{\text{jet}, \text{jet}}|$ would apply only if there are two or more jets, the minimal case being a ~ 110 GeV and a second ~ 30 GeV jet. A clarifying sentence has been

added.

L332-334: It is claimed that a jet pT cut of >110 GeV removes a lot of background and not signal. The plots in Figure 11 don't seem to back this up. It seems like about the same percentage of signal and background events are removed by the cut. Can you quantify the effect of this cut (background rejection and signal efficiency)?

Figure 11 has been updated to show jet pt >30 GeV instead of >55 GeV, as the issue is seen at lower pt. The efficiency of the >110 GeV cut is 84.6% in data and 87.2% for the signal sample shown.

L338-341: Would be nice to show the signal and background distributions for these two variables so we can evaluate for ourselves the effectiveness of the cut. Also, it would be helpful to report the background rejection and signal efficiency for these cuts.

We are producing N-1 plots to show this.

Table 18: Are there no standard Tracking POG quality requirements on the tracks? I think they still have "loose" and "highPurity" requirements based on an MVA. Do you require either of these?

These exist but we do not use them; the standard quality flags do not make requirements on the hit pattern for example.

Table 19: You need to define what sigma is in the last line. Is it the beamspot width, is it the primary vertex transverse position uncertainty, is it the uncertainty on the track position at the distance of closest approach to the beamline or primary vertex, or some combination of these?

Both dxy and sigma here refer to the dxy measurement with respect to the origin. This is made more clear in the AN.

L365-367: You write that "all" muons and electrons are used. I would like to have a more complete description of this. It may help if you add some text around L260 describing muon and electron reconstruction. For muons, Table 13 defines tight and loose ID. Is it simply the OR of these two that you use? Or do you include tracker muons? I think there is also a soft muon ID. Are these included? What about standalone muons? For electrons, Table 12 defines tight and loose ID. Is it simply the OR of these two categories that you use? Or do you loosen up the requirements further? If so, what are the requirements?

Text describing this around line 365 has been added to explain that "all" means all available in MINIAOD, which has a minimal set of slimming requirements which are now provided in the text.

L368-370 and Table 20: Would be nice to see plots of Delta R to see why 0.15 is chosen. I would have expected a smaller value for muons and a larger value for electrons.

We are producing N-1 plots to show this.

L363-370: Just to be clear, there are no requirements on the pT of the leptons? So, if there is a 4 GeV muon within Delta R of 0.15 of a 100 GeV track, then you reject the track?

As above these are all those available in MINIAOD, which has a very minimal set of slimming requirements. For example muons passing the PF ID have no pt requirement whatsoever. If such a muon as you write is near our track, yes we reject it.

L389-91 and Figure 14: It would be good to plot Figure 14 with finer bins (at least between 0 and 10 GeV) to back up this statement.

The statistics are very limited and the intent is to show the separation between 0-10 and above 10. The statement now states "remove almost all background" which Figure 14 supports.

Figure 16: Why is the first bin at ~30% rather than 100%?

Fixed.

AN Table 22: Why don't you apply the full set of vetos for all lepton flavors? Is it an attempt to increase statistics? Can you perform the test with all vetos applied to see if the results are consistent?

All three sets of requirements are applied in the signal region. In measuring each flavor's $P(\text{veto})$, it's necessary to retain the veto against other flavors to maintain purity of the flavor under study. For example when studying muons, one would still require $E_{\text{Calo}} < 10 \text{ GeV}$. So this is tighter than what you suggest, to achieve better purity of each flavor.

L456-460: Your assumption here is that the same-sign sample has the same production rate as the background. Have you verified this? You could verify it with a high statistics sample of dilepton events (or come up with a scaling factor if it is not exactly true). Also, in L457-458 you list three sources of background: DY, non-DY, fake tracks. I don't see how a same-sign sample can be used to estimate the DY background? I would suggest calling DY part of the signal. For di-electron and di-muon events, you also have the possibility of using the sidebands around the Z mass to estimate the background. You could check that this gives consistent results.

This method was suggested by you in the review of EXO-16-044. The language of the AN regarding continuum DY has been updated for clarity as you suggest. It is not relevant to the measurement of $P(\text{veto})$ to estimate the non-Z backgrounds in our tag-and-probe samples, so we do not -- the purpose of the same-sign subtraction is to increase the purity of the lepton flavor under study.

L468-9: It would be helpful to provide a table giving the values for the 4 numbers in Equation 6 for each of the 9 cases (3 leptons * 3 nlayer bins). I would like to get an idea of the signal-to-background ratio. I may also want to calculate the effect of subtracting the background versus ignoring it.

In making this table we found some bugs in the $P(\text{veto})$ script. Firstly the $N_{\{\text{SS T\&P}\}}$ was not being subtracted in the denominator of Equation 6; in $n_{\text{layers}} \geq 6$ this is an extremely trivial issue but is relevant in the newer, shorter categories. Secondly in cases where the numerator of Equation 6 is negative, the $N_{\{\text{SS T\&P}\}}^{\{\text{veto}\}}$ subtraction was ignored when it should be assumed to be $0 + 1.1 - 0$. Both these issues are resolved and we are currently making the table. This slightly changes some estimates.

L472-5: Are these recHits from the tracker or the calorimeter? I don't really understand what you are describing here. Are the electron seeds from the ECAL or pixel detector?

Throughout the documentation, recHits refers to calorimeter recHits. Electron seeds are ECAL superclusters matched to seed tracks in or close to the pixels -- so electron seeds are from both. We've reworded this section for increased clarity.

L482: It looks like the probe tracks already have a p_T cut $> 30 \text{ GeV}$. So going down to 20 GeV is just being extra safe. Is that right?

Correct.

L478-487: It is not clear to me. Is this re-reconstruction needed for the signal region or not? Was it done for the signal region?

It is not needed for the signal region and is not done, as track $p_T > 55 \text{ GeV}$ is above the threshold of 50 GeV .

Figure 19 and Tables 29-31. If I try to integrate the plots in Figure 19, I would estimate that the integral of red/blue is roughly 10^{-6} , 10^{-7} , and 10^{-4} , for electrons, muons, and taus, respectively. I would expect this to be approximately equal to P_{veto} . But in Tables 29-31, I find P_{veto} numbers of 10^{-5} , 10^{-6} , and 10^{-3} for electrons, muons, and taus for $n_{\text{layers}} \geq 6$. So roughly a factor of 10 off. Can you explain this?

Recall in the review of EXO-16-044 you recommended we utilize all possible tag-and-probe pairs in every event, on top of performing the same-sign subtraction. Figure 19 shows all probe tracks in all events; often there are multiple probes that can be chosen as the tag-and-probe combination. As you say these figures are related to the value of $P(\text{veto})$, but are not precisely equal.

Figures 19, 21, 25, and 40 and page 72: I would suggest removing footnote #17 on page 72 and adding that information into the captions for figures 19, 21, 25, and 40. You should also add a similar explanation to the caption of Table 31 indicating that N_{ctrl} is scaled to the signal region luminosity.

We have added the mention to Table 31's caption, but the authors feel the luminosity label is sufficient in those figures.

Figure 22: Would be good to show the results for $n_{\text{layers}}=4$ and $n_{\text{layers}}=5$ (unless there are no entries in which case you should note that in the caption), similar to the way you show the results for tau for $n_{\text{layers}}=5$ even though it is not used.

There are 1 and 2 events respectively for $n_{\text{layers}}=4$ and $=5$ respectively, thus these plots were found unhelpful. A comment has been added to the caption of Figure 22 to mention this.

L514-518 and Tables 29-31: I note that P_{offline} for electrons and muons is very similar, around 80%, while for taus it is much lower, around 20%. Do you understand the difference and do you think it is OK for the method. I can imagine two effects that could cause this. First, it could be that since the pion from the tau decay does not carry all of the tau momentum, the tau candidates from W decays will have a lower p_T than the muon or electron candidates from W decays. So when the tau p_T gets added to $p_{T\text{miss}}$, it will get shifted less than when the electron p_T gets added and so more will fail the Ecalo cut. Based on Figure 25, this seems to be true and is probably innocuous. But I think there is more to it. Comparing Figures 20 and 21, the modified $p_{T\text{miss}}$ for the tau case has a large contribution at the bottom left corner that is not present in the electron (or muon) case. It seems that the electrons and muon are very consistent with the topology of a W recoiling from an ISR jet so $\Delta\phi$ is $\sim\pi$. However, for the tau case, there seems to be many events where the "tau" is part of the leading jet. I guess that since there is a ΔR cut of 0.5, the "tau" must differ from the jet a bit in η . Given this evidence and the fact that we know that the tau purity is much worse than electron and muon purity, it seems likely that many of the events in Figure 21 do not contain taus. I would guess the events are multijet QCD events with either an isolated track by chance or a fake track. So, my hypothesis is that the single tau control region has a large contamination of non-tau events. You use the same sample for measuring P_{offline} and the multiplying by P_{offline} as part of estimating the tau background. So we could consider this estimate as being the tau+single hadronic track+fake track contribution. But there are two problems with that. First, P_{veto} is measured on a much purer sample of taus as it uses Z decays and subtracts the same-sign contribution. So P_{veto} is really measuring taus, not the sum of tau+single hadronic track+fake track. And P_{veto} may not be the same if the other contributions were included. Second, you have a separate measurement of the fake track contribution so you would be double counting. Please let me know what you think.

Our interpretation of the lower modified-MET distribution for taus is the same as yours here, and we too feel it is innocuous.

We intentionally capture the "single hadronic track" component as part of the tau estimate. These do contribute as a background, and are included as part of our "tau background"; the analyzers feel that calling this the "tau and single hadronic track background" is a distraction for the reader, beyond the one mention in L454-455. We capture this contribution in two ways: firstly as other reviewers have noticed, our hadronic tau

ID is fairly loose, and secondly we remove the requirement on $\Delta R(\text{track}, \text{jet}) > 0.5$. Thus the tau $P(\text{veto})$ considers all of these contributions, and you see in Figure 21 that some of them have a lower probability to pass the $\Delta\Phi(\text{jet}, \text{MET})$ requirement which is then included in $P(\text{offline})$. For clarity we have added a reminder of the $\Delta R(\text{track}, \text{jet})$ cut removal to L455 in the AN.

Lastly the "fake track" contribution will not survive the same-sign subtraction, whereas the "single hadronic track" contribution will. So we are not concerned about double-counting the fake tracks.

Section 6.1.3: I think you need to be a little more clear here. I want to confirm that I understand. First, the figures mention HLT efficiency but I think this is really the full trigger efficiency (L1+HLT). Do you agree? Second, the trigger efficiencies shown in Figure 23 and 24 are the actual results from the L1+HLT that was run and the x-axis refers to the actual $p_{T\text{miss,nomu}}$ of the event. That is, the x-axis is not the modified $p_{T\text{miss,nomu}}$ where the electron p_T is added back in. Is that correct? Then, the x-axis of Figure 25 shows the modified $p_{T\text{miss,nomu}}$ with the electron or tau p_T added back in. Is that correct? Try to make the text and figures a bit clearer.

Figures 23 and 24 now are labeled as just "trigger efficiency", and the caption for Figure 25 has been made more clear.

Figure 25 and Tables 29 and 31: Figure 25 seems to show that the electron distribution is shifted higher than the tau distribution. Therefore, once you convolute this distribution with the trigger efficiency, I would expect the electron trigger efficiency to be higher than the tau trigger efficiency. However, the opposite is true. If I naively take the trigger efficiency as a step function which is 0% for $p_{T\text{miss}} \leq 200$ GeV and 100% for $p_{T\text{miss}} > 200$ GeV, I think I get about 30% for electrons and 13% for the tau, compared to 46% and 52%. Can you check the results and if correct, try to explain what I am missing?

Your numbers are correct if you integrate Figure 25 across the entire MET range. However $P(\text{trigger})$ is a conditional probability after $P(\text{offline})$ has already required $\text{metNoMu} > 120$ GeV, so that must be applied.

Table 31: How did you determine the uncertainties for $n_{\text{layers}}=4$? The upper uncertainty of 0 for N^{ctrl} and the upper uncertainty of 0.0058 for estimate seem too small. Actually $n_{\text{layers}} = 4$ and $n_{\text{layers}} = 5$ for both muons and taus (Tables 30 and 31) have yields that are too small to assume Gaussian uncertainties. You should use Poisson uncertainties. You can ask the statistics committee for better advice but I think more correct uncertainties would be: $0 +1.15 -0.1 +1.36 -0.62 +1.52 -0.86$ This comes from the prescription on page 32 of <http://pdg.lbl.gov/2019/reviews/rpp2018-rev-statistics.pdf> but again, the statistics committee may have another prescription. Note that I think this results in an estimate for the tau background for $n_{\text{layers}}=4$ to be $0 +1.9 -0.0$ rather than $0 +0.0058 -0$.

This has been corrected to " $0_{-0}^{+8.2}$ " using poisson errors; the +1.15 must be multiplied by the tau trigger prescale. This is also corrected in the total lepton and total background stat uncertainties. Only the table values needed correction -- the upper limits used the correct values.

Section 6.1.5: There needs to be more information. Do you calculate P_{veto} , P_{offline} , and P_{trigger} for each of the leptons using simulated samples following the same recipe as for data? If so, what simulated samples? Do you just use $Z \rightarrow ll$ for P_{veto} and $W \rightarrow l\nu$ for the others? Does the single lepton control region come from just $W \rightarrow l\nu$ events? Or do you include all the background samples in Tables 7-8? I think using all of the samples from Tables 7-8 for every calculation would make the most sense.

With the modifications given, we calculate the background estimates in precisely the same way as in data. The AN has been clarified on that. For the lepton closure we use only the $t\bar{t}$ samples because the other samples do not affect the statistics for this study due to the small $P(\text{veto})$.

Section 6.1.5: It seems like even with the relaxed selection criteria, you are still quite lacking in MC statistics for this test. You mention that you only include $t\bar{t}$ events. Is this just the $t\bar{t}$ semileptonic sample?

Although this may be the largest single sample, it seems like you could also include other samples. Most importantly would be the $W \rightarrow \nu$ and $Z \rightarrow$ invisible (including the HT-binned samples) as these seem to be the largest source of background in Figures 14 and 15. Is there some reason you didn't include these? If not, I suggest you go ahead and do this.

All three $t\bar{t}$ samples are used, but the di-leptonic $t\bar{t}$ sample contributes the most. Keep in mind that $P(\text{veto})$ is a tag-and-probe selection, and samples such as $W \rightarrow \nu$ and $Z \rightarrow$ invisible will not significantly contribute. The $Z \rightarrow ll$ sample should contribute, but the size of those samples is considerably smaller than the $t\bar{t}$ samples.

Figure 26: Are all three results properly normalized to 41.5 fb^{-1} ? If so, it seems like we should be including 3 layer tracks in the signal region because I can exclude a $ct=10\text{cm}$ with this plot alone (observe 350 with a prediction of 125), which you can't do with the whole analysis.

Yes, the results are properly normalized, but there is no observation in Fig. 26 -- all the entries are MC. If we were to include data, we would expect the fake contribution to the 3 layer tracks to be orders of magnitude larger making any exclusion difficult without a dedicated analysis.

Figure 27: Would be good to add the $n_{\text{layers}}=4$ result as is done in Figure 28. Should also say what simulated samples are included here.

All samples are used, the caption is updated. The cleaning cuts are only relevant for the $n_{\text{layers}}=3$ category which is only used in an appendix (after the edits suggested below). Showing this for $n_{\text{layers}}=4$ would provide the same information as Figure 31.

Section 6.2: This needs to be cleaned up and explained better. Here are some specific comments/suggestions - I think that L566-569, Figures 26-28, and L602-615 can all be removed. It seems like they have nothing to do with the analysis that is done. They just lead to confusion. If you want to move this material to Appendix C, that is fine. But don't clutter up this section.

These have been moved to Appendix C.

- I'm confused by the transfer factor. I assume that the fit in Figure 29 is actually a Gaussian + flat line. Is that correct? What is your hypothesis about what is contained in the Gaussian area and what is contained in the flat line area? I would have assumed that Gaussian contribution indicates real tracks (since they peak at $d_0=0$) and the flat line contribution indicates fake tracks. But this doesn't seem to match your hypothesis. Can you say exactly what the fit in Eq. 14 is doing? In L629-630 you quote a single transfer factor for each Z mode. Shouldn't there be a different transfer factor for each of the 9 sideband regions?

a) Correct, the fit is a gaussian + constant. The AN has been clarified.

b) Our hypothesis is that this is a bias in the track-fitting algorithm, where short tracks with very few hits have the importance of the primary vertex inflated, drawing tracks closer to the PV. Figure 27 shows this also occurs in SM background MC, and in those MC samples none of the tracks are near to any hard interaction truth particle; this is precisely our consideration of "fake" in MC truth.

c) The purpose of the transfer factor is only to normalize the sideband rates to the signal region. We must describe this normalization in a way that does not depend on observing the signal region count, because in $n_{\text{layers}}=5, \geq 6$ the statistics do not allow for that. That is what the fit does in Eq. 14.

d) L629-630 quotes the transfer factor for the baseline sideband (0.05, 0.10) cm, so only one. The authors felt that Table 35 was large enough already, but we now provide an additional table listing the $P^{\text{raw}}_{\text{fake}}$ and transfer factors.

- What best describes your assumption of the fake track rate as a function of d_0 . Is it uniform (flat), Gaussian, Gaussian+flat, or something else?

Gaussian + flat.

- I don't see the advantage of having 9 different sideband regions. Simply take the sum of events from 0.05-0.5 and multiply by the overall transfer factor. This should minimize the statistical uncertainty. In fact, I would suggest combining the Z->mumu and Z->ee samples as well. Also, remember to use the correct Poisson uncertainties (as discussed for Table 31) when you only have a handful of events. If you somehow think it is a good idea to have 18 different measurements instead of 1 and you are using a transfer factor with an uncertainty, make sure to properly account for the fact that this uncertainty is correlated for different bins.

As discussed in the first ARC meeting, we've used a single larger sideband. The fit uncertainty is employed as a nuisance parameter 100% correlated between bins.

- L635-638: As mentioned above, I would suggest combining the Z->mumu and Z->ee results to get the final estimate, seeing as you are statistics limited. You can still use the difference between the two as a systematic uncertainty (but see below).

Doing as you suggest is acceptable, but will not change the estimate very much. This was discussed in the first ARC meeting and will be revisited.

- L640-645: It is obvious that Z->mumu and Z->ee events are quite similar. They have the same production mechanism, they are selected by single lepton triggers, etc. So, it is not much of a test to show that they give the same result. On the other hand, your signal region requires large missing ET, a high p_T jet, and a high p_T isolated track that is neither a muon or electron. One might worry that the fake track rate depends on the amount of hadronic activity in an event, which is likely higher in the signal region than in Z events. One might also worry that the fake track rate depends on pileup, and the signal trigger/selection may be more susceptible to pileup than the single lepton trigger/selection. Ideally, I would suggest that you perform the same measurement on a QCD dominated region (like requiring a dijet or quadjet trigger or just high HT). You can require $p_{T,miss, no \mu} < 100$ GeV to ensure no signal contamination. If this is not possible, then you could consider taking what you have and either reweighting the pileup and HT distribution to match the signal region or checking that the fake rate is independent of these quantities.

See the above response (from "ARC action items from July 25 meeting"). Small differences exist between pileup in these different samples, but reweighting for those differences doesn't change estimate.

- L649-653: I don't understand how these numbers are consistent with Figure 29. In Figure 29 (left) it seems there are about 9 events with $|dx_{y1}| < 0.02$ cm and about 15 with $0.05 < |dx_{y1}| < 0.10$ cm to be compared with 32 events and 68 events. There is a similar discrepancy for electrons. I guess the plots have been scaled for some reason as the entries are not integers. Please fix the plots and verify the results are consistent.

The scaling of the plots is now fixed, and agrees with the text.

- Figure 29: Why do you not fit the region $|dx_{y1}| < 0.1$ cm? If you fit out to $|dx_{y1}| = 1.0$ cm, please show the entire range in the plots. It would be nice to see the results for $n_{layers} = 5$ and 6 as well so we can evaluate the extent to which a fit may or may not be possible and whether the shape is consistent with $n_{layers} = 4$.

The AN has been updated to correctly reflect the fit extending to $|dx_{y1}| < 0.5$ cm, the range of the plots. Should the d_0 peak actually contain real tracks, it would peak more narrowly than observed in the sidebands; so $|dx_{y1}| < 0.1$ cm is excluded from the fit, and the count of $n_{layers} = 4$ tracks in the signal region is checked against the fit prediction, and agrees.

Shown below is the $n_{layers} = 5$ d_0 distributions, with the fit from $n_{layers} = 4$ overlaid. The $n_{layers} \geq 6$ samples

have one (three) events in ZtoMuMu (ZtoEE), so no fit is possible.

ZtoMuMu NLayers5	ZtoEE NLayers5
------------------	----------------

Section 6.2.2: In Table 36, it would be enlightening to show the same results as in Table 35. That is, I am curious as to how P_{fake} compares between data and MC. Are these results normalized to 41 fb^{-1} ? If so, then it seems like the MC predicts about 1/5 as many fake tracks as data. It is hard to be confident that the MC tells us anything if that is so.

We felt that Table 35 was distractingly large, and the low MC statistics makes things worse for Table 36. That we are moving to one single sideband, Table 35 is less relevant too. Table 36 is normalized to $41/\text{fb}$. The value of Table 36 is the test of closure in the fake estimate method, rather than the absolute rate of fake tracks in simulation which has always been an issue, thus the data-driven estimate is a must.

Section 6.2.2: Your hypothesis is that the fake track rate is independent of selection so you can use the Z data to estimate the fake track rate in your signal region. I have suggested that you could also measure the fake track rate in QCD events to verify this. You can also check the effect in MC. I guess in Section 6.2.2 you apply the same criteria to MC as you do for data (selecting Z events). However, if your hypothesis is true, then you should also get the same fake rate if you use any MC sample. What happens if you use all the samples in Section 3.3 but remove the Table 33 and 34 requirements so you are using all events? If P_{fake} changes significantly, this is cause for concern. If not, then that is good. In either case, it still may not prove anything if the MC is really predicting 1/5 the amount of fake tracks.

As above the absolute rate of fake tracks in simulation is not well trusted, so the comparison of 1/5 to data does not concern us. Certainly one could use additional MC samples and change the selection, but this then deviates from the treatment in data and in principle is not the same closure test. If a third selection/sample were used then the closure test in MC would also need to be used.

Figure 31: Would be good to have a plot for $n_{\text{layers}}=5$ as well.

We are producing this plot.

Figure 35: Please include the ratio of the two since this provides the scale factors that are used. It may be better to simply include the region of 50-300 GeV on a linear scale.

We are producing this plot.

L752-754: While this signal yield reduction is interesting, just as interesting would be the change after all cuts are applied (with $n_{\text{layers}} \geq 4$). Can you provide this as well?

Producing this.

L760-762: How sure are we that the Z can be used to measure ISR difference for the signal model? I generally agree with the statement that both recoil off ISR but it would be nice if this could be confirmed somehow. Does the p_T distribution for a 100 GeV chargino look similar to a Z in Pythia8? Does the ISR reweighting work for $t\bar{t}$ events or diboson events?

We are proceeding through the SIM/RECO steps for a small portion of the Pythia8 Drell-Yan sample generated for Figure 38. By applying the ISR weights and checking against the reconstructed MadGraph sample, this will be an effective check of the method. For the p_T distribution we are producing this for 100 GeV; but for 1000 GeV for example this surely will be different. The relevant issue is not that the Z's distribution looks like the electroweak-ino pair's, but that the simulation underestimates the ISR shape compared to data for the same event content.

L772-773: Please expand on "is applied to the simulated signal samples". Do you reweight the events using the ISR jet in the event or the net momentum of the produced SUSY particles or something else.

The vector sum p_T of the gen-level electroweak-ino pair is used to evaluate the weights. The AN is clarified.

Section 8.2.2: Please expand on this. I am very surprised that this is such a small effect. Given the problems encountered with the Phase 1 pixel detector (problems with timing levels 1 and 3, way more noise than expected in layer 1 causing high thresholds, DC-DC converter failures, etc.) I would have expected big differences between data and simulation on quantities requiring hits in the pixel detector. I know the tracking reconstruction was changed at HLT and offline to keep track reconstruction efficiency high but this doesn't remove the problem of missing pixel hits. So please expand on how you measure these uncertainties. Do you just use tracks with $p_T > 55$ GeV that are associated with a muon? One problem with using muons to evaluate tracking efficiency is that there are special track reconstruction techniques developed to recover muons missed by the standard tracking. These tend to use wider windows to discover silicon hits and so may not reflect the track reconstruction of "standard" charged particles. You could perhaps remove the electron and tau vetoes to see what you get in those cases.

THE AN describes the process correctly. The global tag used for signal was formed well after data-taking was completed for 2017, and has updated hit efficiencies. Further this efficiency is very high which affects the scale of this value. For missing middle hits, the *inefficiency* has a 4.5% difference between data and MC, but it is the *efficiency* which is 0.02% different.

Before the chargino decays, signal tracks are muon-like and would be treated the same way. The muon control region is still a track selection ($p_T > 55$, $MET > 120$, jet $p_T > 110$), and the reconstruction is done only with the tracker information. The electron/tau vetoes need to remain so that this sample is dominated by muon tracks, and is comparable to the signal tracks before they decay.

Section 8.2.5: This seems like an underestimate of the systematic uncertainty. If you had infinite data and MC statistics, your systematic uncertainty would be 0. As mentioned above, this doesn't address whether measuring the ISR using $Z \rightarrow \mu\mu$ decays translates exactly into the ISR for the signal process. The paper mentions this is up to a 350% correction, so it is a big effect. I am very worried that the systematic uncertainty does not cover all that we don't know. I note that Figure 37 shows results with p_T and p_{Tmiss} . Why did you use p_T ? Perhaps p_{Tmiss} could also be used as a systematic check.

Most of this uncertainty comes from the data/MC correction at lower sum- p_T 's, where we do not have infinite data statistics. Moreover these statistical uncertainties are largest where our signal populates the least, which lowers this systematic uncertainty. The p_{Tmiss} is a useful cross-check which was requested by conveners, but the sum- p_T of a diMuon system has a much better resolution than p_{Tmiss} ; it is a very common tool in characterizing the hadronic recoil in many analyses.

Figure 43: I would suggest including the same comparison vs p_T from the document you reference. This shows that for $p_T > 55$ GeV, the differences are similar as for lower p_T .

Added.

Section 8.2.11: Please explain this better. My understanding is that various prescales were in place. In the original measurement of the trigger efficiency in Section 7.3, you rely on being above the track requirement plateau to measure the trigger efficiency versus p_{Tmiss} in data (which has all of the various prescales naturally included) and compare to MC (which just has an OR of all trigger paths, I think). This is the main reason why the data efficiency is lower than the MC efficiency. Is this correct? Now, in this section, you are measuring the trigger efficiency solely with MC, which is an OR of all trigger paths. So if the trigger path with the track requirement fails, the MC might still find that an MHT only trigger will fire, while in data the MHT-only trigger may be prescaled. Isn't this a problem? Can you perhaps repeat the exercise using only triggers in MC that were never prescaled (as the opposite extreme to assuming there was never any prescale)?

Also, why would you average over the chargino lifetimes? Shouldn't this systematic uncertainty depend very strongly on chargino lifetime?

Several triggers were disabled in portions of 2017, so not precisely "prescaled" but we understand. The main difference between data and MC in Section 7.3 is that this history was not in the simulation; this operational history is averaged over in data and applied to the simulation with these weights.

For this section, consider a simple worst-case scenario: the HLT_MET105_IsoTrk50 path which was disabled for 2017B (10% of the lumi), and a 100% enabled path HLT_MET120. In 2017B conditions the efficiency is solely that of HLT_MET120, and you end up over-estimating the trigger efficiency by the difference in efficiency between HLT_MET120 and the OR of the two. Figure 44 and the systematic in Table 45 show this difference to be very small (~1%), and this would only apply to 10% of the data in 2017B, very contained by this systematic. Ignoring the IsoTrk50 path, the triggers dominating the MET turn-on from Table 9 are HLT_PFMET(noMu)120_PFMHT(noMu)120_IDTight which is very similar to this simple worst-case example.

Lastly if the charginos are reconstructed at all they would be reconstructed as muons and will not be included in metNoMu. The only way they can contribute to the metNoMu is by affecting the recoil of the ISR jet, which is why we average over chargino lifetime but measure this systematic separately for each chargino mass.

Section 8.2.11: Per my discussion of pixel issues in 2017. It is relatively easy to get 5 pixel hits with only 4 pixel layers as in order to make a hermetic cylindrical detector with flat sensors, you need to have overlaps. These overlaps are largest in the first layer, which is where there were significant issues with the Phase 1 pixel detector. So I am concerned that if the MC is optimistic about layer 1 hits, then relying on the MC may not be wise. Maybe you can check the following. Take good tracks (not muons but large number of hits with $p_T > 50$ GeV). Check the fraction of tracks that have two layer 1 pixel hits compared to one layer 1 pixel hits between MC and data. Or, more generally, the average number of pixel hits. If they differ, then you could see how many times you would go from 5 pixel hits to 4 pixel hits in data vs MC and use this difference as another estimate of the difference in trigger efficiency.

We do not have the entire hitPattern contents histogrammed so this is not quickly answerable as suggested. However in the electron control region ($pt > 55$ GeV) with $n_{layers}=4$, 11% of tracks have more than 4 pixel hits whereas signal ranges from 8-12%. This comparison however would only describe the offline association of hits to tracks, which is known to be better than the online association -- so one would still need to examine the trigger fires as Section 8.2.11 does to get a clear view of the difference in trigger efficiencies.

Here are also some brief comments on the paper:

Whenever you have a range, it should be written in regular (not math) mode with a double hyphen in LaTeX and no spaces. That is, "1--2". Done correctly in L44, . Incorrect in L186, L285, L321, L326, L327, L331

Fixed.

In Section 2, I think it would be good to give more information about the tracker, especially the Phase 1 pixel detector. It is pretty important to know that we expect particles to pass through 4 pixel layers.

Lines 28-31 should establish the extra categories that are possible thanks to the upgrade. A sentence listing the positions of the layers/disks has been added to Section 2.

Should mention the difference between number of hits and number of layers with a hit.

Lines 185-187 have been expanded to mention this.

L60-67: At the end you talk about physics quantities like $\tan\beta$, μ , and the chargino-neutralino mass difference. In principle, I believe the lifetime is set by the masses (mainly mass difference) of the chargino and neutralino. I think you need to be clear that the lifetimes are changed arbitrarily and also give the mass difference (could just say 0.2 GeV).

Reworded to mention that more clearly. Typically the mass values would be included in the HEPdata entry because they vary, and space in a letter is too limited to include a full table.

L113: $p_{T\text{miss}}$ and $p_{T\text{miss}}$ no μ should be vectors

Fixed.

L128: Should say why $\text{le}_{\text{t}} < 2.1$ is used.

Added.

L131: need to specify the η and p_T requirements on the jets, perhaps in L103-107.

The p_T is specified now. The jet η requirement is $\text{le}_{\text{t}} < 4.5$, which for tracks with $\text{le}_{\text{t}} < 2.1$ is all of them so it's left out.

L157: Should describe hadronic tau reconstruction. Could be at the end of L91-102 where electrons, muons, and charged hadron reconstruction is described.

The PubComm does not give a recommendation for this, and typically hadronically decaying taus are included in the mention of charged hadron reconstruction.

L168,L177: Given that your special procedure removes 4% of the signal tracks, it is natural to wonder what fraction of the signal tracks are removed by the requirements of L159-168.

Added, it's 20%.

L179: Commas after "Firstly" and "Secondly"

Fixed.

L190: Should make it clear that leptons here refers to electrons, muons, and taus.

Okay.

L194, 196, 222, 231: The ordering of P_{offline} and P_{trigger} in L194,196 is different than in L222,231. Better to be consistent.

Fixed in in L193-197.

L204: I think you mean "excepting" rather than "expecting"

Fixed.

L214: I don't think you need the subscript "invmass" given that you define it that way in L213.

Removed.

L222: Change "condition" to "conditional"

Fixed.

L227: p_T^{\perp} should be a vector

Vectorized.

L234-238: This will need to be expanded to make it clear

Reworded somewhat to improve clarity.

L247: I don't think it is useful to mention a closure test with 2% of the data. I mean a 2% test may reveal something that is horribly wrong but it is not going to convince anyone that you know what you are doing.

Removed.

L339 and Table 3 caption: Suggest changing "signal yields" to "signal efficiencies"

Changed..

Table 4: I guess to match the text it should be "spurious tracks" instead of "fake tracks"

Changed..

Lots of the references need minor fixing. The main problems are - volume letter needs to go with title and not volume number: refs 2, 8, 26, 27, 30, 39, 40, 41 - only the first page number should be given: refs 2, 19, 30 - no issue number should be given: refs 8, 13, 31 - PDG should use the Bibtex entry given here: <https://twiki.cern.ch/twiki/bin/view/CMS/Internal/PubGuidelines> - ref 40 needs help

All addressed.

Questions from Juan Alcaraz (July 5)

Show Details Hide Details

Regarding the Z (or ewkino pair) recoil correction, are you really performing the following two steps for the signal: 1) reweight from Pythia8 to MG as a function of the recoil p_T ; 2) reweight again the resulting signal MC according to the data/MC observed recoil spectrum in Z- \rightarrow muon events? Also, let me ask again (probably you answer that at the pre-approval meeting, but I forgot): the data/MC discrepancy at low dimuon p_T was just due to the lack of MC dimuon events at low invariant mass? (this should be irrelevant given the ISR jet cut used in the analysis, but just to understand).

This is correct, we apply both weights. In the ARC review of EXO-16-044 (in which Kevin Stenson was chair and will recall), it was noted that only applying the data/(MG MC) correction would only correct the MG distribution to that seen in data. As our signal is generated in Pythia, we need to correct Pythia's distribution to that of MG's first, otherwise the first correction is not applicable.

Yes, the discrepancy at low dimuon p_T is driven by the drell-yan samples; in 2017 the samples available were $M > 5$ GeV, and $M > 10$ GeV in 2018. Yes, this is irrelevant given the ISR cut.

If one of the trigger paths has a tighter cut (5 hits) than the offline cut, why did not you redefine the offline cuts and required ≥ 5 hits when ONLY that trigger path is fired? I do not see any right to assume that we can count on an extra efficiency that does not really exist, even if it is small. Am I missing anything?

There is a non-zero probability that a track has multiple hits associated to it in the same pixel (9% in one signal sample for example). This allows tracks with only 4 layers to have ≥ 5 hits and fire the IsoTrk50 leg.

As you suspect, the addition of this trigger has a small effect on the efficiency for the nlayer = 4 bin as shown in the left plot below (of course for the nlayers > 6 bin the effect of is much more larger, as shown in the right plot).

☐

On a tangent matter: when can we expect to have any kind of 2018 results ? Despite the suggestion from the EXO conveners I am a bit uncomfortable with considering this step as a trivial top-up operation in an analysis like this one. We know by experience that each new year can give rise to new features and then change significantly the rate of pathological background events that we have to consider...

The above section will for now provide immediate updates for 2018 results. See also this recent update [☞](#) with recent updates in 2018 ABC.

Pre-approval

Additional pre-approval followup Ivan Mikulec HN June 21 [☞](#)

Show Details [☐](#) Hide Details [☐](#)

In your answer to (4) the plots show only the correction related to Fig. 37 in the AN. First, it is a bit surprising that there is a residual MC overprediction at high pT (right plot on the twiki) and the effect on recoil (left and middle figure) is marginal. Second, and more importantly, in the Pythia/MG part of the correction which is in Fig. 38 of the AN, it seems that Pythia does not generate enough high recoil events, so the resulting weight on high recoil signal (>~250 GeV) seems completely saturated. Do you have convincing arguments that this is not an issue?

The residual MC overprediction is an artifact of the fact that we reweight the background MC evaluating them as a function of GEN-level electroweak-ino pair pT in our AMSB signal, not as function of reconstructed di-muon pT that is plotted.

For Pythia/MG, you are correct that Pythia does not generate enough high recoil events. This is a well-known feature of Pythia and one of the main reasons why MadGraph was developed, and why such a correction is necessary to correctly describe the AMSB hypothesis. Yes, it does result in weights of 3--4 for events >~ 250 GeV, but this is a necessary correction, so we don't see it as an issue.

We find the first paragraph of the answer to (5) confusing. If most of the signal events are in the plateau, why not cutaway the turn on in the selection? Anyway, according to Fig. 36 in the AN, quite some part of signal is in the turn on. If this is the case, we find unbelievable that you can be confident about your efficiency in the middle of the steep turn on to relative uncertainty of the order of 0.5%. We still think that a check with different datasets might provide some handle on the related systematics (position and slope of the turn on). We hope that ARC can pay attention to this issue. We are fine with the second paragraph of the answer.

Since MET from ISR is only needed for the trigger strategy, as a search for the disappearing tracks signature we wish to keep as much acceptance as possible. The small uncertainties you mention are those from the statistical uncertainties in the data and SM background MC efficiency measurements, and are small due to those samples being very large. We recently added Section 8.2.11 and Figure 44 to the AN in version 7 which introduces a signal systematic for the shorter (==4, ==5 layers) track categories due to the turn-on region for those. Only about 10% of the signal is on the turn-on so even a 10% uncertainty in the turn-on region only results in a 1% yield systematic -- this new AN section resulting in a 1.1% and 0.5% systematic for ==4 and ==5 layers respectively. In the next version of the AN we will combine all of the trigger signal systematics into one section to make it easier to read.

☐

Also as requested we measured the trigger efficiency in data using electrons instead of muons; see the below plots. One can take the ratio of these efficiencies and apply them as a weight (as a function of MET) to derive another signal systematic on the signal yields. This would give a 2.7-3.2% downwards systematic across the

NLayers categories, using 700GeV 100cm charginos as an example. The analyzers feel however that this is not appropriate to use, because the chargino signature is muon-like in the tracker and electrons introduce hit pattern effects due to conversions and bremsstrahlung which would not affect the signal.

Data comparison	Ratio (electrons/muons)
------------------------	--------------------------------

Questions from pre-approval EXO June 1 [↗](#)

Show Details [▢](#) Hide Details [▢](#)

Thanks a lot for a comprehensive preapproval presentation. Overall the analysis is in good shape. Here is the list of comments/questions that came up during the preapproval

(1) The MET+IsoTrk trigger requires at least 5 hits on the isolated track whereas the analysis starts with short tracks with 4 hits. Please show the trigger turn-on curves for the signal for the different bins of number of tracker layers considered in the analysis, and compare with the turn-on you get with the single-muon events. It would also be good to see the turn-on curves separately for the MET+IsoTrk trigger and the other MET(NoMu) triggers.

See below for several plots, which will be added to the analysis note. Some of these are relevant for (5) below.

MET leg of HLT_MET105_IsoTrk50	OR of all used paths (AMSB 700GeV 100cm)
---------------------------------------	---

(2) The uncertainty on the P(veto) estimate is set to ~10-15%. However, we cannot verify this in the closure test due to lack of statistical power. Please demonstrate that the uncertainty on P(veto) is sufficient. Also, assess the impact of this uncertainty on the analysis. Do the results change significantly on inflating this uncertainty ?

We studied the ratio of the pt of tracks after to before applying the lepton veto, and found that the statistics were too poor to determine a dependence of P(veto) on track pt; all pt-binned values are consistent with the average over all pt. We attempted anyways to fit the pt-binned ratios to a linear function, and found in one case a linear dependence could increase one background by 17.4%, but in all other cases the result was actually a decrease in backgrounds or no change at all. The fit uncertainties were very large. Even in the worst case assumption of a $\pm 17.4\%$ uncertainty on all lepton background estimates, there was no discernible change in our upper limits.

As such we find that the pt-average value of P(veto) used for the estimates is consistent with any possible pt-dependence in the statistically limited data we have available.

(3) Slide 23 : The 1.9% uncertainty can be dropped. Please rebin the track d0 distribution when showing the Gaussian fit.

The 1.9% has been dropped and the plots rebinned in the AN.

(4) Slide 28 : Show the data/MC comparison of the recoil (METNoMu) distribution before and after the corrections.

We attach those plots here. Showing after the corrections is actually not completely trivial, because we apply the ISR weights as a function of the sum-PT of the electroweak-ino pair of which there is no pair in SM background MC. We felt the most correct procedure to show these figures after a correction was to correct only drell-yan, which has a clear gen-level muon pair to evaluate the weights at. As such the plots after the correction will not agree by construction as one might expect.

Before ISR weights applied	Re-weighted DY - metNoMu	Re-weighted DY - di-muon pt
-----------------------------------	---------------------------------	------------------------------------

(5) Please reassess the systematic uncertainty on the trigger efficiency - it seems to be too small. One

possibility would be to compare the trigger turn-ons between single-muon and single-electron events. Also make sure that any potential systematic due to the fact that the number of hits requirement in the IsoTrk trigger is tighter w.r.t. offline selection is also taken into account.

Taking a closer look at the values involved, we believe they are correct despite being small. This is because the bulk of the selected signal events have larger MET well onto the efficiency plateau where the scale factors and SF uncertainties are smallest – see the below plot for one signal sample. In that plot for example the data efficiency just above our offline requirement at ~ 122 GeV, the data efficiency is 5.680 ± 0.026 , a relative 0.46% error; however a very small fraction of the accepted events receive that scale of systematic, so the total systematic is very small.

For the tightness of the IsoTrk50 leg wrt the 4 and 5 layer categories, we do see a difference in the turn-on since the IsoTrk50 path has the lowest MET threshold and has reduced efficiency. Unfortunately however we cannot measure the trigger efficiency in muon data and SM background MC as 4- and 5-layer muons tend not to exist with reasonable statistics. It's not precisely the correct thing to do, but without a data measurement in 4- and 5-layers the best option we see is to take a very, very conservative systematic based on the difference between trigger efficiencies in signal for the 4/5 layer categories to the 6+ layer category -- e.g., that would cover any difference between data/simulation in those samples if the difference was as large as it was between nLayer categories, which is quite conservative. This would result in an average signal systematic of 1.1% (4 layers) and 0.5% (5 layers).

▣

(6) Currently, a veto is applied on candidate tracks overlapping with any reconstructed leptons. Please quote the signal efficiency for this veto. It would be good to ensure that we do not spuriously lose events in data by matching tracks to some mismeasured objects that are not simulated well. Please check the impact on $P(\text{veto})$ of requiring at least some loose selection (e.g. requiring the muon to be a PF muon or a loose muon).

This suggests, quite correctly, that we use a lepton selection that will have some scale factor between simulation and data for its efficiency. We measure this scale factor relative to the scale factors published by the Muon and EGamma POGs for the loosest available approved selections: loose muons and veto ID electrons. As we measure a scale factor relative to the loose SFs, the product of these two are the relevant factor.

We propose using these measured scale factors as a systematic on the signal yields. In the 4- and 5-layer bins these would be at most -6% and -3% systematics respectively, and the rest being well below -1%. Compared to the ISR weight systematic of roughly 9%, these will be largely inconsequential.

(7) For the strong production switch to the NNLO cross sections.

We have switched to NNLO(approx)+NNLL listed here

Concerning the the next steps, as discussed at the preapproval, once you address the comments on the $P(\text{veto})$ uncertainty and the impact of some loose ID on leptons used for veto i.e. (2) and (6), we can proceed to the unblinding of 2017 data.

Comments on the paper

Juliette Alimena (v0) May 22 [↗](#)

Show Details [▣](#) Hide Details [▣](#)

Thank you for the paper draft v0. I find it to be fairly complete (besides the 2018 data, which of course we know about). I have just a few very minor comments (which should not be taken as requirements for preapproval, although having them implemented by the time of the ARC review would be great).

I understand that you want to target PLB, and so have restricted the number of figures, but I think it's a little unfortunate that the only figures currently in the draft are limit plots. Maybe the easiest thing to do is to consider what figures you might also want to be made public in supplementary material. Think of what figures it would be nice to have when presenting this search at a conference. You might consider: Feynman diagrams, a sketch and/or event display of the signal in the detector (figure 3 or 4 in the AN?), 1 or 2D histograms of key variables (perhaps Figure 1 in the AN?), etc.

Figure 4 of the AN is a fairly classic one to present when discussing this, and as opposed to a Feynman diagram of AMSB is more signature-driven. But as you've pointed out we've possibly limited ourselves to much in our first draft, and will be considering what to add to it or what supplemental material is best.

The last 2 sentences of the paragraph starting at L60 and the one starting at L68 are nearly identical (see L64-67 and L72-75). Please consider writing the information once, but making clear for which signals it is applicable.

Some sentences in this section have been rewritten for increased clarity and without repetition.

You need to define PF as an abbreviation for particle-flow when it is first mentioned on L95.

The acronym is now included/defined on line 95.

Figure 1 is not mentioned in the text.

This has been added.

Although it could be nice to mention the new 2017+2018 results in the Summary section, I think this section should again mention the full Run 2 integrated luminosity and quote your results in that case.

A short paragraph has been added to the summary to re-mention the Run II combination and the mass exclusions from the combination. I've left the results of the 2017+2018 in however as it helps clarify what is new about this publication.

References 10, 21, 37: You have mistakenly written Collaboration Collaboration .

Fixed.

References 27, 39, 40: We list only the first page number, not the range (see the style guide).

Fixed.

Reference 34: \sqrt{s} mistakenly used parentheses instead of curly braces

Fixed.

Reference 37 seems unfinished.

Fixed.

Object reviews

(🔴 - tobedone 🟡 - inprogress 🟢 - OK) 🔴 Reset

- Click me to change my state from 'tobedone' to 'inprogress'
t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'tobedone'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'OK'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 [X]   

 jets

- Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-12-23 - 04:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'tobedone'
 t: 2019-12-23 - 04:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-12-23 - 04:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'OK'
 t: 2019-12-23 - 04:46 - Main.TWikiGuest, last state: tobedone
 [X]   

 MET

- Click me to change my state from 'inprogress' to 'OK'
 t: 2019-12-22 - 05:22 - Main.TWikiGuest, last state: inprogress
 Click me to change my state from 'inprogress' to 'tobedone'
 t: 2019-12-22 - 05:22 - Main.TWikiGuest, last state: inprogress
 Click me to change my state from 'inprogress' to 'inprogress'
 t: 2019-12-22 - 05:22 - Main.TWikiGuest, last state: inprogress
 Click me to change my state from 'inprogress' to 'OK'
 t: 2019-12-22 - 05:22 - Main.TWikiGuest, last state: inprogress
 [X]   

 electrons (high ET)

- Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'tobedone'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'OK'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 [X]   

 muons

- Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'tobedone'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'OK'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 [X]   

 taus

- Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'tobedone'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone
 Click me to change my state from 'tobedone' to 'inprogress'
 t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'OK'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 StatComm questionnaire

- Click me to change my state from 'tobedone' to 'inprogress'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'tobedone'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'inprogress'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

Click me to change my state from 'tobedone' to 'OK'

t: 2019-11-10 - 19:46 - Main.TWikiGuest, last state: tobedone

[X]   

 combine datacards check

Anshul Kapoor (Electron object review) [↗](#)

Show Details [▢](#) Hide Details [▾](#)

You use Ele35 trigger and your minimum pT cut for electrons is also 35 GeV. Isn't that cutting too close? Since you have a compound procedure for applying trigger scale factors, I am not sure if can judge whether this choice of 35 GeV is trigger-safe? Do let me know if I am not understanding how these electrons are used in this analysis.

We are a bit on the electron trigger turn-on, yes, but our signal selection is only based on MET+track paths. The scale factors you mention do not concern electrons, we instead use muons as proxies for the track with $p_T > 55$ to be on the track leg plateau. When we measure only the MET/track legs of our main HLT_MET105_IsoTrk50_v* path for TSG studies, we use an orthogonal method again with muons as proxies for the track.

In the background estimation where we do use electrons, we find no dependence on pt of any measured quantity, so any shaping of the low-pt distribution due this turn-on issue will not affect the estimate. As long as the electrons we select are of a high purity -- we use a tag-and-probe technique with opposite sign subtraction to ensure this -- then it does not matter to us if we've missed a small number of electrons close to 35 GeV.

Few unrelated typos I noticed while reading the AN,

Abstract: calorimter -> calorimeter

Line 1077: entireity -> entirety

Line 893 and 978 (two places): between -> between

And thank you for the careful reading for typos, when the AN is un-frozen after pre-approval we will correct these.

Benjamin Radburn-Smith (Muon object review) [↗](#)

Show Details [▢](#) Hide Details [▾](#)

I have gone through your AN2018_311_v6. I cannot see any issues from the muons side and therefore give a green light from us. However, for my interest, I cannot see why you choose muons with $p_T > 96$ GeV for your TnP muon reconstruction inefficiency study (Table 16). Could you please explain this pT choice?

"pt > 96 GeV" in Table 16 is an unfortunate typo, and was also noticed by another object review. The real value is "pt > 29 GeV", the same value used throughout for muon-related selections. Sorry for the confusion!

Klaas Padeken (Tau object review) [↗](#)

Show Details [▢](#) Hide Details [▾](#)

Hi, this is a very nice analysis and a thorough handling of the tau veto. I would just point out one thing of which you are probably aware, that requiring the tight isolation with cutbased and MVA discriminators does not have the highest efficiency. But since you are vetoing on taus, this means a higher background contamination. I also see that this allows the background studies, which use the tau trigger to use the same selection, which is needed. So as I said I just wanted to point this out.

But I did find one information missing in the AN. When you use the leptons as a veto, which pt threshold is used. There is an implicit cut, if you assume that the missing signal track is part of the tau, but it would be great to document this explicitly.

One other not tau related question and coming from my pixel perspective, do you see the effects of the stuck pixel TBMs and the dead DCDCs?

In the lepton veto, we do not explicitly make any pt or ID requirements and reject tracks that are near to **any** PF lepton. However indeed there is an implicit cut for the object to exist in MINIAOD in the first place. For taus we use "slimmedTaus" which requires pt > 18 GeV and passes "decayModeFindingNewDMS". We typically think in terms of "any lepton" but we see that we can mention the PAT slimming cuts explicitly in the AN, and will add this.

On the efficiency of our hadronic tau ID, an imperfect efficiency rather works in our favor to include the contamination you mention. We apply the electron and muon vetoes when studying the tau background, so we are not concerned about cross-contamination from the backgrounds we do present. This non-tau contamination could still fake our signal in the same way as real hadronic taus; our P(veto) estimate would include this type of event, and our tau control region we use to normalize the overall rate will as well. We just do not make any distinction about the tau purity and just call it the tau background.

For the pixel perspective, we definitely lose candidate tracks due to stuck TBMs and dead DCDCs due to our very tight hit pattern requirements, but these holes should appear fairly randomly and in small amounts compared to the whole. We don't see any dependence of efficiency on pileup for example, as a stuck TBM early in a fill could bias us to before it became stuck. We haven't examined the offline track occupancy on a fill-by-fill basis for our selection to see stuck TBMs being cleared, as another example, since that would just repeat the work DQM does. We also correct the missing hits distributions of our signal samples to the data, which would account for this effect in the simulation.

Chad Freer (MET object review) [↗](#)

Show Details [▢](#) Hide Details [▾](#)

This analysis looks very nice. I have a question regarding Figure 25. Can you explain the discrepancy between the total yields from the SingleElectron and the SingleMuon. The SingleElectron is emulating the same selection by removing the electron from the MET calculation so it is good that the distributions look similar, but i don't understand the difference in yields considering the trigger efficiency is high in both datasets. Is this coming from the EcaloDR<0.5 selection?

The largest difference between single-electron and single-muon selections in the search are due to the triggers available in 2017; electrons we require pt>35 GeV and muons we require pt>29 GeV as per the leptonic sections Tables 15 and 16 (we see now a typo in Table 16, it's 29 GeV not 96 GeV!). There are other differences in efficiency from ID/isolation/more but the PT is the largest.

Also since you are using Type-1 corrected MET can you out the version of JECs that you use?

For the JECs, we get them from the event setup in data using global tag 94X_dataRun2_ReReco_EOY17_v6, and in simulation using global tag 94X_mc2017_realistic_v15, with the AK4PFchs payload. According to CondDB this should retrieve the tag

"JetCorrectorParametersCollection_Fall17_17Nov2017BCDEF_V6_DATA_AK4PFchs" for data and "JetCorrectorParametersCollection_Fall17_17Nov2017_V8_MC_AK4PFchs" for simulation.

Eirini Tziaferi (Jet object review) [↗](#)

Show Details [▢](#) Hide Details [▢](#)

As the EXO Jet Object contact I have read through your document AN2018_311_v6. Things looks ok as concern jets, however since I could not find this information in the AN, I would like to ask you what is the JEC and which are the resolution scale factors you used.

In 2017 data we use the global tag 94X_dataRun2_ReReco_EOY17_v6 which retrieves from the event setup: JetCorrectorParametersCollection_Fall17_17Nov2017BCDEF_V6_DATA_AK4PFchs We just use "slimmedJets" from MINIAOD without additional manipulation. In 94X simulation we use the global tag 94X_mc2017_realistic_v15 which gets: "AK4PFchs" -- JetCorrectorParametersCollection_Fall17_17Nov2017_V8_MC_AK4PFchs "AK4PFchs_pt" -- JR_Fall17_25nsV1_1_MC_PtResolution_AK4PFchs "AK4PFchs" -- JR_Fall17_25nsV1_1_MC_SF_AK4PFchs.

Questions from subgroup conveners

Juliette Alimena email comments on AN v2, 19 Feb 2019

Show Details [▢](#) Hide Details [▢](#)

L51: The symbol `\ptmiss` is not defined.

We've now defined this at its first mention.

Including the PF reference at L147, and the anti-kt and fastjet user manual references at L195, would make your paper preparation that much easier. Also `PF` has not been defined on L147.

These references have been added, and PF is defined at its first mention.

Table 9 caption: You write that these are the triggers used for the background estimation and systematic uncertainty evaluation, but aren't at least the MET triggers also used to collect the data in your signal region?

You are right, and the caption has been made more clear in stating which data sets are used for which purposes.

L313, L328, L650: missing figure references

These lines have been modified to avoid the missing references.

Could you gain sensitivity if you re-optimized any of your selections? I imagine the last time they were tuned was for the previous version of the analysis. For example, if I look at Figure 14: could you bin more finely in E_{calo} and cut out more (W to lnu) background? Or if not on this variable, are there any other likely candidate cuts that can be returned?

As our background is data-driven, Figure 14 is mostly informational rather than a method to optimize; the

poor statistics makes this fairly difficult to use for that purpose. Optimizing Ecalo in data however, we see as not promising because it primarily removed electrons and is already very strict at 10 GeV, and changing by a few GeV would mostly just be cutting into pileup calorimeter deposits rather than real backgrounds. There's also the practical issue for us that changing the Ecalo cut would change our fiducial maps, which would have a large impact on our selection and the workflow of the analysis. To answer the more open-ended question of how we can overall increase our sensitivity, we look firstly at our large fake track estimate in the shorter nLayers bins -- especially considering ATLAS's 2015-6 results right now are still better at low lifetimes than even our 2015-6-7 combined result. We've found a series of well motivated cuts that increases the purity of fake tracks in our 3-layer transfer factor control regions, and leads to a much more reasonable estimation of this background. We plan to update you on the details of this on Friday.

Please add a legend to Figure 16.

Done.

You have some sort of typo in Section 7.2.1 (after L682) that makes most of the section unreadable.

There was an errant "\$" causing this, and that's now fixed.

Section 7.2.9: You say you don't yet have the track reconstruction efficiency study from the tracking POG for 2017. Would the study presented in <https://indico.cern.ch/event/768528/timetable/?view=standard#2-tracking-in-run2-performance> meet your needs? Hopefully it will soon be approved by the Tracking POG.

This is precisely what we need. For now we have added the relevant plot, but since it's not yet in CDS we just have a footnote to this presentation. Eventually we will promote that to a CDS reference.

Figures 42-45 inclusive are not described in the text.

A brief description of these figures has been added to the text.

All the appendixes except B.1 are empty? Appendix D would be particularly interesting to see.

They were indeed empty, but have now been given text.

I'm trying to compile what you still have left to do on the analysis (before unblinding). Is my list below complete? finish some systematic uncertainties implement trigger scale factors (L615)

The trigger scale factors are actually implemented in the signal yield, it seems we saw "XX%" and overlooked them thinking they were systematic uncertainties -- these numbers on L615 are now updated. The remaining analysis tasks are to investigate possible optimizations as mentioned above, and to finish the signal systematics.

Please add what you told me in your answer about the Ecalo selection to the note, namely that it primarily removes electrons and optimizing this criterion to increase the search sensitivity will be ineffective because it will be primarily be cutting into pileup calo deposits rather than real backgrounds. Adding this to the note will help anyone else who reads it and could have the same question that I did.

We've added an additional comment towards the end of Section 4.3 to clarify this.

Steven Lowette email comments on AN v2, 23 Feb 2019

Show Details ▢ Hide Details ▢

On another topic:

I'd like to understand how much time you think it would take to add 2018 data, keeping in mind this can happen during the review that we have started now (in particular since you are so data-driven). I don't see an obvious showstopper from reading the analysis note, but I remember there were technicalities involved, though forgot the details. Did you already request 2018 signal MC? As you know, the MT2 analysis is moving forward rapidly with a paper that will have full 2018 data. It's a different interpretation, but it's fishing in the same pond, and referees (and maybe internal in CMS too) may ask to compare.

To have arguments for or against adding 2018, it would also be necessary to understand the comparison with the latest ATLAS result.

As Yuri mentioned on Friday, it would probably take in the ballpark of 6 months to complete the analysis of 2018 data. We have 2018 signal MC samples already requested, although they have not yet begun on even 2017 samples. The technical issues affecting our timeline are a few:

We use a custom data format which essentially saves the generalTracks, and computes the ECalo for our tracks from AOD. There is a first step to the analysis, to calculate the fiducial maps for our selection. This adds another round of jobs over the single lepton datasets. We clean our lepton background estimates by pulling the small number of selected events from the RAW data tier. For 2017 data this took longer because the RAW datasets were not hosted anywhere we could submit jobs.

As for the comparison to ATLAS' latest public result, our updated expected exclusions shown on Friday compare powerfully to ATLAS' 2015-6 results that they've published. With even just 2017 data, we very slightly beat them at $c\tau = 1\text{cm}$, and are much, much more sensitive to higher lifetimes than that as we were before. The combination with 2015-6 results improves this slightly as well. So currently above 1cm we are better.

L67: nitpicking detail, but it reads as if just an extra layer was added, while the whole detector was replaced with all layers in different positions.

A good detail however; this is now noted in the text.

Fig 5, caption: "with the dependence ... as described in the text" -> where is this described?

The caption now correctly points to Section 6.3.

Fig 5: the turnon gets fully efficient in the plateau; so is there no inefficiency from the track leg vs some offline selection? Or is this factored out and described elsewhere?

The track leg's efficiency is not at 100%, which is a known issue for some time now. But Figure 5, and the corrections applied to signal described in Section 6.3, are calculated after requiring a $p_t > 55\text{ GeV}$ track and the OR of all the triggers. So any loss in efficiency due to the track leg has already been taken into account by the selection.

Figs 9, 10: what's the difference between the open and closed circles? I didn't see that described.

There are actually no closed circles, if you look closely the color of some bins are just close to the color of the circles. I've made this easier to see by making the circles green, which should stand out more.

L273: actually, this cannot be seen. I guess the figure has the majority of them already cut out.

At the time we only had this plot with the jet requirements already made, however this plot is now correct and the caption has been changed to reflect that. The edge you saw at jet $p_t 110\text{ GeV}$ from the requirement is no longer present.

Tab 17: what are the "lepton veto requirements"?

The JetMET-recommended jet ID is referred to as being tight with lepton vetoes, or "TightLepVeto". Over the tight ID it additionally cuts on the muon energy fraction and the charged EM energy fraction.

L298: is this primary vertex always matched to the main track of interest? It would be good to make this explicit.

It is. L298 is however describing a slightly different requirement on other tracks included in the track isolation sum, so we feel a more appropriate place to clarify this is in Table 18; "(w.r.t. primary vertex)" has been added to the two vertex requirements there.

%Green%L301: "We show several plots" -> where?

This paragraph referenced several figures we removed; the text is now cleaned up.

Tab 19: wouldn't it be useful to keep good tracks from b decays, where $dxy > 5\sigma$?

Since this selection applies to tracks included in the isolation energy sum, this would indeed ignore nearby displaced tracks from heavy flavor decays. However we do need some method of mitigating pileup, and this method actually includes slightly more than the standard PFisolation methods since those match vertices by reference instead of dxy . What should remove tracks too close to heavy flavor decays then is another cut, the ΔR between the candidate track and the nearest jet.

L313: broken \ref

The reference has been removed.

L328: Figs 15 and ?? -> 14 and 15

Fixed.

L337-338: you speak of "very few layers" but actually the biggest effect is for the long lifetimes. How does that add up?

I'm not sure I understand your comment about the "biggest effect" -- the version of the AN you had for this question did not have a legend for Figure 16, perhaps. But Figure 16 does show that for 10cm samples there are drops in efficiency for the p_T , number of valid pixel hits, and track isolation requirements, and these are indeed because not all of the shorter tracks are successfully reconstructed. We can remove the word "very" as this is a relative quality.

Fig 16: the 1jet $p_T > 110$ cut would be more logical much earlier in the chain, when you apply the other jet cuts.

This was an issue in our implementation of the MC-smear p_T cut, when replacing the regular non-smear $\text{jet.pt}()$ cut. This has been fixed now and the " \geq jets with smeaPt > 110" cut is in the correct place.

L350, Sec 5: do you expect no background from charged hadrons with little calo deposit because of dead ECAL cells etc?

As we veto tracks near dead ECAL cells (see Figure 8) we do not expect this background.

Fig 19: why does the control sample in muons have much more events than electrons; because of MetNoMu?

The principle reason is the trigger thresholds available in 2017. Due to these we need to require $p_T > 35$ GeV for electrons but only $p_T > 29$ GeV for muons. The muon selection efficiency was slightly higher in 2015-6 but

this additional difference increases the disparity in Figure 19.

Fig 19: why is the rejection so much stronger for the muons than for the electrons?

Recall the rejection is from vetoing *any* quality lepton, so it is an extremely loose definition of electrons/muons. This selection efficiency has been seen in 2015-6 as higher for muons than electrons, and we see the same in 2017 data.

Fig 22: I suggest you add the red boundaries also here.

Added.

Fig 25: can you add the muon and tau projections too (the right hand side plot I mean, the left one is not needed).

This figure now just has the three lepton flavors' projections, instead of the illustration concept.

L463: "N^l_ctrl": maybe this question should be obvious, but I tried to wrap my head around it and couldn't figure it out: it seems to me that you are missing in the estimate the leptons that did not pass the lepton iso/ID of the control sample selection, but that do lead to a track passing the selection. If it would be included but I missed it, then I could imagine the P_veto to be different for these leptons, indeed the veto to work worse and more background to pass?

From an operative perspective of just how to accomplish that, without the presence of a reconstructed (however poor quality) lepton, we can't infer the presence of a lepton in data to make any real distinction. The only place you can make a distinction is in SM background MC, which we do in the closure test.

From the perspective of overall strategy, the leptons you refer to should be present -- in a rate proportional to the total rate of single lepton events, and with probabilities to pass the search requirements equal to the probabilities listed on line 460. So the estimate in equation 11 (line 466) is designed to estimate these events which are really and truly the background to our search, where N^l_ctrl is not really a background since their tracks are from well-reconstructed leptons. Put one more way, N^l_ctrl is only necessary to normalize the background estimate to the expected number of single lepton events.

L511-513: even after reading, I still don't understand why this dxy sideband is needed. Maybe you can demonstrate how the non-displaced selection is not working?

We discussed this at length in our update on Friday. We were previously using the BasicSelection (the MET dataset) to derive a systematic on the fake estimate, so you needed to avoid using the well-vertexed tracks since that would be our signal region. As we now avoid using that due to potential signal contamination, we will need to avoid this peaking element in the 3-layer track sample at low fabs(dxy). The question was asked, why can we not just use the number of Z+track events while applying the fabs(dxy) cut, and our answer (which we've now confirmed as true) is that this would give precisely the same answer, but with very poor statistical uncertainty. So in short, using the 3-layer track samples greatly improves our statistical uncertainty while giving the same answer, and in order to use the 3-layer track sample we must use the dxy sideband to avoid this tracking issue only present in 3-layers. %ENDCOLO%

Tab 34: Is it a typo that the Z->ee P_fake for n=4 is a factor 10 off wrt the Z->mumu and the basic one? Isn't the point that these are expected to be the same?

This was indeed a typo. The Z->ee value for n=4 should have been $3.0 * 10^{-5}$. This is now fixed, but also changed since our fake estimate method has changed.

Tab 34: what are the MC truth values for these P_fake values - or do you lack statistics to estimate them?

We had not done a closure test of the fake estimate in MC, but now that one signal region is for 4-layer tracks, the statistics is better and we now are able to do this. We have added a section to the AN detailing this, and provide the P_{fake} values in MC truth you suggest.

Tab 35: averaging the 3 estimates seems a bit random. The "basic" category is the one that really matters for the analysis, if I understood correctly, so why not take that one as your real estimate, and use the other two for validation and systematics? Or if you want to avoid the statistical uncertainty, use $Z \rightarrow e\ell\ell$ with the basic as validation.

As mentioned on Friday, the basic selection is now very dangerous to use since 3-layers and 4-layers are not so different, and this presents signal contamination. Also with our updated fake estimate method, the systematic is now much lower and we do not have to use the average value to obtain a reasonable systematic.

Tab 35: also here, for information: what are the truth MC values?

See the above answer; these values are now in the AN.

Fig 28: can you have same backgrounds have same color, and order them the same between the plots?

Fixed.

L579: "excellent agreement": rather "better" or "good agreement"

Changed to "good".

Tab 37: I'm surprised you don't also apply the $p_T > 110 \text{ GeV}$ requirement, so that you are closer to your analysis phase space when you estimate the correction factors. The corrections are big, and I'm worried the inclusive phase space you consider may miss a dependence on the further analysis selections. Can you check whether adding the cut makes a difference?

If we require the jet requirements here, that would bias the MET distribution and we would not as successfully measure the trigger efficiencies. You are right that with the jet requirement, our samples populate a different region of Figure 30, but what we need in Table 37 is to measure the overall efficiency of the trigger requirement to derive scale factors well into the turn-on curve.

Fig 33: this correction is huge and it leaves me a little uneasy (but maybe I shouldn't be?). Was it not an option to have your signal generated with an extra jet in MadGraph?

Our choice of Pythia8 is mostly historical, but we also have not validated anything in MadGraph. As for the corrections, their size actually does not depend on the application (or not) of a gen-level ISR filter in our signal samples. The corrections are due to an apparently sizable difference in the hadronic recoil between Pythia8 and MadGraph within the same tune (CP5). So applying a gen-level ISR filter would only increase our signal efficiency/statistics in regards to what is generated in the samples, but those filtered events would still populate the same, higher recoil phase space where the disagreement between Py8/MG seems to be larger. In the end we feel the gen-level ISR filter is not necessary, as we are content with the statistics selected from these signal samples already.

L647: the 28% is smaller than the stat uncertainty. Does the stat uncertainty still go separate as well?

With our updated fake estimate method presented on Friday, these values all decrease. But for your question we consider the statistical and systematic uncertainties separately and completely uncorrelated when put into the limit datacards.

L650: ?? -> 26

Fixed.

L652: "100%": this demonstrates the randomness of this systematic. You can always change the dxy range until you run out of statistics, and take a big hit in the systematics. I think this 100% creates a fake "we're conservative" feeling, while I'd argue it's not a real systematic: all the values in Fig 26 for same color are compatible with one another, so why assign an uncertainty at all? I think what does require a thought-through systematic, on the other hand, is the use of the transfer factor measured in ==3 tracker layer tracks. There's an assumption that goes in here that the measured value is applicable in your phase space, but there is no validation or systematic to deal with that.

Since our fake estimate has changed and has different considerations to make, this 100% is reduced. As well rather than progressively narrow the sideband which as you point out decreases statistics, we now do the opposite and progressively widen it to test for this dependence. We have also decided to move away from using the ==3 layer tracks, as the presence of this central bias peak makes comparison with ==4 layers very difficult. Using ==4 layer tracks for the transfer factor for ==5 and >=6 numerically does not change the estimate in any case, and improves this particular plot.%ENDCOLOR

L671: my memory tells me we expect on average about 3GeV deposit for a muon traversing the calorimeter, so it may not always be completely negligible. But ok, smaller than the electron case, agreed.

In the future this is something to investigate, but for now we exclude it as it will be a very very small uncertainty.

p67, bottom: math mode mess

Fixed.

L718-719: how much? Please mention here.

These values hadn't yet been calculated. They are now and are quoted in the text. They are below 1.2% for the ≥ 6 layer bin and less than 0.003% in the =4 and =5 layer bin.

L723: I didn't really understand the procedure. What is this 1 sigma?

This is 1 sigma of their statistical uncertainties, so the error bars in Figure 30. The text has been changed to say this more clearly.

L731: that must be a huge factor, and I don't think it makes sense to take the change as a systematic. The pythia description of the ISR is just not good at such high pT, so why base a systematic on it? I'm sure it's overly conservative.

The AN actually was incorrect here. Rather than removing the weights, the systematic here is taken by fluctuating the weights shown in Figs 32 and 33 up and down by their statistical uncertainty, and comparing the change in signal yields. So the uncertainty is not on applying or not applying this, but in the sample sizes used to derive the weights. This does end up as a roughly 7% uncertainty.

Table 45: the n=4 region has a large number of expected background events, with a large systematic. In a simple $S/\sigma(B)$ sensitivity metric, you get $S/\sqrt{B+DB^2} = S/\sqrt{191+6400}$. Thus, the systematic completely dominates the sensitivity. In general, underlined by my comments above, I have the impression it's hard to do better than guesstimating that systematic - but the sensitivity at low displacement crucially depends on it (shown also in Fig. 42). It's an uncomfortable situation. Better would be if the background could be significantly further suppressed. And if I look at Fig 29, it seems to me there is an obvious kinematic variable you can use further: the track pT. Although I'm not sure how that looks like for fake tracks (can you add a plot?). Can you further suppress the background this way, or in another way, and put similar or better

sensitivity on more solid ground for small displacements?

As mentioned on Friday we have also noticed this and have made a successful effort to reducing the overall background as you say. So perhaps this question is not so needed with this updated estimate. On the possibility of using track pt as a discriminating variable, this would be certainly be effective for some chargino masses, but we consider a wide range where the track kinematics may not benefit from this in some regions. We also try to keep the search quite general, and make an effort to remind readers that the AMSB limits we show are provided only as a benchmark and not a specific motivation for analysis optimization.

-- BrianFrancis - 2019-03-07

This topic: Main > TmpEXO19010

Topic revision: r4 - 2019-08-27 - BrianFrancis



Copyright &© 2008-2020 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.
Ideas, requests, problems regarding TWiki? Send feedback