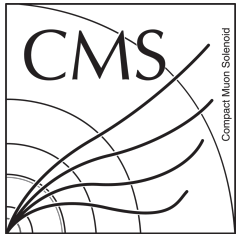


UCSD

Limits update

Anthony Vizcaíno Aportela
05/03/23

A preamble

I was having a bit of an organizing and naming problem, with codes and samples from different periods. To make this a little easier, I tried to standardize the naming scheme a bit for myself.

Let's name things by the "era" they were produced and used. They correspond to the dates of the talks Aram and I gave in the EXO meeting.

- `oct_2022`
- `jan_2023`

So the cutflow codes, the signal samples, and the background, each have two(ish) versions corresponding to those dates given above. I also define a **third date** for the background (`apr_2023`) since later on Aram produced an updated version of the bkg using his bigNtupler. **Tables on next page**

Some maybe helpful tables

	% of b-parking run on	nEvents
oct_2022 bkg	2%	2.57e+07
jan_2023 bkg	???	4.74e+06
apr_2023 bkg	2%(?)	2.57e+07

	~nEvents
oct_2022 sig	5e+05
jan_2023 sig	1.5e+06

	oct_2022 cutflow code	jan_2023 cutflow code
oct_2022 sig + oct_2022 bkg	✅	❌
jan_2023 sig + jan_2023 bkg	✅	✅
jan_2023 sig + oct_2022 bkg	✅	✅
jan_2023 sig + apr_2023 bkg	✅	✅

The BR limit calculation

l = luminosity
σ = cross section
$\epsilon_\mu \cdot \epsilon_{LLP}$ = muon and LLP efficiency

$$BR \approx 2\frac{\sqrt{N_{bkg}}}{N_{sig}}$$

$$N_{bkg} \approx N_{Bparking}\frac{N_{sample\_bkg\_cut}}{N_{sample\_bkg}}$$

#events in pbarking

bkg events in cutflow after and before cuts

$$N_{sig}(c\tau) \approx \epsilon_\mu \cdot \epsilon_{LLP} \cdot l \cdot \sigma \cdot \frac{N'_{sample\_sig\_cut}(c\tau)}{N_{sample\_sig}}$$

sig events in cutflow after and before cuts, reweighted to take SF into account and also to extrapolate to other ctaus

$$N'_{sample\_sig\_cut}(c\tau) = \sum \epsilon_{\mu\_sf}\frac{c\tau_{old}}{c\tau}\exp\left[c\tau_{LLP}\left(\frac{1}{c\tau_{old}} - \frac{1}{c\tau}\right)\right]$$

Reweighting digression

Each event has single LLP and therefor a single LLP lifetime, $c\tau_{LLP}$. Each event also has a muon scale factor, ϵ_{μ_sf}. Then we sum over the events to reweigh the samples accordingly:

$$N'_{sample\_sig\_cut}(c\tau) = \sum \epsilon_{\mu\_sf} \frac{c\tau_{old}}{c\tau} \exp\left[c\tau_{LLP}\left(\frac{1}{c\tau_{old}} - \frac{1}{c\tau}\right)\right]$$

If the sample in question was generated with ctau=100cm, then that is what is put in place of $c\tau_{old}$

Mistake #1

I did this:

$$N_{bkg} \approx \epsilon_{\%Bparking} N_{Bparking} \frac{N_{sample\_bkg\_cut}}{N_{sample\_bkg}}$$

Instead of this:

$$N_{bkg} \approx N_{Bparking} \frac{N_{sample\_bkg\_cut}}{N_{sample\_bkg}}$$

Thinking that $\epsilon_{\%Bparking}$ was the percentage of the b-parking dataset run through to generate the background sample.

Mistake #1 cont.

Last Friday I was told that

$$\epsilon_{\%Bparking} \sim N_{sample\_bkg\_cut} / N_{sample\_bkg}$$

and therefore I was double accounting for that correction.

With this extra factor, the BR limit is of course **under-estimated**.

Mistake #2

I did not properly take the LLP gen filter into account when calculating the BR limits with Aram's jan2023 samples, i.e. I was missing the ϵ_{LLP} correction in the signal yield calculation.

$$N_{sig}(c\tau) \approx \epsilon_{\mu} \cdot \boxed{\epsilon_{LLP}} \cdot l \cdot \sigma \cdot \frac{N'_{sample\_sig\_cut}(c\tau)}{N_{sample\_sig}}$$

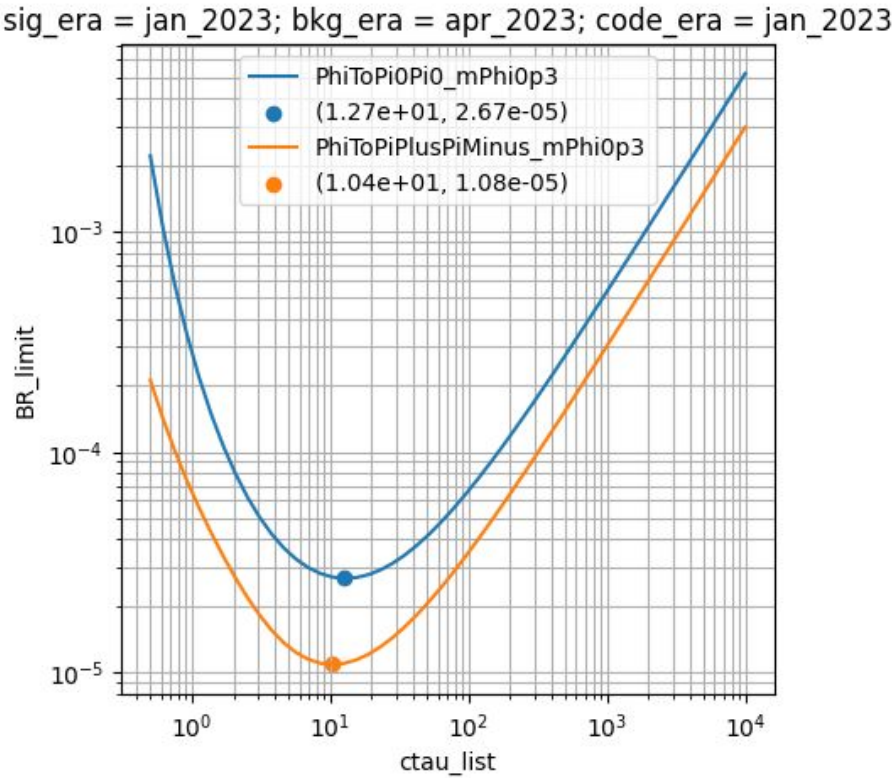
Without this correction, the BR limit is of course **under-estimated**.

Mistake #3

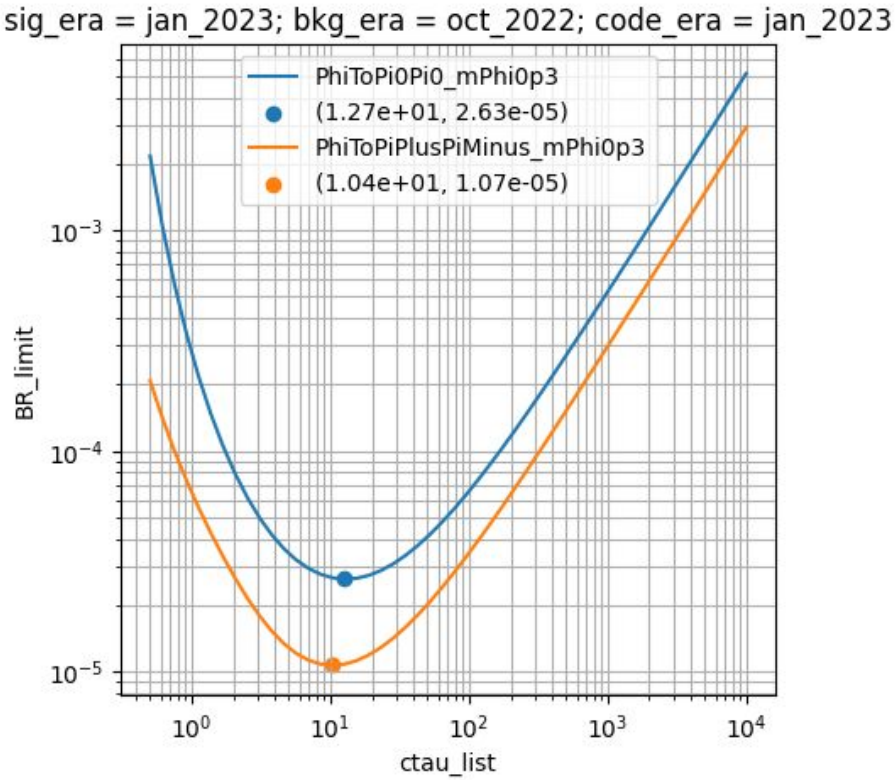
There was a misunderstanding in the details of `jan 2023` bkg sample which only had 4.74e+06 events.
I don’t quite understand what is “wrong” if anything with this bkg, but Aram ended up reproducing the bkg (`apr_2023`) to have 2.57e+07, ie to have the same number of events as Christina’s `oct_2022` bkg. I think in principle `oct_2022` and `apr_2023` should be the same statistically, but I’m not sure.

I think it would be really helpful to document these differences when producing these sample in case anyone else has to go through this kind of archeology again…

Finally, with all these changes taken into account

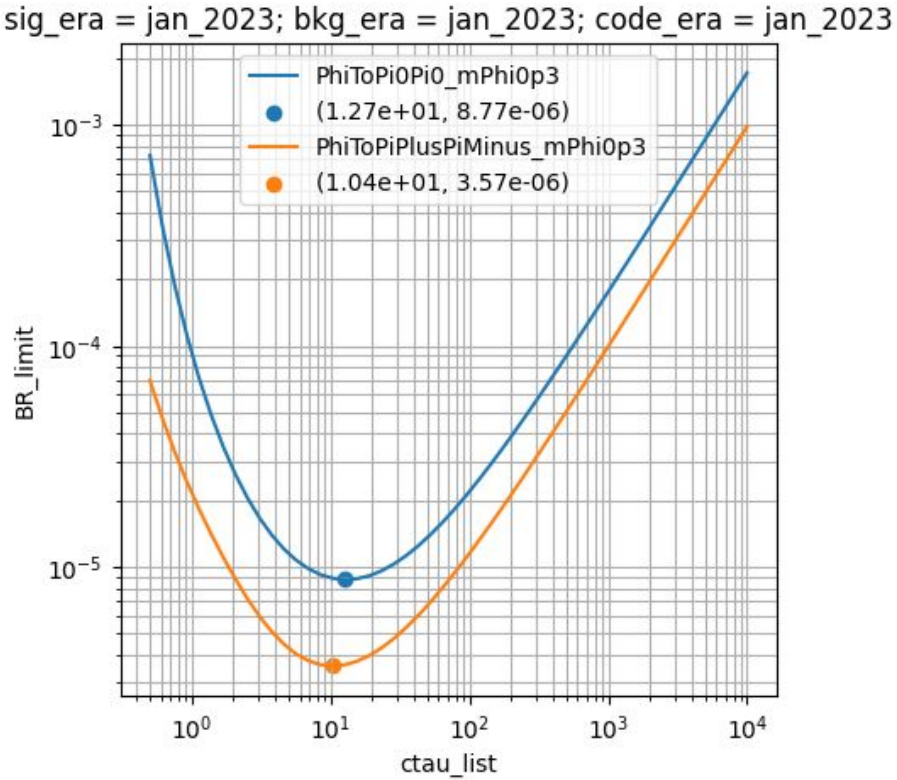


Same thing but with bkg_era=oct_2022



Same thing but with bkg_era=jan_2023

This drives the point home that something about the `jan_2023` bkg is different, driving the BR limit down by a factor of 3.



I don't actually understand how bkg and sig are made from the analyzer. Makes me think that `jan_2023` bkg produced was somehow too small for the sig of that era

Conclusion

This was somewhat difficult to do mostly because my own code and process was not well archived and documented. What I learned most from this was that I didn’t take enough lab notes.

Something else that also could make something like this easier in the future is to have an archive of the samples with **a short readme detailing where the samples come from, what analyzer was used, what variables were skimmed, produced, removed, cuts, etc**