Physicist’s Summary PhystatDM 2019

Hugh Lippincott (Fermilab)

Aug 2, 2019

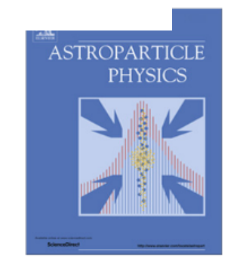
Summary talk

My first "PHYSTAT" – Durham 2002

COVERAGE OF CONFIDENCE INTERVALS FOR POISSON STATISTICS IN PRESENCE OF SYSTEMATIC UNCERTAINTIES

J. Conrad, O. Botner, A. Hallgren, C. P. de los Heros
High Energy Physics Division, Uppsala University, Sweden.

ELSEVIER
Contents lists available at ScienceDirect
Astroparticle Physics
journal homepage: www.elsevier.com/locate/astropart



Review

Statistical issues in astrophysical searches for particle dark matter

CrossMark

Jan Conrad

Oskar Klein Centre for Cosmoparticle Physics, Physics Department, Stockholm University, Albanova University Centre, SE-10691 Stockholm and Imperial College London, London SW7 2AZ, UK

Limits and confidence intervals in the presence of nuisance parameters

Wolfgang A. Rolke[a,*], Angel M. López[b], Jan Conrad[c]

[a]*Department of Mathematics, University of Puerto Rico - Mayagüez, P.O. Box 5959, Mayagüez, PR 00681, USA*
[b]*Department of Physics, University of Puerto Rico - Mayagüez, Mayagüez, PR 00681, USA*
[c]*PH-Department, CERN, CH-1211, Geneva 23, Switzerland*

Received 7 February 2005; received in revised form 13 May 2005; accepted 18 May 2005
Available online 11 July 2005

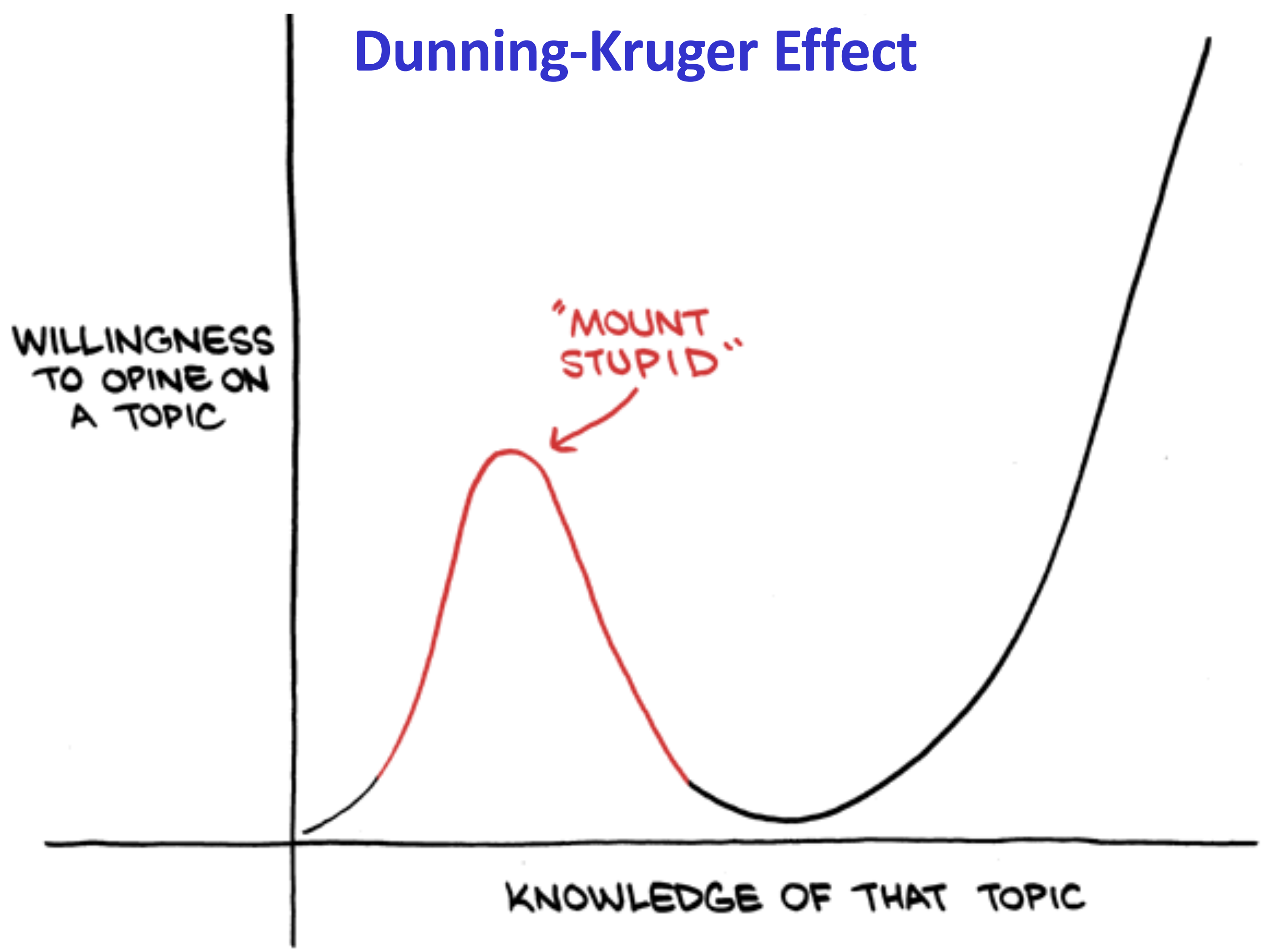
Statisticians summary — *Alessandra Brazzale*

Stockholm university — 15:40 - 16:10

Physicists Summary — *Hugh Lippincott*

Stockholm university — 16:10 - 16:40

- **Hopefully relevant and correct, but likely not at the same time**



Carl Larsson

Statisticians summary	*Alessandra Brazzale*
Stockholm university	15:40 - 16:10
Physicists Summary	*Hugh Lippincott*
Stockholm university	16:10 - 16:40

- **Hopefully relevant and correct, but likely not at the same time**

L. Lyons, reporting on the Durham meeting

a specified range. Instead, a frequentist would be prepared to use probabilities only for obtaining different experimental results, for any particular value of the parameter of interest. The frequentist restricts himself to the probability of data, given the value of the parameter, while the Bayesian also discusses the probability of parameter values, given the data. Arguments about the relative merits of the two approaches tend to be vigorous.

$P(A|B) \neq P(B|A)$

Remind Lab or University media contact person that:

Prob[data, given H0] is very small

does **not** imply that

Prob[H0, given data] is also very small.

e.g. Prob{data | speed of $\nu \leq c$}= very small

does **not** imply

Prob{speed of $\nu \leq c$ | data} = very small

or Prob{speed of $\nu > c$ | data} ~ 1

Everyday situation:

p(eat bread|murderer) ~ 99%

p(murderer|eat bread) ~ 10^{-6}

From Louis' intro talk on Wednesday

P(A|B) ≠ P(B|A)

Remind Lab [And Self] or University media contact person that:

Prob[data, given H0] is very small

does **not** imply that

Prob[H0, given data] is also very small.

e.g. Prob{data | speed of $\nu \leq c$}= very small

does **not** imply

Prob{speed of $\nu \leq c$ | data} = very small

or Prob{speed of $\nu > c$ | data} ~ 1

Everyday situation:

p(eat bread|murderer) ~ 99%

p(murderer|eat bread) ~ 10^{-6}

From Louis' intro talk on Wednesday

P(A|B) ≠ P(B|A)

Reminder:

annalen der physik
www.ann-phys.org

del, which

s the elec

(10)

(11)

ion of 5th

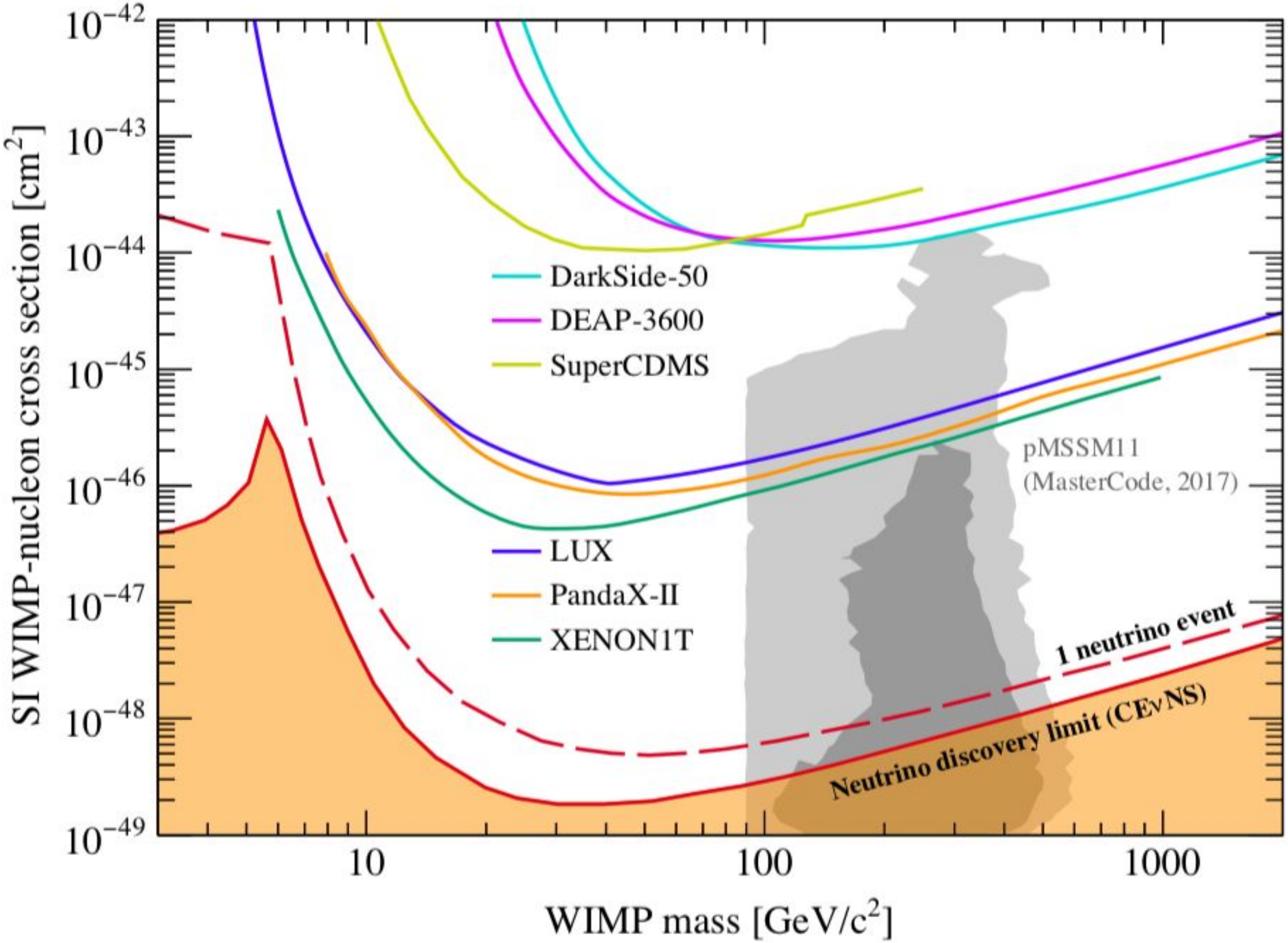
o shows a

commonly used in low-energy experimental tests of fundamental physics, an event's probability is defined by the frequency of its occurrence in a large number of trials. The high-energy physics community, however, has largely moved away from frequentist statistics when deriving upper limits and uses methods known as CL_s or Power Constraint Limit (PCL)[42,43] instead.

In the context of our measurement, suppose there is a dark photon theory predicting one value for the electron's magnetic-moment anomaly $\mu = \delta a_{e,\text{ th}}$, while we observe another value $\hat{\mu} = \delta a_{e,\text{ obs}}$, where $\delta a_{e,\text{ obs}} < \delta a_{e,\text{ th}}$. Under the frequentist paradigm, we can calculate a p-value p_μ for the theory to be compatible with the experiment (assuming a normal distribution of the data)

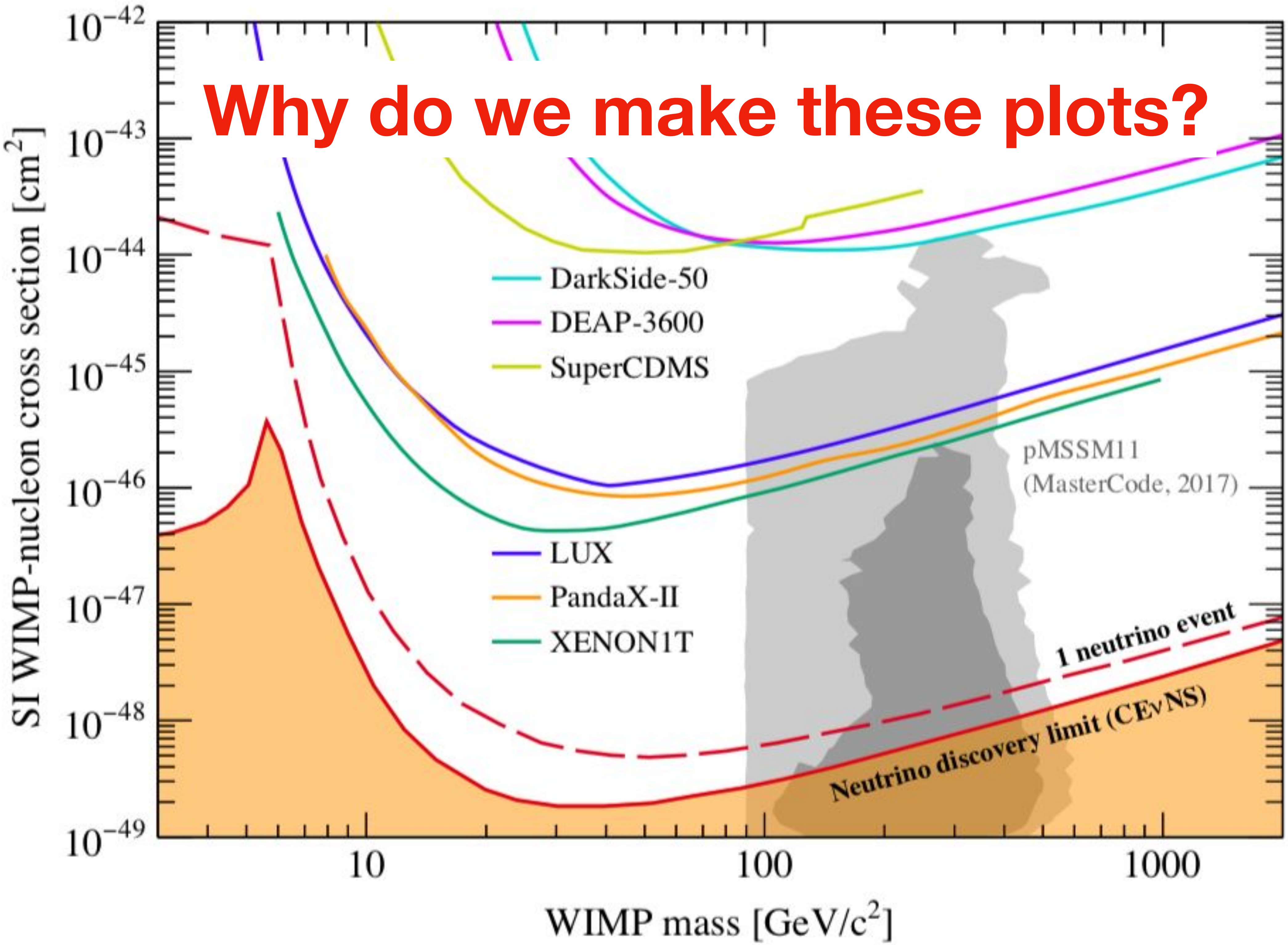
$$p_\mu = P(\hat{\mu} < \delta a_{e,\text{ obs}} | \mu = \delta a_{e,\text{ th}}) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\delta a_{e,\text{ obs}}} d\hat{\mu}\, e^{-\frac{(\hat{\mu}-\mu)^2}{2\sigma^2}}$$

K. Mora



- **Our primary product sets us up for misinterpretation**
- **90% of people seeing our talks assume that models above the lines are excluded**

Why do we make these plots?

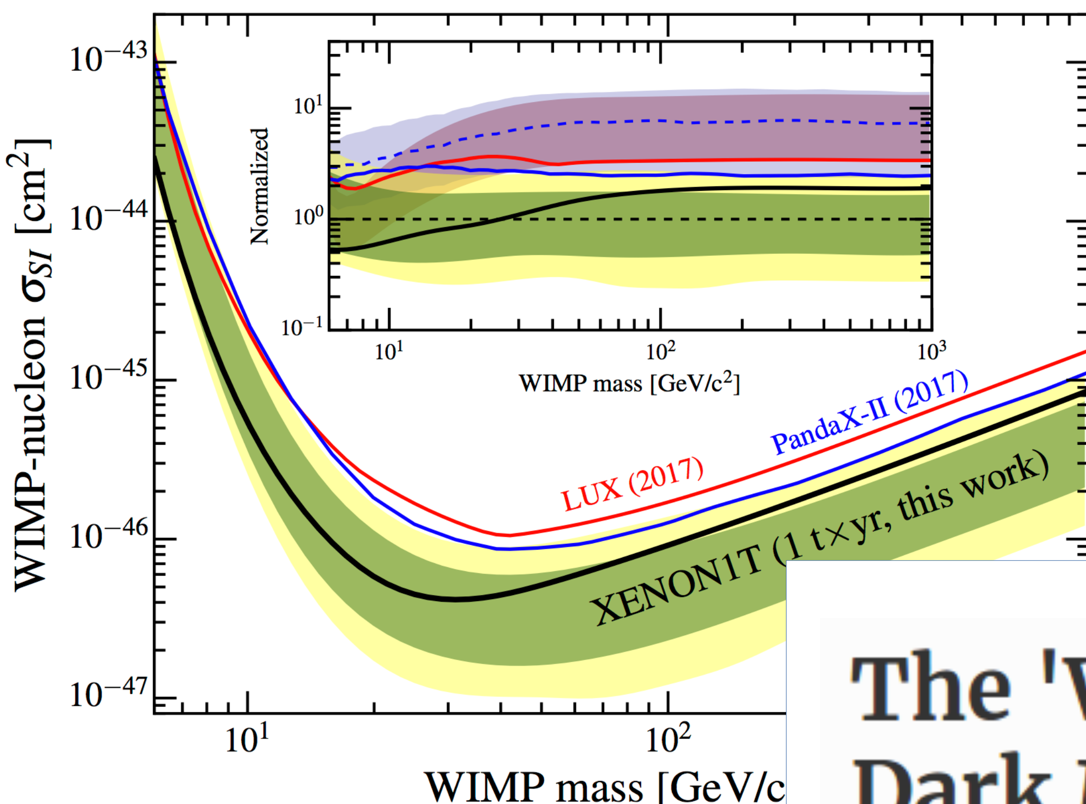


- Our primary product sets us up for misinterpretation
- 90% of people seeing our talks assume that models above the lines are excluded

The idealized case

- **To say something about particle physics and guide what we do next**

Where are the WIMPs?



So far we have not observed any conclusive evidence for WIMPs in direct, indirect or collider experiments

Frankfurter Allgemeine

Dunkle Materie

Wo sind die Wimps?

The 'WIMP Miracle' Hope For Dark Matter Is Dead

Forbes

SCIENTIFIC AMERICAN.

In the Dark about Dark Matter

Recent disappointments have physicists looking beyond WIMPs for dark matter particles

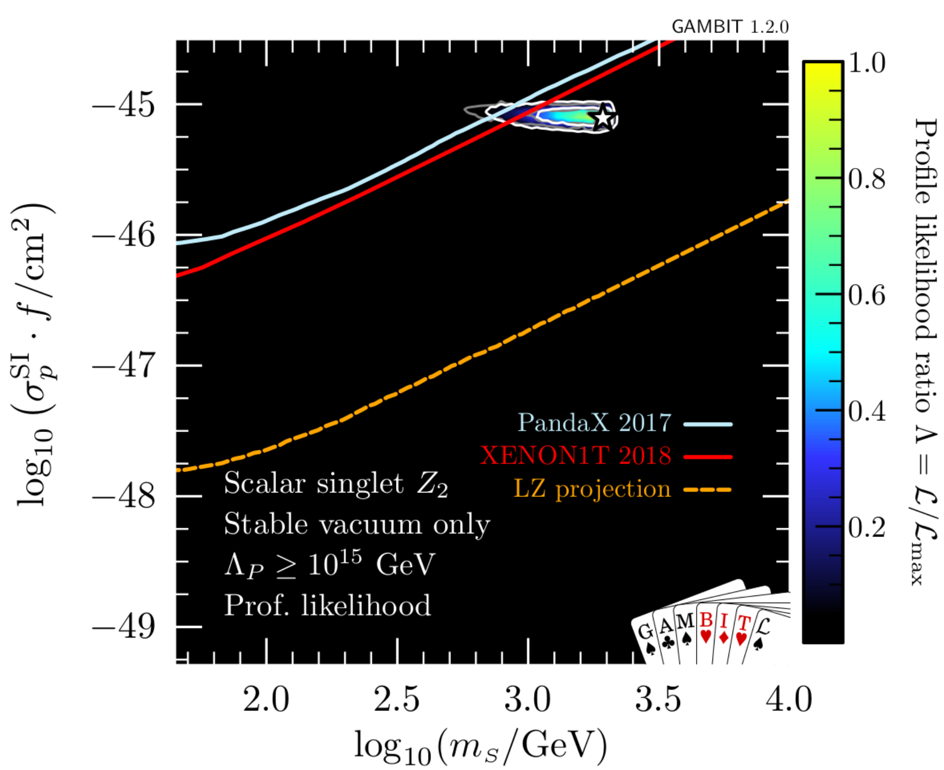
F. Kahlhoefer

The idealized case

- To say something about particle physics and guide what we do next
- We need an appropriate statistical treatment to do this properly - P(A|B) is not P(B|A)

Scalar Higgs portals

- Higgs portal coupling is dimensionless → model **fully renormalisable**
- Scalar Higgs portal models remain valid and perturbative up to the Planck scale (at least in some regions of parameter space)



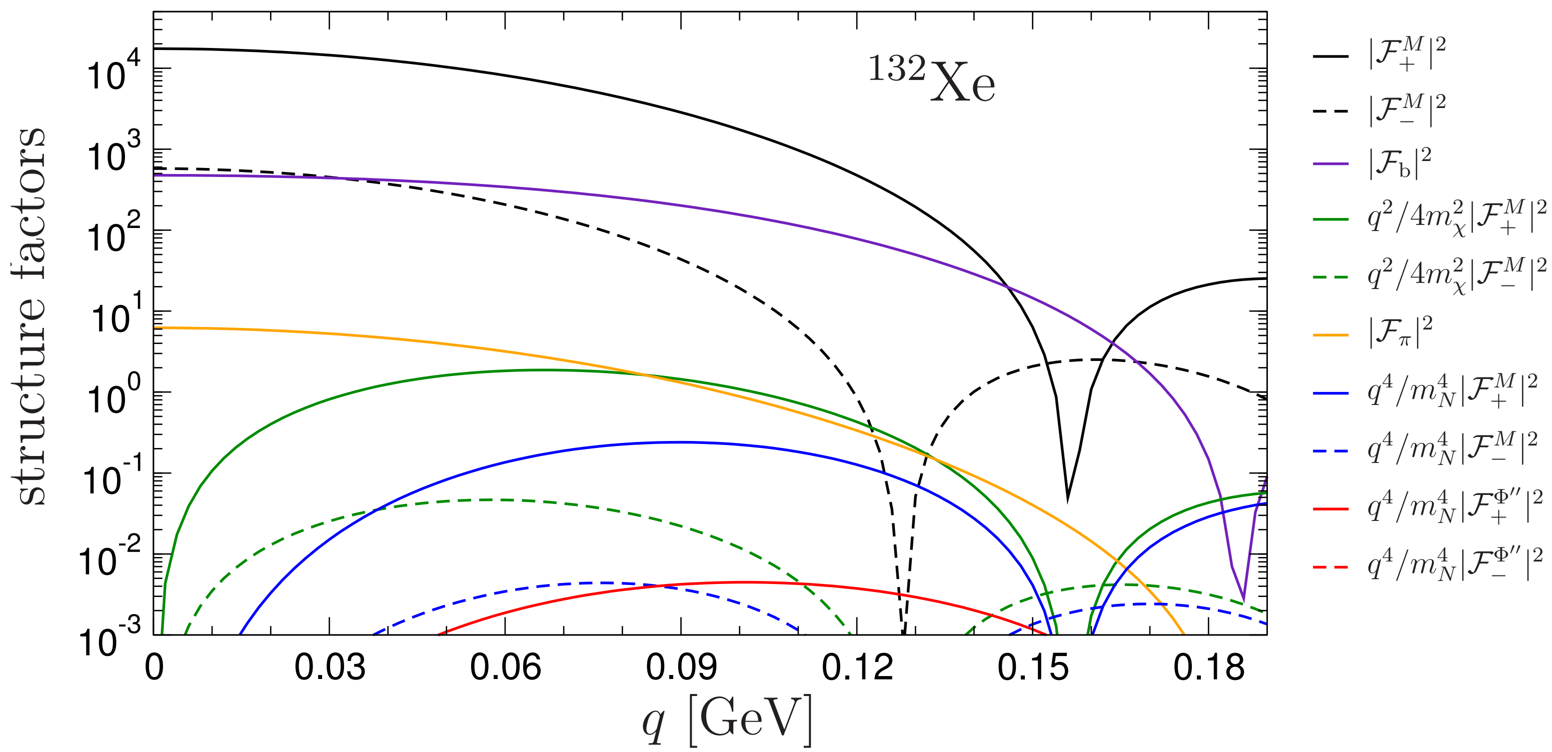
- The scalar DM particle prevents the Higgs self-coupling from running to negative values and thus **stabilises the electroweak vacuum**
- Small remaining parameter region where scalar singlets **can be all of DM**, evade all experimental constraints and stabilise the vacuum

F. Kahlhoefer

The idealized case

- **To say something about particle physics and guide what we do next**
- **We need to appropriately account for stuff going on in the background of our standard plot**

Full set of coherent contributions



$$\frac{\mathrm{d}\sigma_{\chi\mathcal{N}}}{\mathrm{d}q^2} = \frac{1}{4\pi v^2}\left|\sum_{I=\pm}\left(c_I^M - \frac{q^2}{m_N^2}\dot{c}_I^M\right)\mathcal{F}_I^M(q^2) + c_\pi\mathcal{F}_\pi(q^2) + c_b\mathcal{F}_b(q^2) + \frac{q^2}{2m_N^2}\sum_{I=\pm} c_I^{\Phi''}\mathcal{F}_I^{\Phi''}(q^2)\right|^2$$

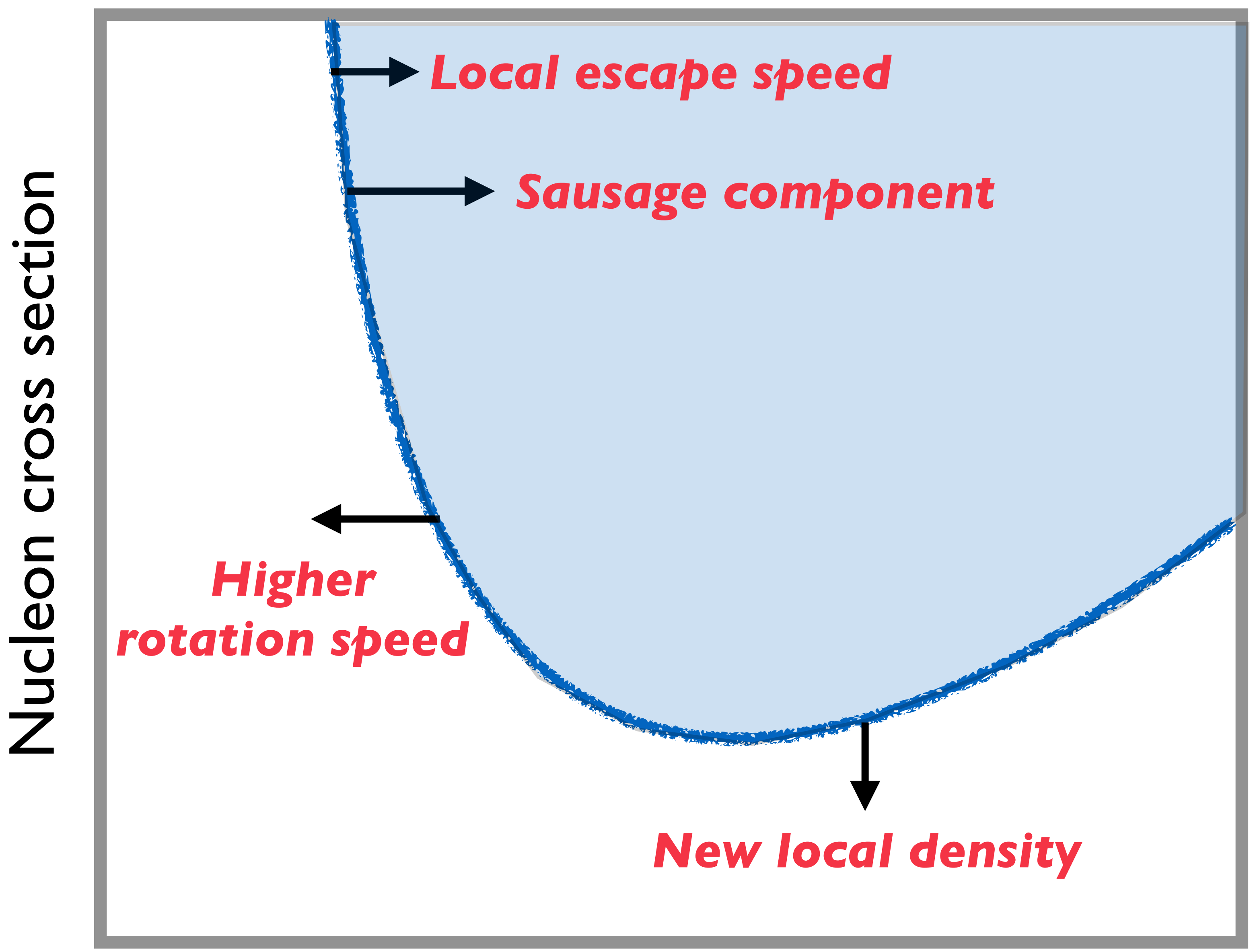
$$+ \frac{1}{4\pi v^2}\sum_{i=5,8,11}\left|\sum_{I=\pm}\xi_i(q, v_T^\perp)c_I^{M,i}\mathcal{F}_I^M(q^2)\right|^2$$

$$+ \frac{1}{v^2(2J+1)}\left(|a_+|^2 S_{00}(q^2) + \mathrm{Re}\,(a_+a_-^*)S_{01}(q^2) + |a_-|^2 S_{11}(q^2)\right)$$

M. Hoferichter

The idealized case

- To say something about particle physics and guide what we do next
- We need to appropriately account for stuff going on in the background of our standard plot

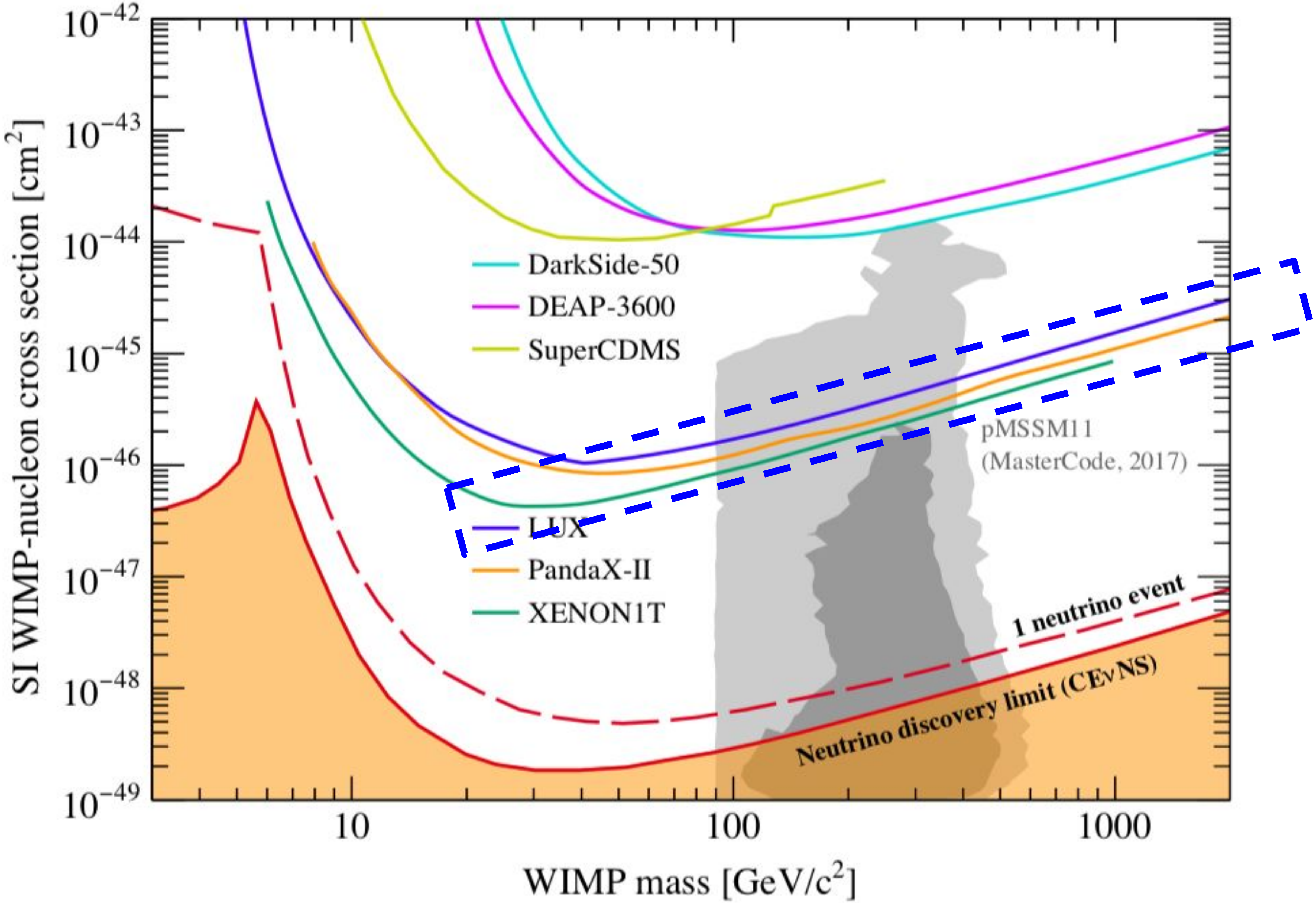


C. McCabe

The realist case

- We need to be able to compare experiments
  - Particularly important in the absence of discovery
- Competition is fierce
- Funding is limited
- Professional prestige/success on the line

My experiment is better than yours!



- **In N (1-2?, 20) years, no one will care**
- **But at the time, we care desperately**

Statistical treatments

- Experiments are “fixed” - once it is built, there are limited opportunities for upgrading
- Statistical methods can provide a way to get every last piece of information out of the data
  - This is a good thing!
- And in some cases, being smarter can let you win

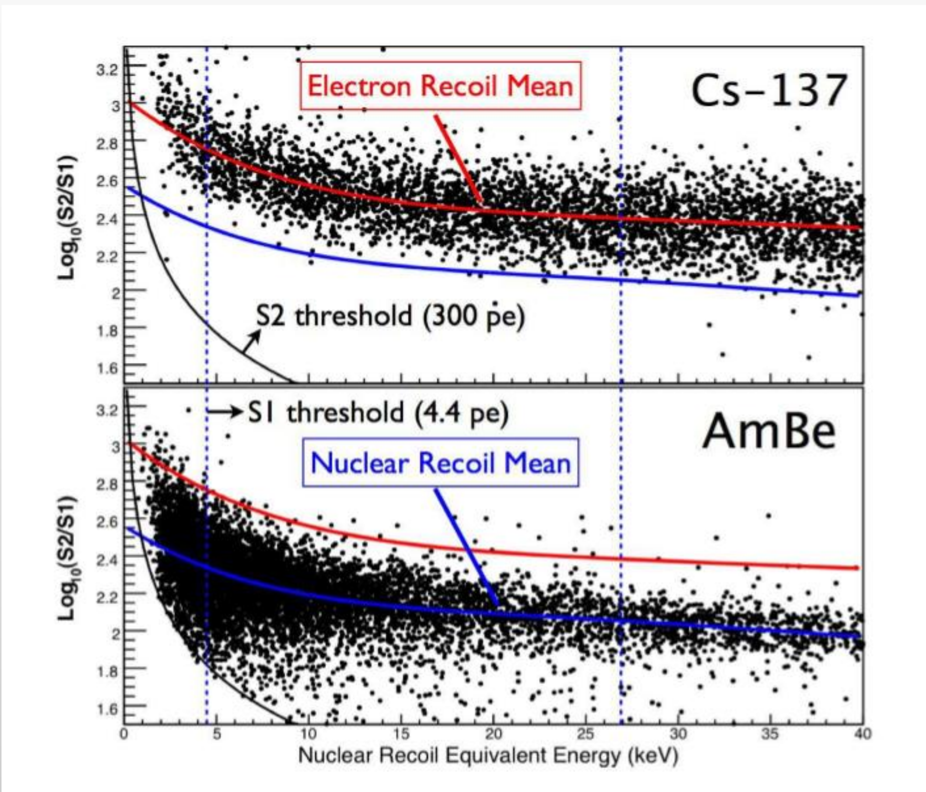
Region Of Interest

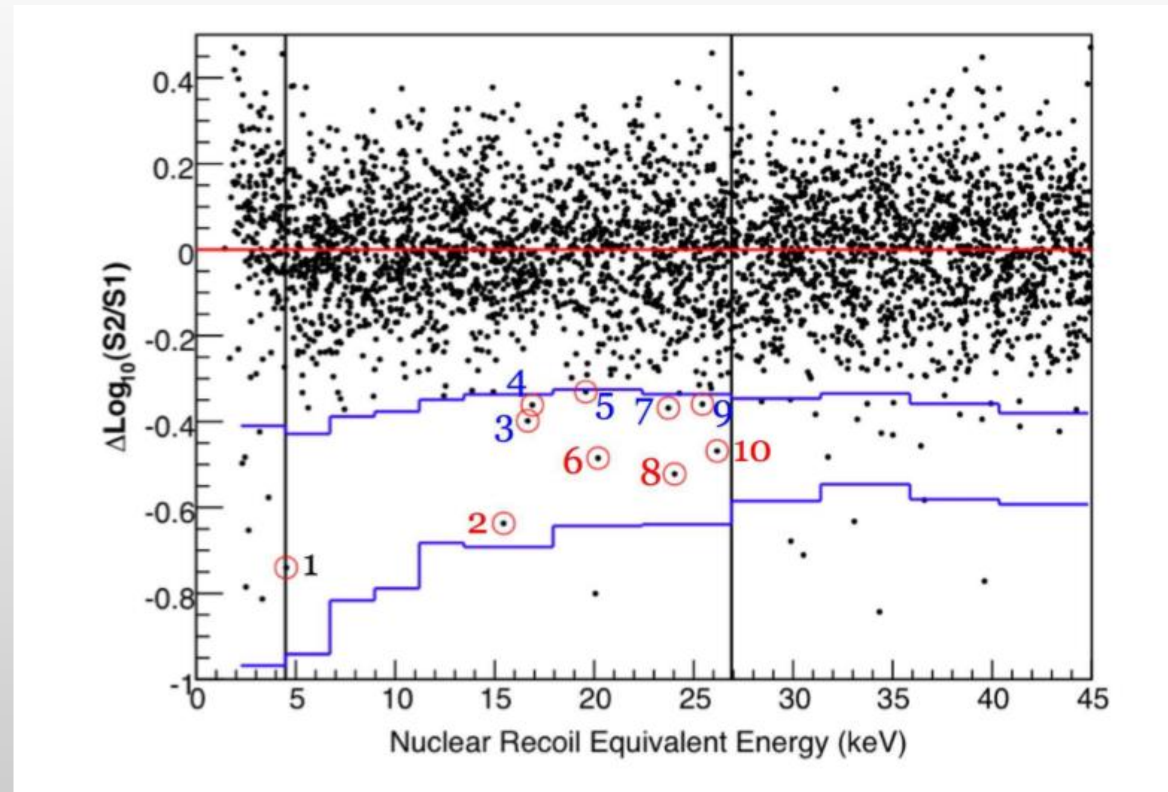
PRL 100, 021303 (2008) PHYSICAL REVIEW LETTERS

First Results from the XENON10 Dark Matter Experiment at the Gran Sasso National Laboratory

J. Angle,[1,2] E. Aprile,[3,*] F. Arneodo,[4] L. Baudis,[2] A. Bernstein,[5] A. Bolozdynya,[6] P. Brusov,[6] L. C.
C. E. Dahl,[6,8] L. DeViveiros,[9] A. D. Ferella,[2,4] L. M. P. Fernandes,[7] S. Fiorucci,[9] R. J. Gaitskell,[9] K.
R. Gomez,[10] R. Hasty,[11] L. Kastens,[11] J. Kwong,[6,8] J. A. M. Lopes,[7] N. Madden,[5] A. Manalaysay,[1,2]
D. N. McKinsey,[11] M. E. Monzani,[3] K. Ni,[11] U. Oberlack,[10] J. Orboeck,[2] G. Plante,[3] R. Santorelli,[3] J. M.
P. Shagin,[10] T. Shutt,[6] P. Sorensen,[9] S. Schulte,[2] C. Winant,[5] and M. Yamashita[3]

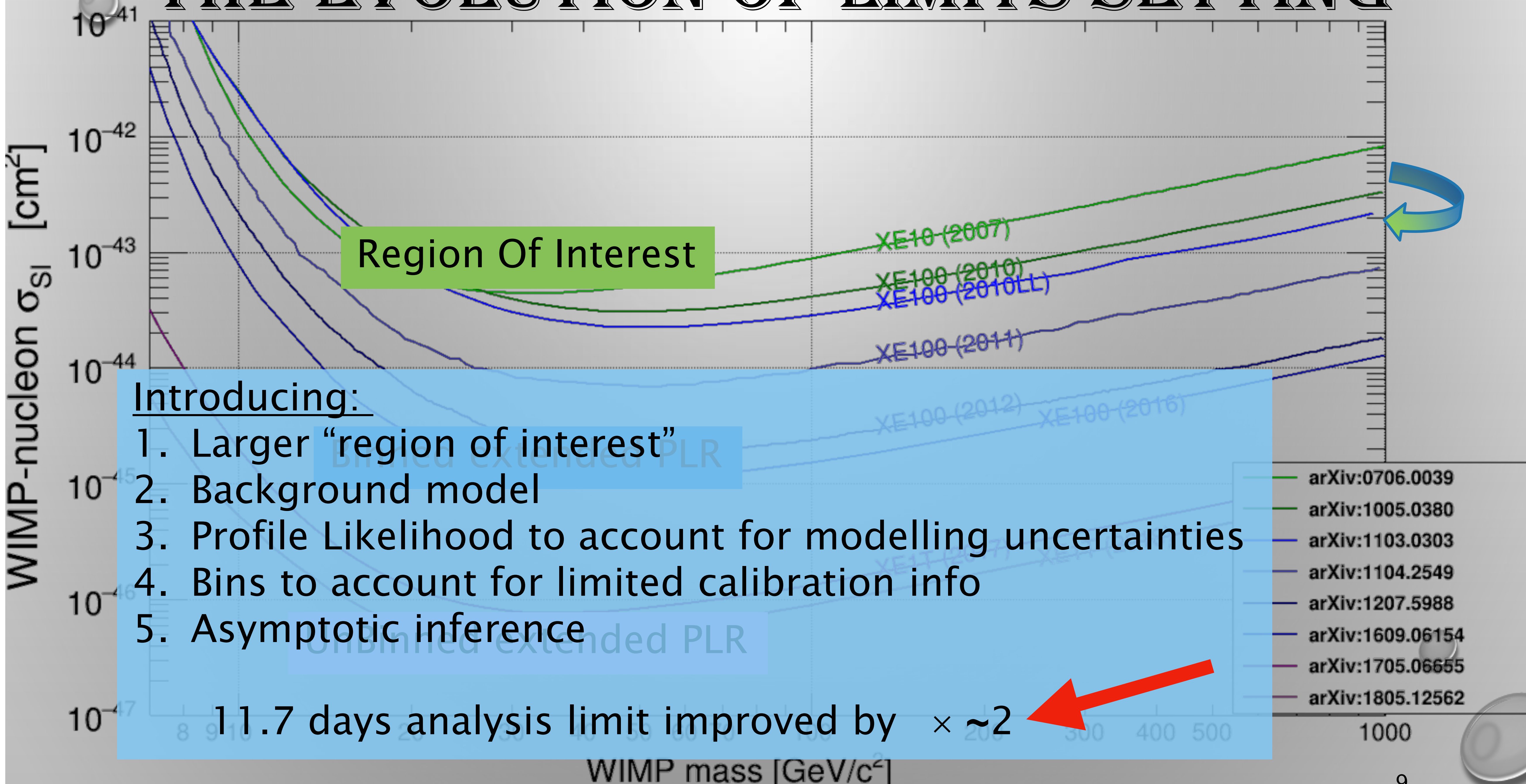
XENON10 2005-2007





"However, the uncertainty of the estimated number of leakage events for each energy bin in the analysis of the WIMP search data is currently ***limited by available calibration statistics****. Based on the analysis of multiple scatter events, no neutron induced recoil event is expected in the single scatter WIMP-search data set. To set conservative limits on WIMP-nucleon spin-independent cross section, we consider all ten observed events,* ***with no background subtraction.****"*

XENON
THE EVOLUTION OF LIMITS SETTING



<u>Introducing:</u>

1. Larger "region of interest"
2. Background model
3. Profile Likelihood to account for modelling uncertainties
4. Bins to account for limited calibration info
5. Asymptotic inference

11.7 days analysis limit improved by $\times$ ~2

Statistical treatments

- Experiments are “fixed” - once it is built, there are limited opportunities for upgrading
- Statistical methods can provide a way to get every last piece of information out of the data
  - This is a good thing!
- And in some cases, being smarter can let you win
- But statistically speaking, you’re only smarter until you publish, at which point everyone copies you

Statistical treatments - the dangers

- **Lots of choices to make (NB: not just statistical)**

What have some recent experiments done*

- **LUX:** 2017 combined WS2013+WS2014–16 limit, **PLR** with **CLs+b, two-sided** (not stated in paper), **power constrained** at -1σ level, nuisance parameters as gaussian constraints [arxiv/1608.07648]
- **PandaX-II:** 2017 54 tonne.day limit, **PLR** with **standard TS** (not clear if one/two-sided), considered power constraint but not applied as close to -1 σ, gaussian nuisance params [arxiv.org/1708.06917]
- **XENON1T:** 2019 1 tonne.year limit, + signal-like "safeguard" term, report only upper limit if < 3σ, discovery, blinding+salting [arxiv.org/1902.11297]
- **DEAP-3600:** 2019 231-day exposure limit, 90% CL upper limit using **Highland-Cousins method** [arxiv.org/1902.04048]
- **CRESST-III:** 2019 limit, used **Yellin optimum interval algorithm**, no background subtraction [arxiv.org/1904.00498]

* my understanding based on writeups

Statistical treatments - the dangers

- **Lots of choices to make (NB: not just statistical)**

What have some recent experiments done*

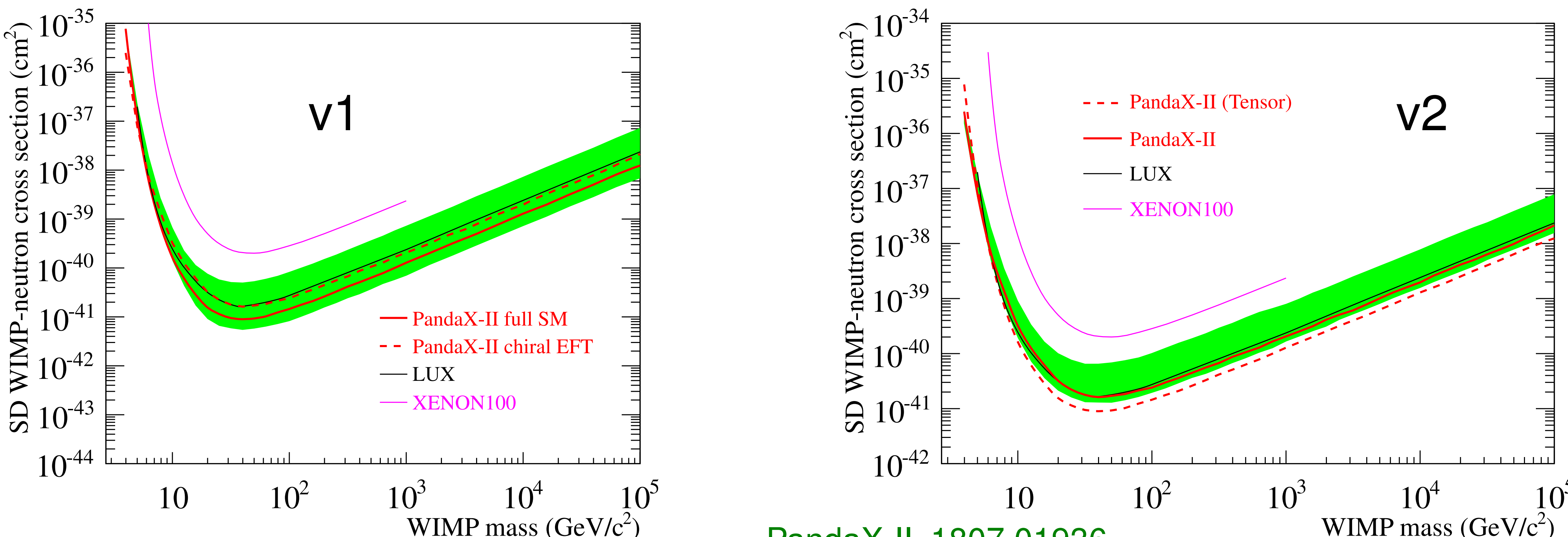
- **SuperCDMS:** sensitivity projection using an optimum interval calculation [arxiv.org/1610.00006]
- **DarkSide-20k:** sensitivity projections, not clear what was used [arxiv.org/1707.08145]
- **XENONnT:** sensitivity projections, **PLR** with **one-sided TS**, **CLs**, gaussian nuisance params [arxiv.org/1512.07501]
- **LZ:** sensitivity projections, **PLR** with **one-sided TS**, **CLs+b**, power constraint at -1σ, gaussian nuisance params [arxiv.org/1802.06039]
- Plus others that I've probably missed...

* my understanding based on writeups

Statistical treatments - the dangers

- Lots of choices to make (NB: not just statistical)
- Some case studies

Case 1: Panda X SD



PandaX-II, 1807.01936

- Convention for SD scattering goes back at least to Engel, Pittel, Vogel 1992

  $\hookrightarrow$ **axial-vector–axial-vector** current $\bar{\chi}\gamma^\mu\gamma_5\chi\,\bar{q}\gamma_\mu\gamma_5 q$ (motivated by SUSY)
- In QCD

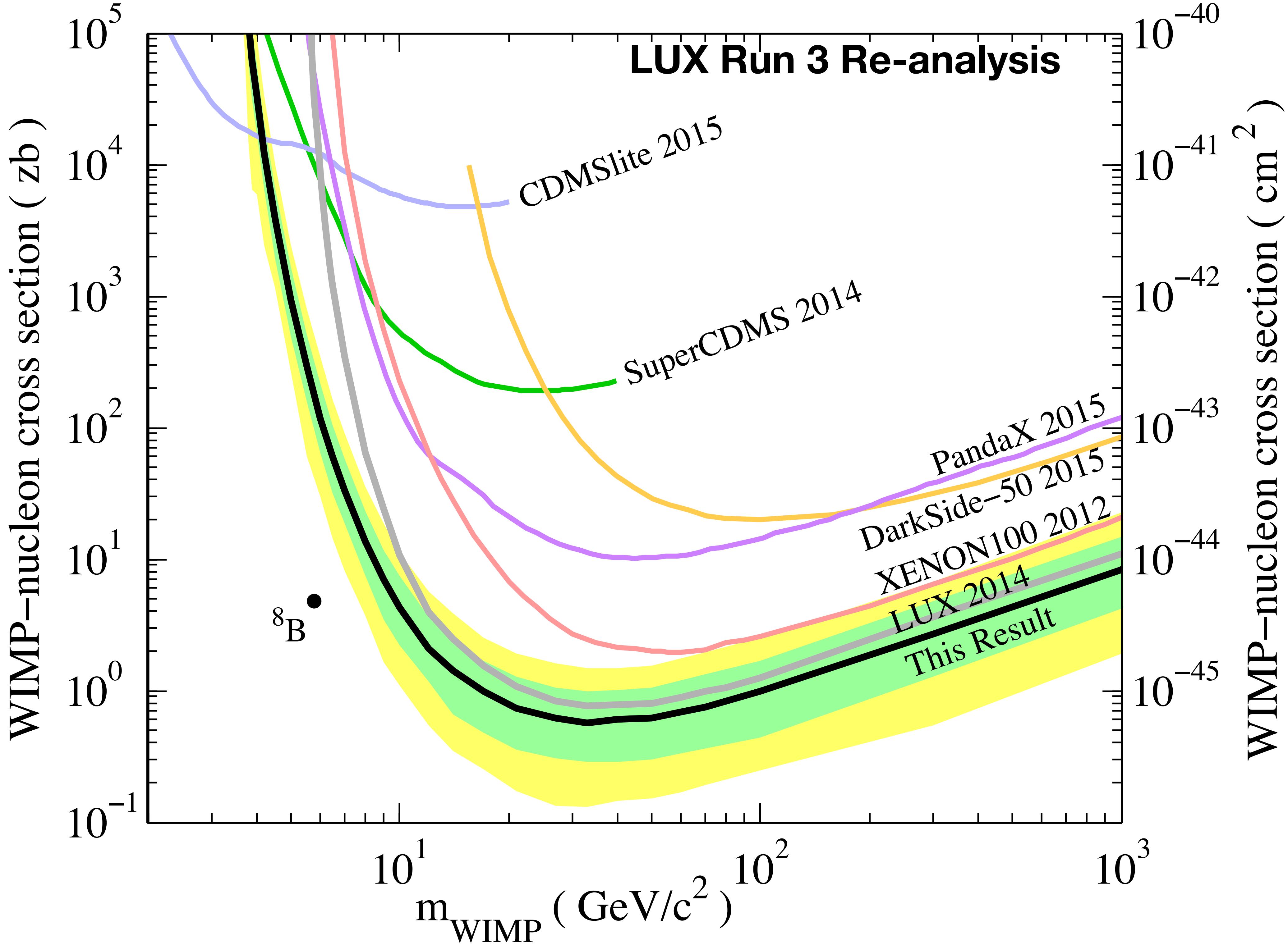
$$\langle N(p')|\bar{q}\gamma_\mu\gamma_5\tau^3 q|N(p)\rangle = \langle N(p')|\gamma^\mu\gamma_5 G_A(q^2)\tau^3 + \gamma_5\frac{q^\mu}{2m_N}G_P(q^2)\tau^3|N(p)\rangle$$

$$G_A(0) = g_A \qquad G_A(q^2) - \frac{q^2}{4m_N^2}G_P(q^2) = \mathcal{O}(M_\pi^2)$$

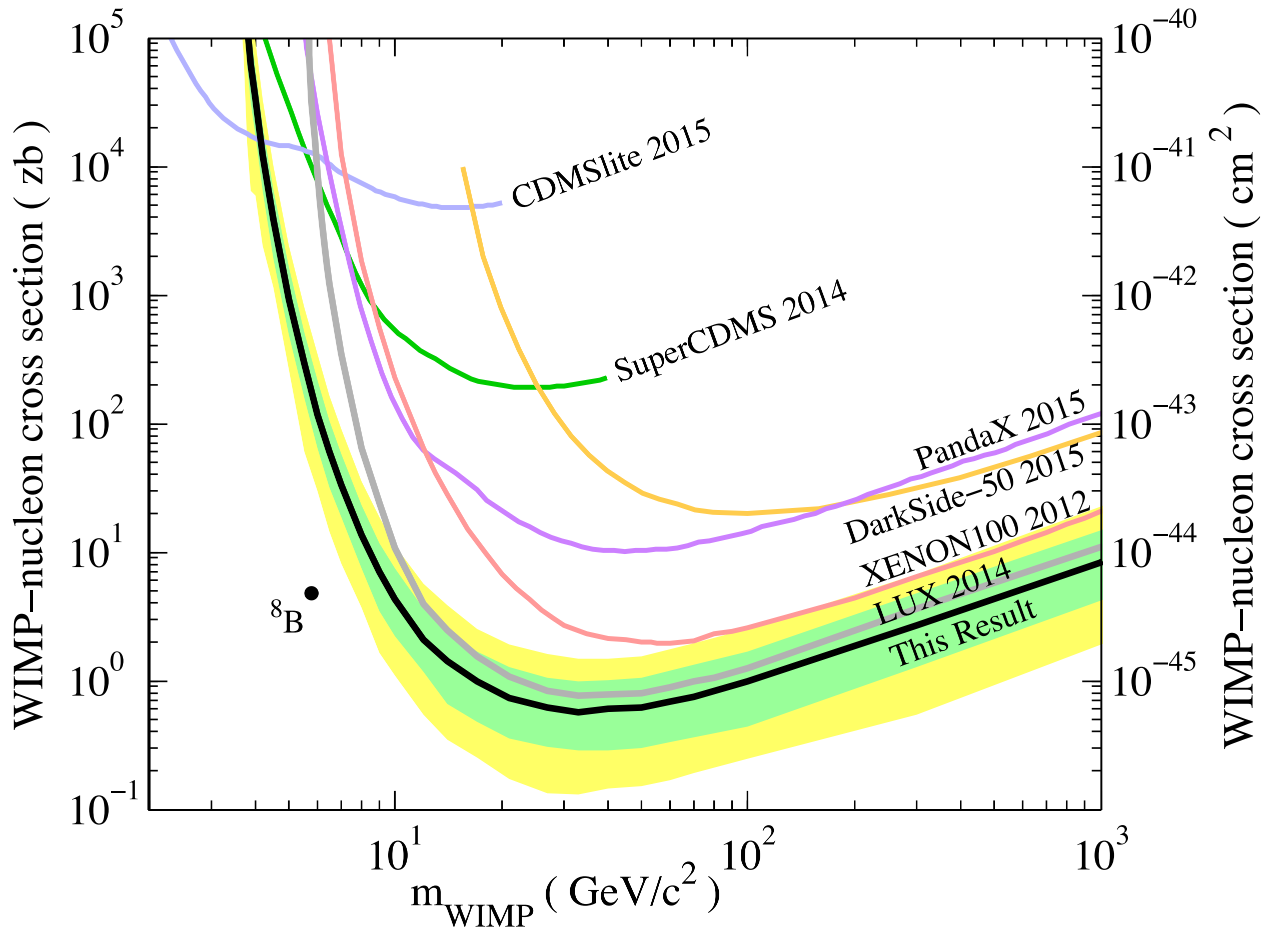
- **Induced pseudoscalar** $G_P(q^2)$ neglected in v1, "improving" the LUX limits

  $\hookrightarrow$ need to use consistent conventions for meaningful comparison!

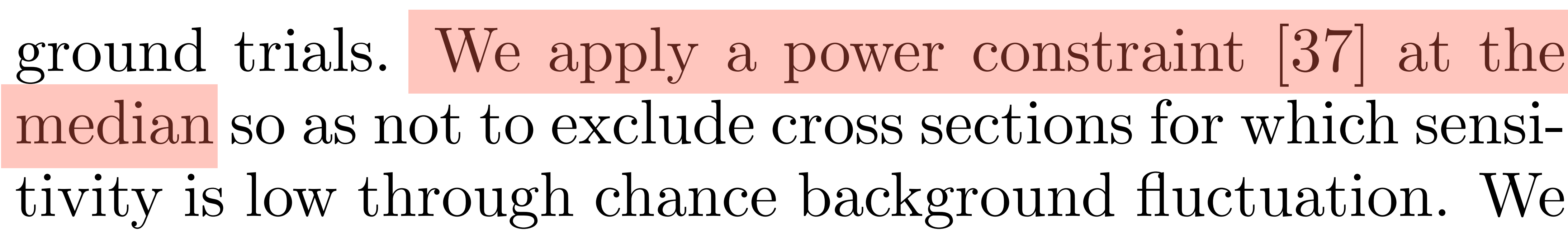
Case 2: LUX, Run 3 + 4



Case 2: LUX, Run 3 + 4

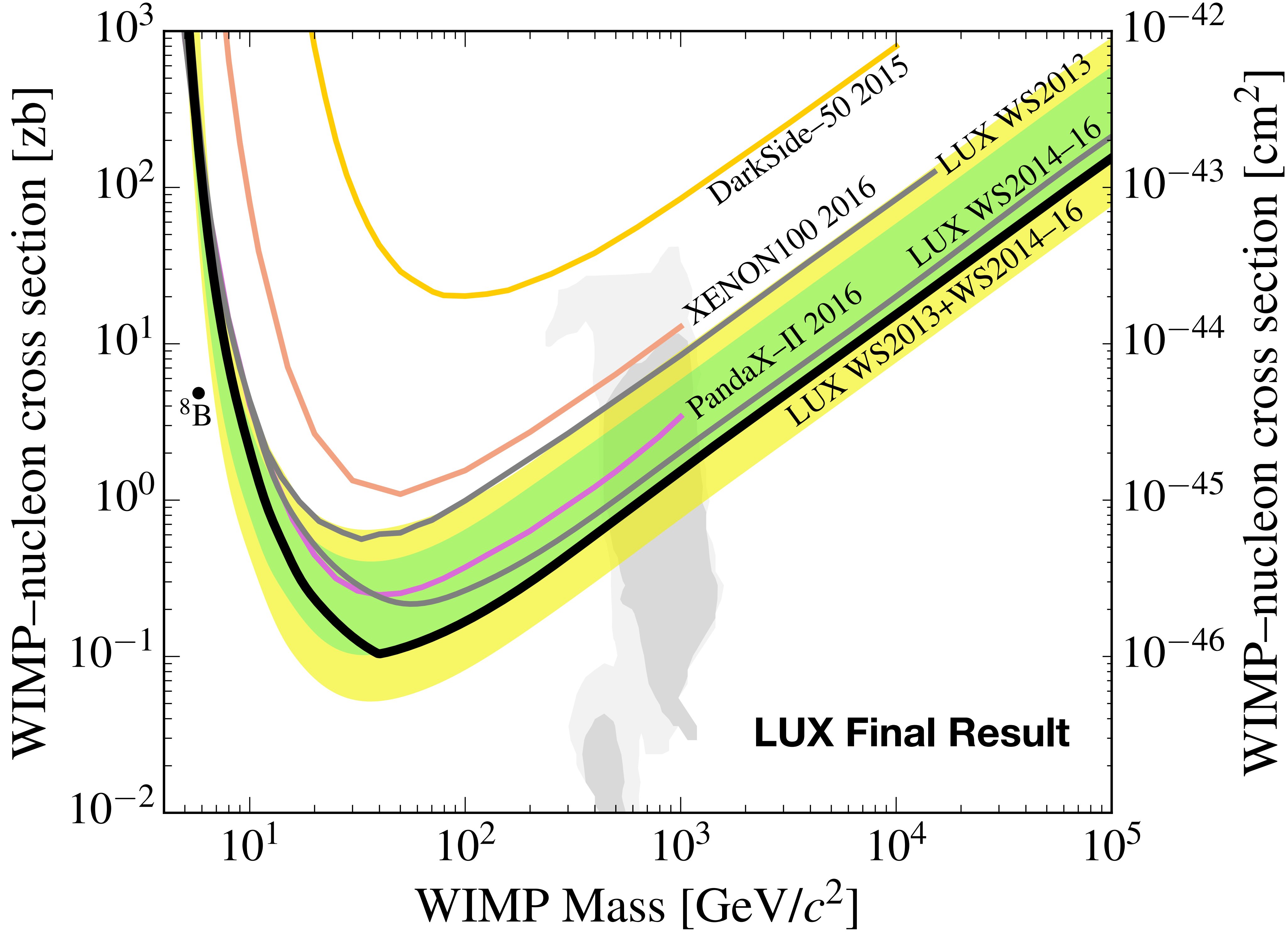


LUX Run 3 Re-analysis 2015

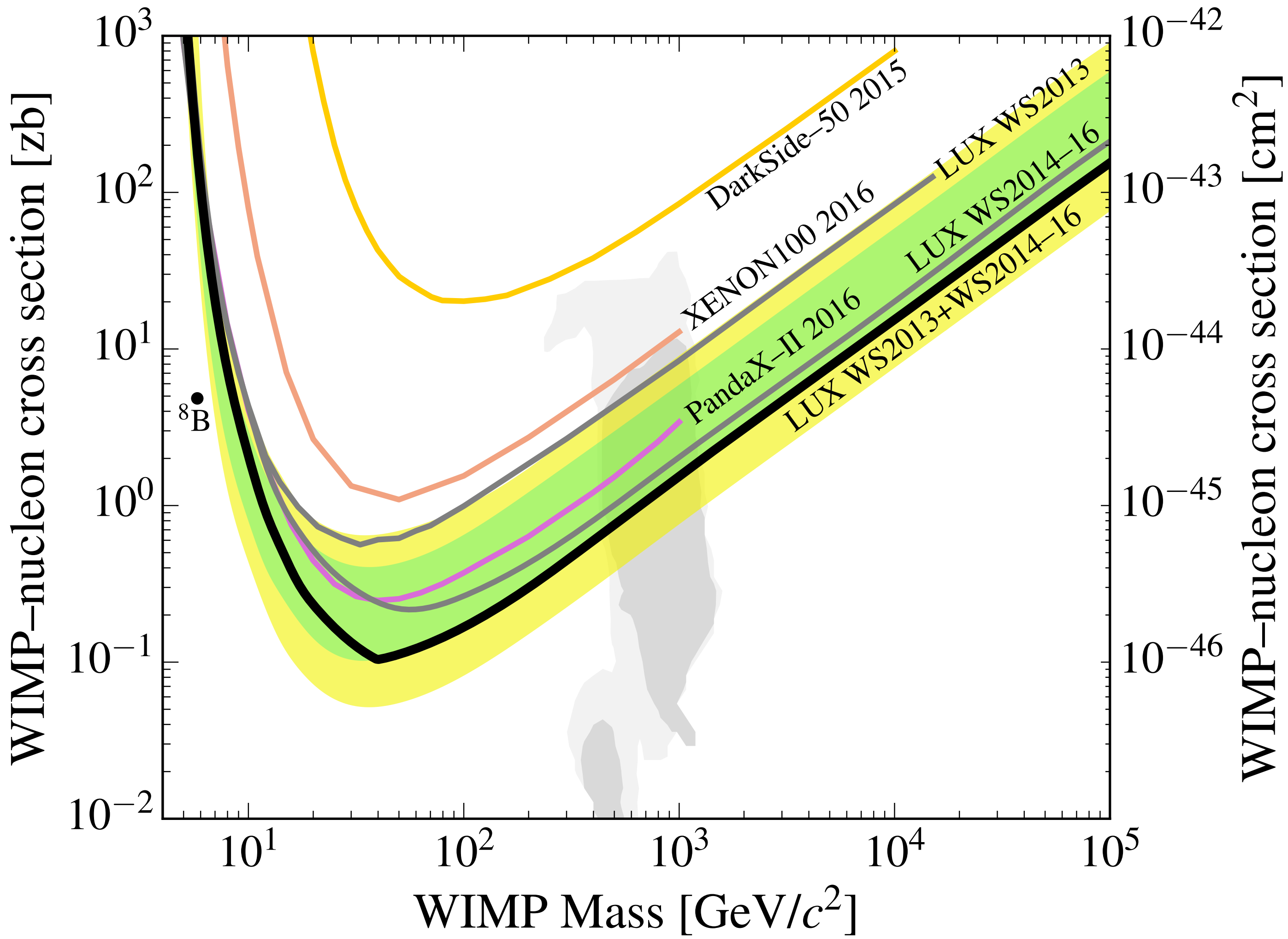


ground trials. We apply a power constraint [37] at the median so as not to exclude cross sections for which sensitivity is low through chance background fluctuation. We

Case 2: LUX, Run 3 + 4



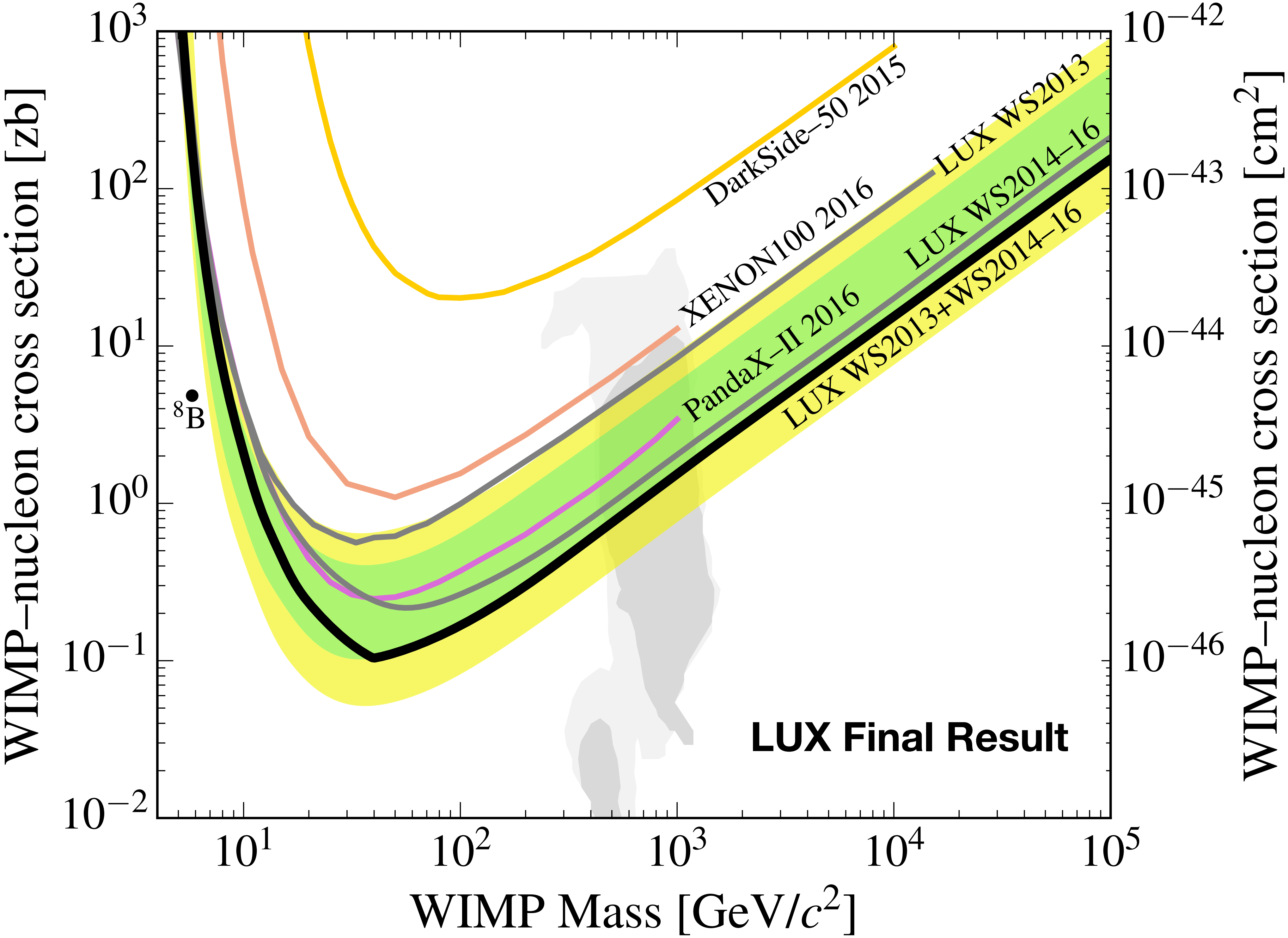
Case 2: LUX, Run 3 + 4



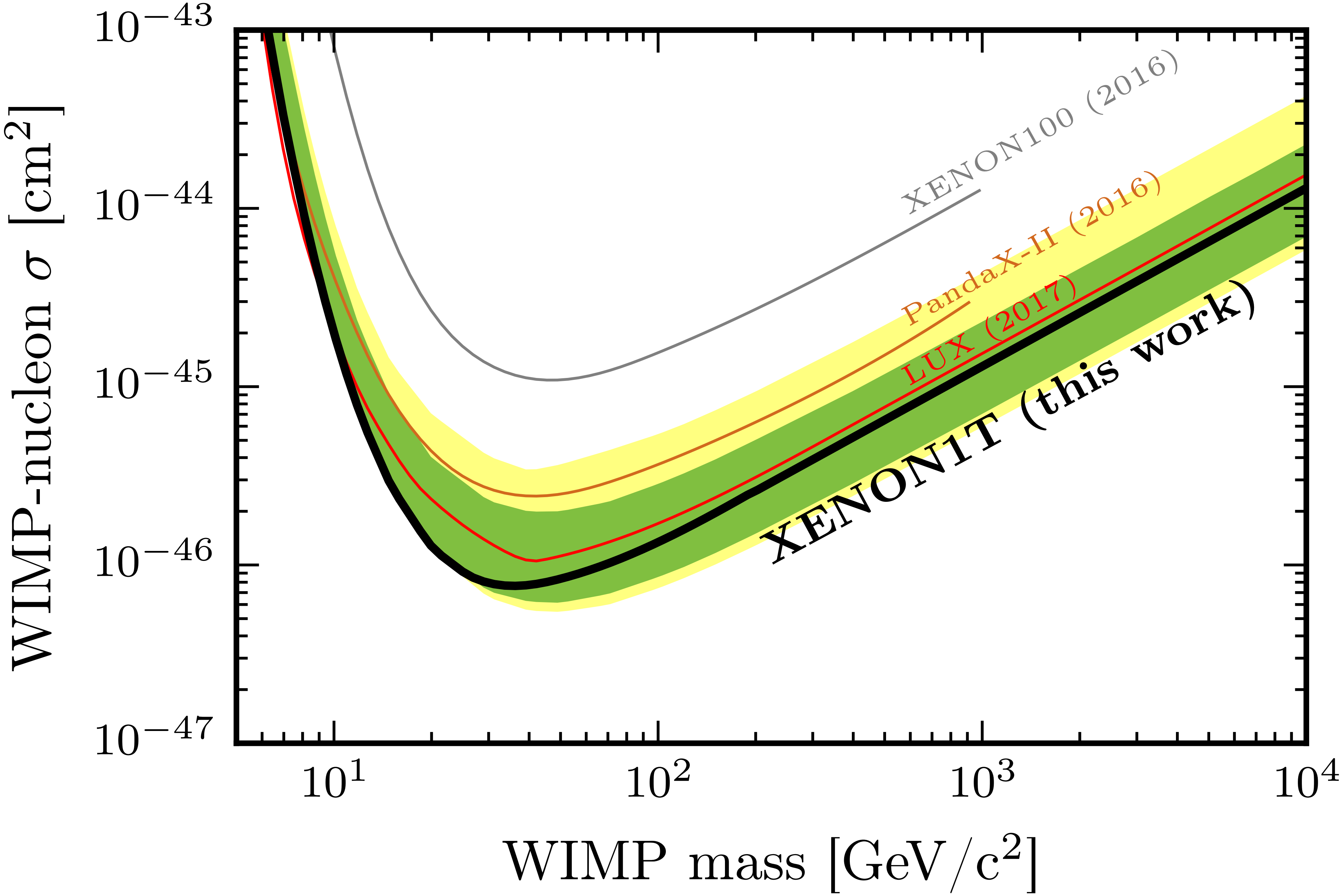
LUX Final Result

exception of the Lindhard k parameter. We conservatively apply a power constraint [49] at the -1σ extent of the projected sensitivity in order to avoid excluding cross sections for which the sensitivity is unreasonably

Case 2: LUX, Run 3 + 4



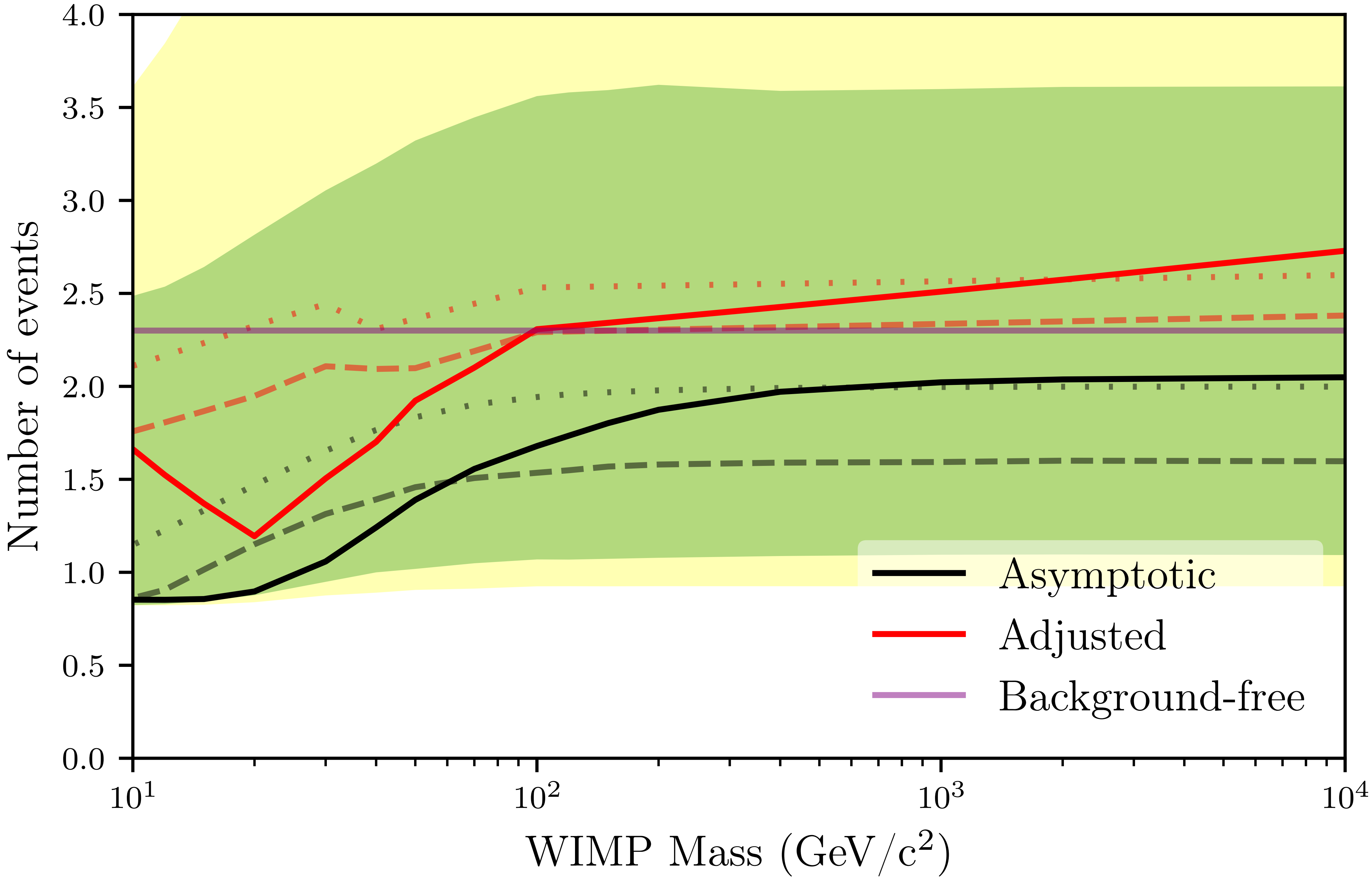
Case 3: XENON1T, SR0



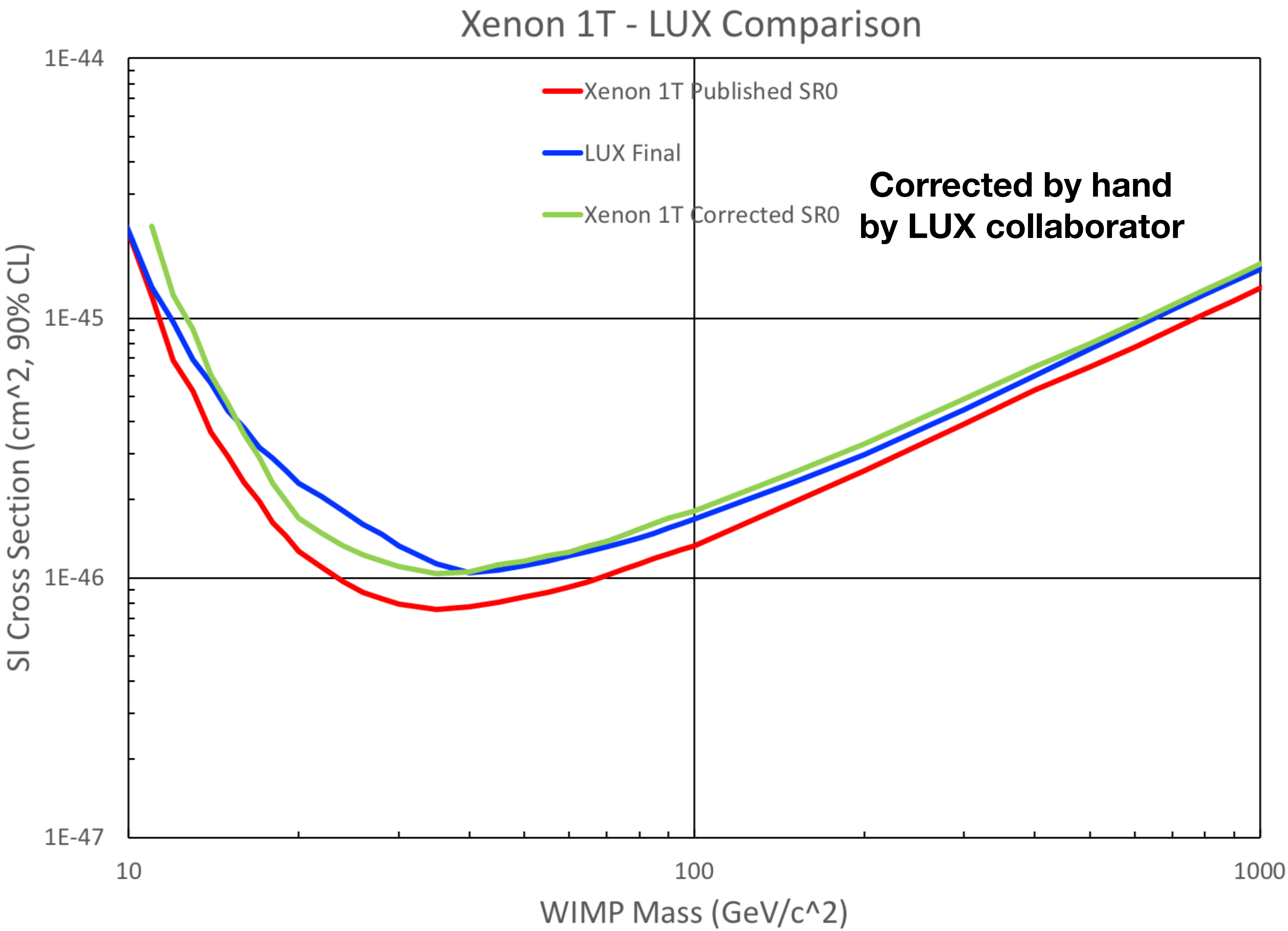
included. The likelihood ratio distribution is approximated by its asymptotic distribution [25]; preliminary

Case 3: XENON1T, SR0

J. Aalbers Thesis: “It is not clear that the asymptotic distribution assumption is accurate

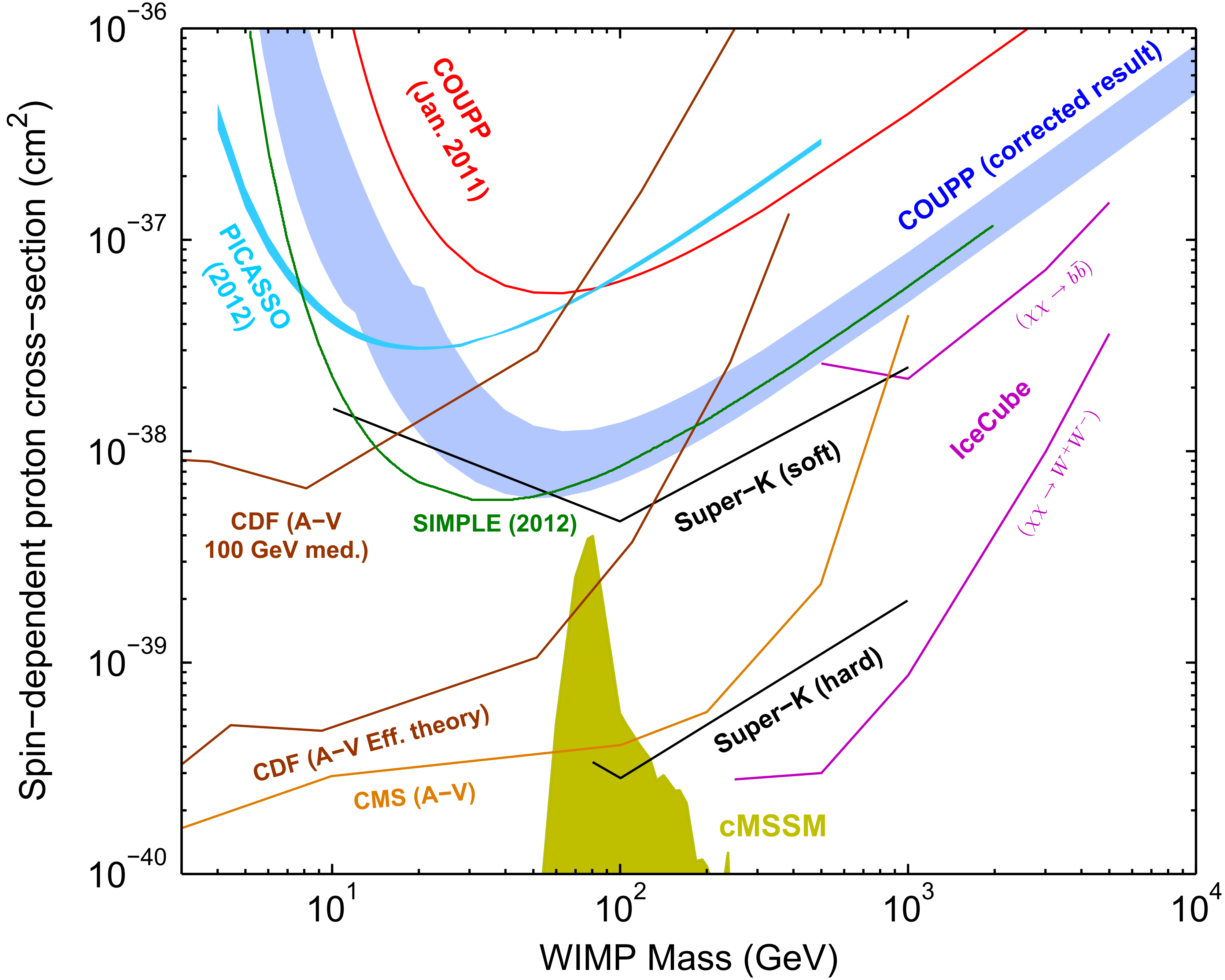


Case 3: XENON1T, SR0

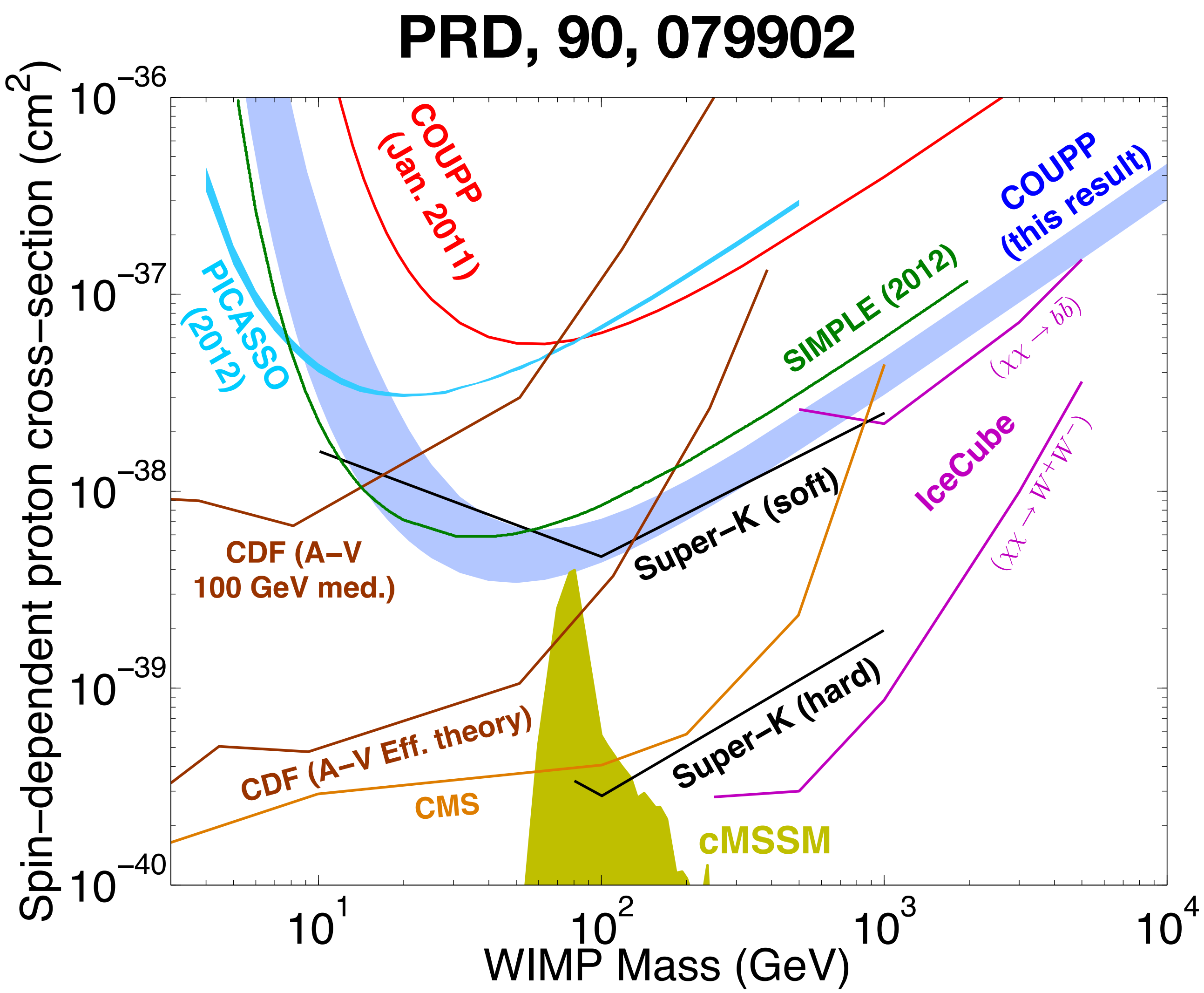


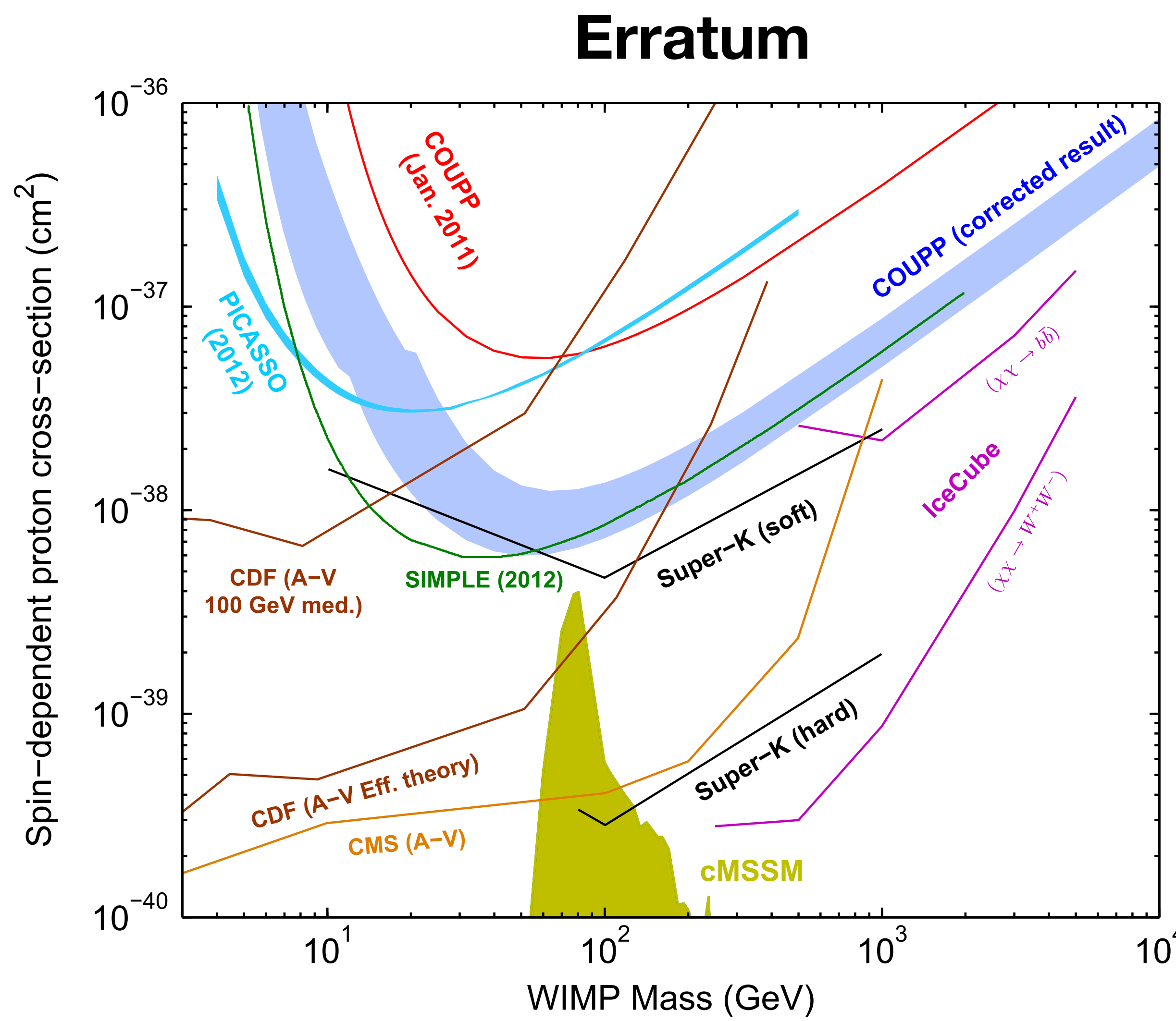
Case 4: COUPP4, 2012

PRD, 90, 079902, Erratum



Case 4: COUPP4, 2012





- **I had combined results from three different runs incorrectly**
- **I did not intentionally do this - but the fact that we were better than SIMPLE would possibly have led me not to cross check as thoroughly as if it had been reversed**
- **This does not address the "band" that was an attempt (incorrect as it turns out) to represent uncertainty on our efficiency**

Bias Mitigation

- **Two good talks from M. E. Monzani and B. Loer on bias mitigation in analysis**
- **Maybe we could “blind” what statistical choices we make…**
  - **“Upon unblinding, we discovered that we had used the CLs statistic with an alpha of 0.05”**

Bias Mitigation

- Two good talks from M. E. Monzani and B. Loer on bias mitigation in analysis
- Maybe we could “blind” what statistical choices we make…
  - “Upon unblinding, we discovered that we had used the CLs statistic with an alpha of 0.05”

But seriously, how to mitigate?

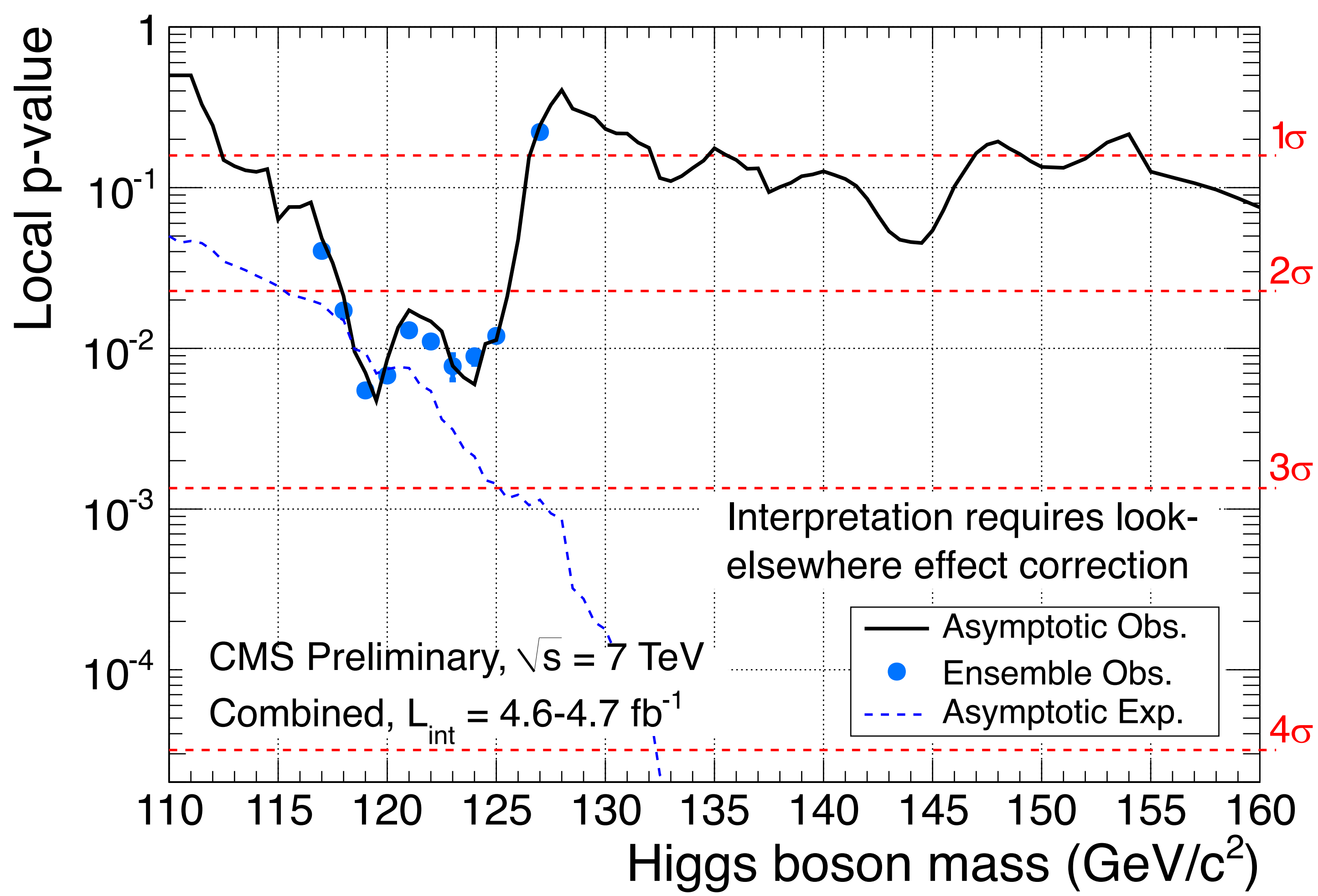
- Limit the number of choices to be made -> white paper?
- Discuss our methods as much as possible
  - Publish likelihood, publish tests, publish nuisance parameters (possible recommendation of white paper?)
  - New developments are immediately treated with suspicion - “are they going to do better with a trick?”
    - Discussion of the “safeguard term” here was great

Discovery

- The goal is to discover dark matter, not set limits
- Statistical techniques help make absolutely sure we don't miss something
  - In a likely case where you need to build another detector to verify/increase significance, they help provide the motivation

Comparisons - Significance

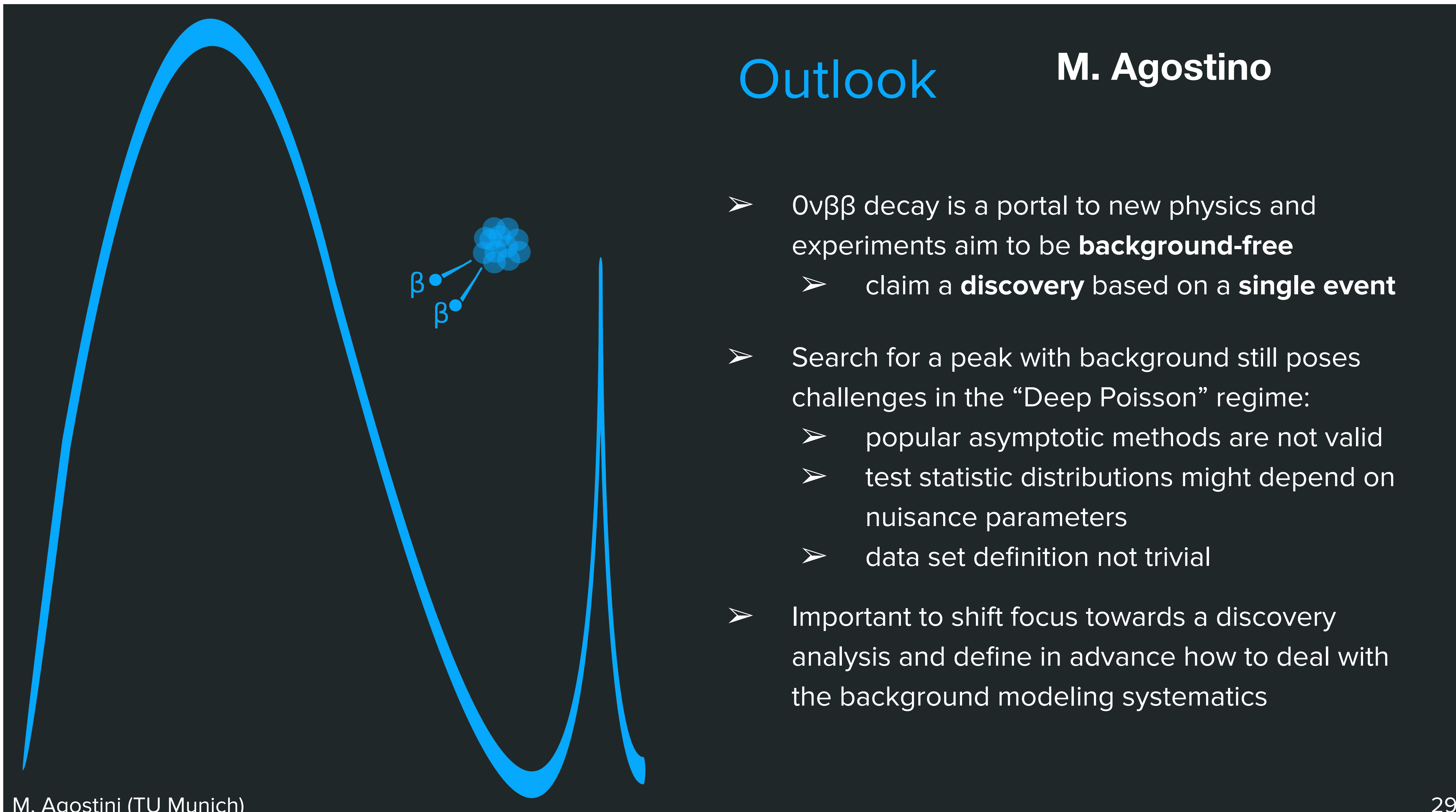
CMS-PAS-HIG-11-032



N. Wardle

Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**



Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**

Systematic Uncertainties

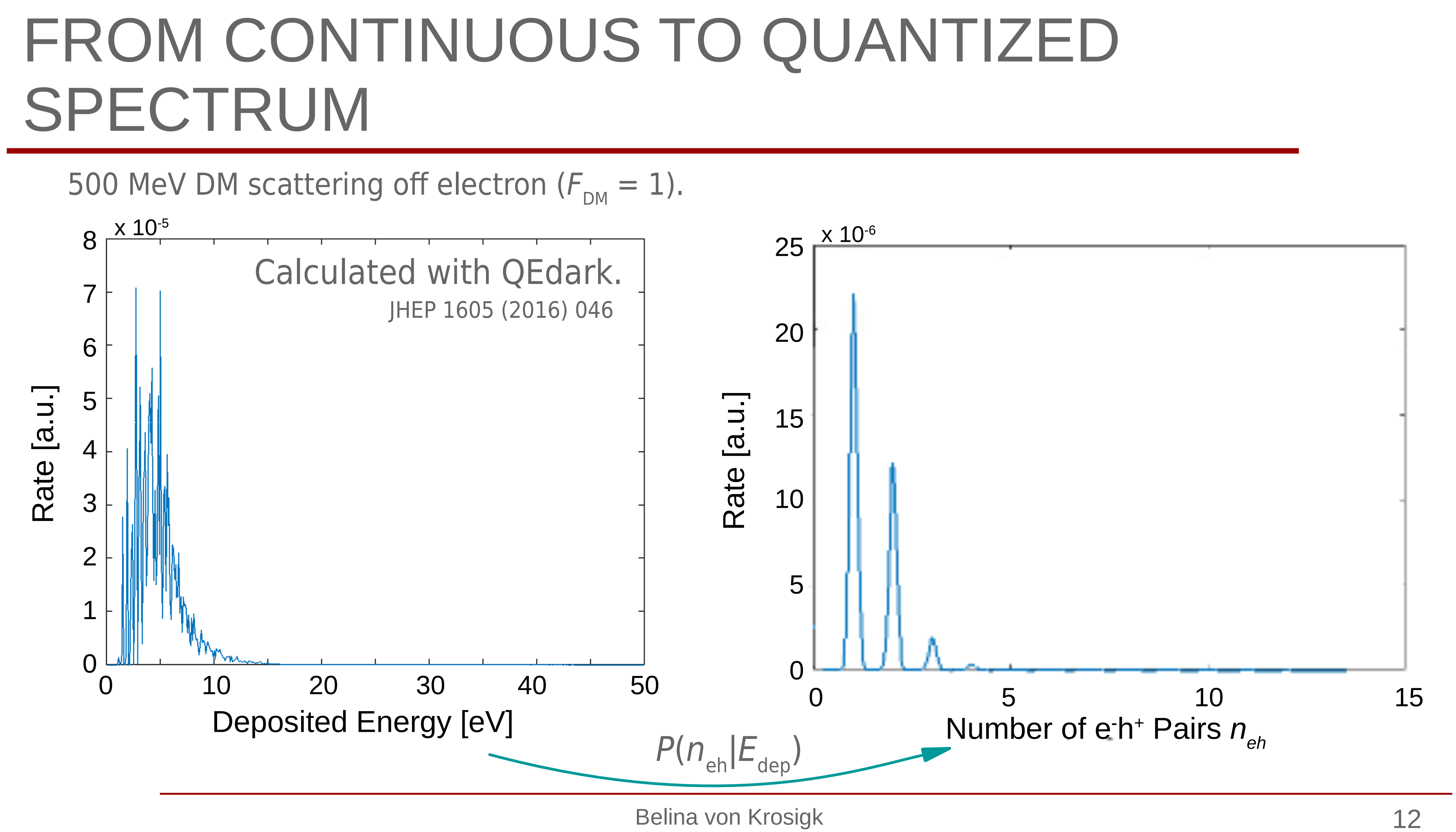
- Statistical uncertainty can affect the result by a factor 2 or 3
- Systematic uncertainties typically affect the result by ≲10%
- Accounted by nuisance parameters and pull terms (auxiliary data) or priors
- Sources:
  - background modeling
  - energy scale and resolution
  - signal detection efficiency (active volume & analysis cuts)
-

Background modeling is troublesome in case of a **discovery based on 1 single event**:

- Gas/Liquid detectors
  - complicated background modeling
  - all components considered?
  - shapes correct within uncertainties?
- Solid state detectors
  - granular design -> many pixels
  - is background homogenous?
  - how to create data sets?

Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**
  - **Similar issues in the presence of backgrounds**
- **Need background (and signal) model that you trust**

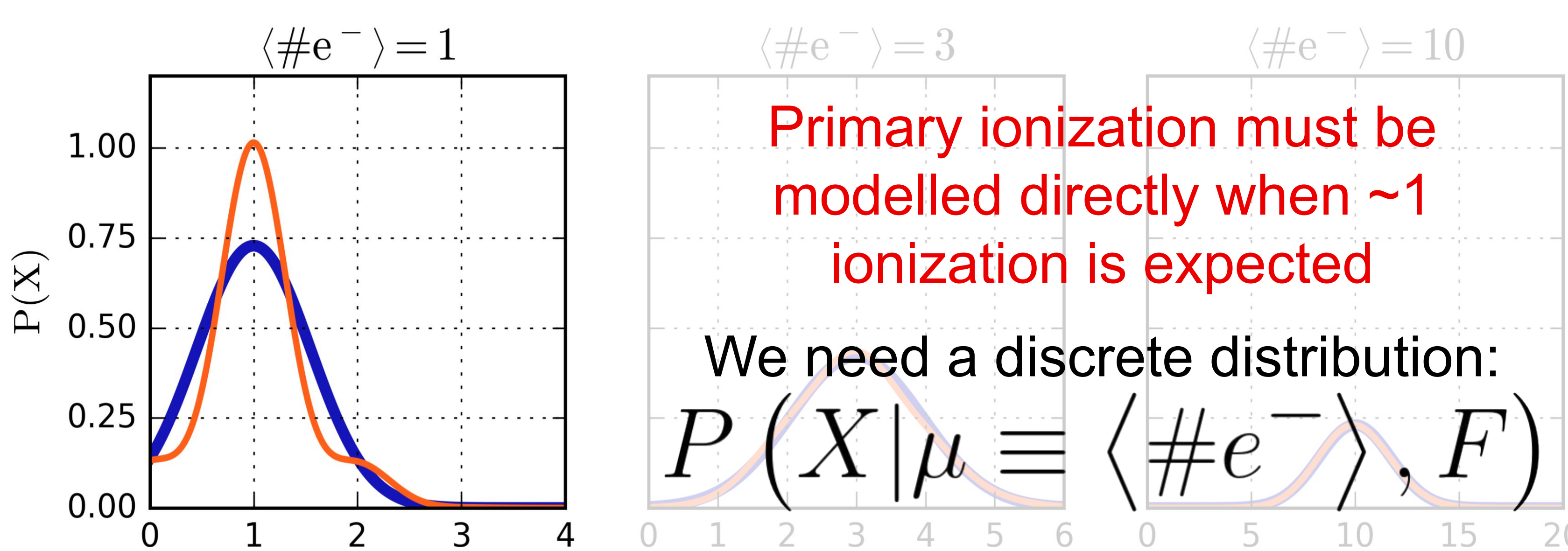


Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**
  - **Similar issues in the presence of backgrounds**
- **Need background (and signal) model that you trust**

The issue of modelling ionization

At high energies, it is valid to fold ionization fluctuations in with resolution effects (i.e. baseline noise), to give an overall model with an effective F



Primary ionization must be modelled directly when ~1 ionization is expected

We need a discrete distribution:

$$P\left(X|\mu \equiv \langle \#e^- \rangle, F\right)$$

F = 0.2 for ionization + Gaussian noise with σ = 0.5

Gaussian with same “effective F”

D. Durnford

Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**
  - **Similar issues in the presence of backgrounds**
- **Need background (and signal) model that you trust**

The backgrounds in the detector are usually not mono-energetic, so build a sum over the contributions from all energies to a given TotalPE slice.

$$R(f; N_t) = \Sigma_E \, T(E) \circ N_E(N_t) \circ N_t \circ P_E'(f; N_t)$$

Or use an 'effective model'

For approximately flat spectra and monotonic resolution function:

$R(f; N_t)$ = Fully correlated Hinkley function (with arbitrary 'width' parameters)
= Gamma distribution ⊙ Gaussian

T. Pollman

Future of backgrounds in one word

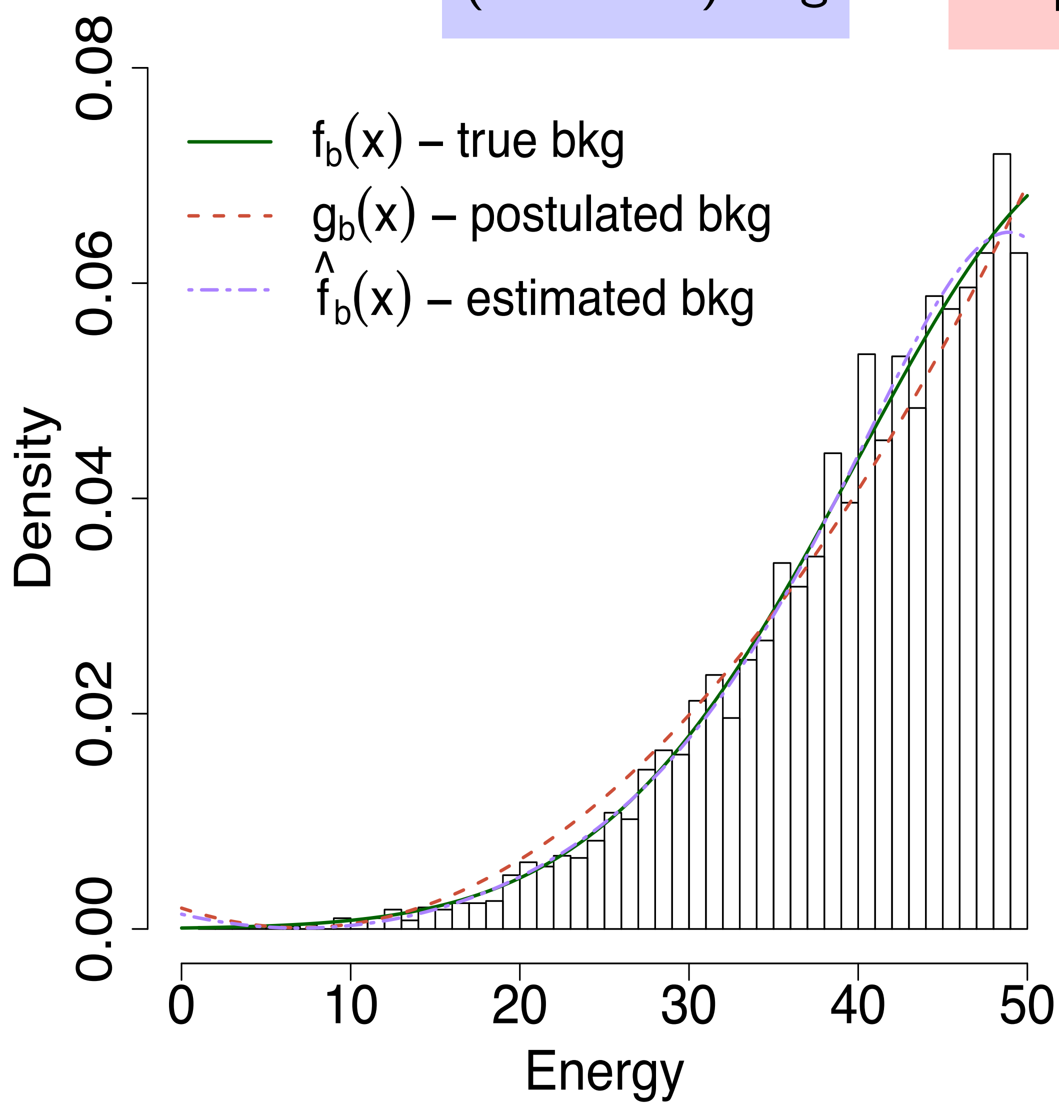
- How do
  signal?
- Claim di
  - Similaı
- Need ba

R. Calkins

In the background problem...

We can exploit the skew-G density model and write

$$\underbrace{f_b(x)}_{\text{true (unknown) bkg}} = \underbrace{g_b(x)}_{\text{postulated bkg}} \underbrace{d(G_b(x); G_b, F_b)}_{\text{comparison density}} \quad (6)$$



We then obtain the LP skew density estimator $\hat{f}_b(x)$ of $f_b(x)$ as discussed in the previous slide.

$\Rightarrow$**Q1** (Modelling) ✓

Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**
  - **Similar issues in the presence of backgrounds**
- **Need background (and signal) model that you trust**

Highlights importance of calibration!

Discovery

- **How do we convince ourselves and others that we have seen a real signal?**
- **Claim discovery based on single event?**
  - **Similar issues in the presence of backgrounds**
- **Need background (and signal) model that you trust**
- **How significant is a result? What significance do you need?**
  - **Nuisance parameters**

THE DANGERS OF NUISANCE PARAMETERS

- When the number of nuisance parameters is appreciable (relative to the number of independent observations), profile maximum likelihood typically produces severely biased estimators of interest parameters (Bartlett, 1937)[a].
- There are comparable difficulties with Bayesian inference based on high-dimensional flat priors.

H. Battey

Discovery

- How do we convince ourselves and others that we have seen a real signal?
- Claim discovery based on single event?
  - Similar issues in the presence of backgrounds
- Need background (and signal) model that you trust
- How significant is a result? What significance do you need?
  - Nuisance parameters
  - Asymptotic approximations

LARGE-SAMPLE THEORY

(asymptotic approximation)

$\Downarrow$

A. Brazzale
and
I. Volobouev

SMALL-SAMPLE THEORY

(higher order asymptotics)

Discovery

- How do we convince ourselves and others that we have seen a real signal?
- Claim discovery based on single event?
  - Similar issues in the presence of backgrounds
- Need background (and signal) model that you trust
- How significant is a result? What significance do you need?
  - Nuisance parameters
  - Asymptotic approximations (we will MC as long as we can)

LARGE-SAMPLE THEORY

(asymptotic approximation)

$$\Downarrow$$

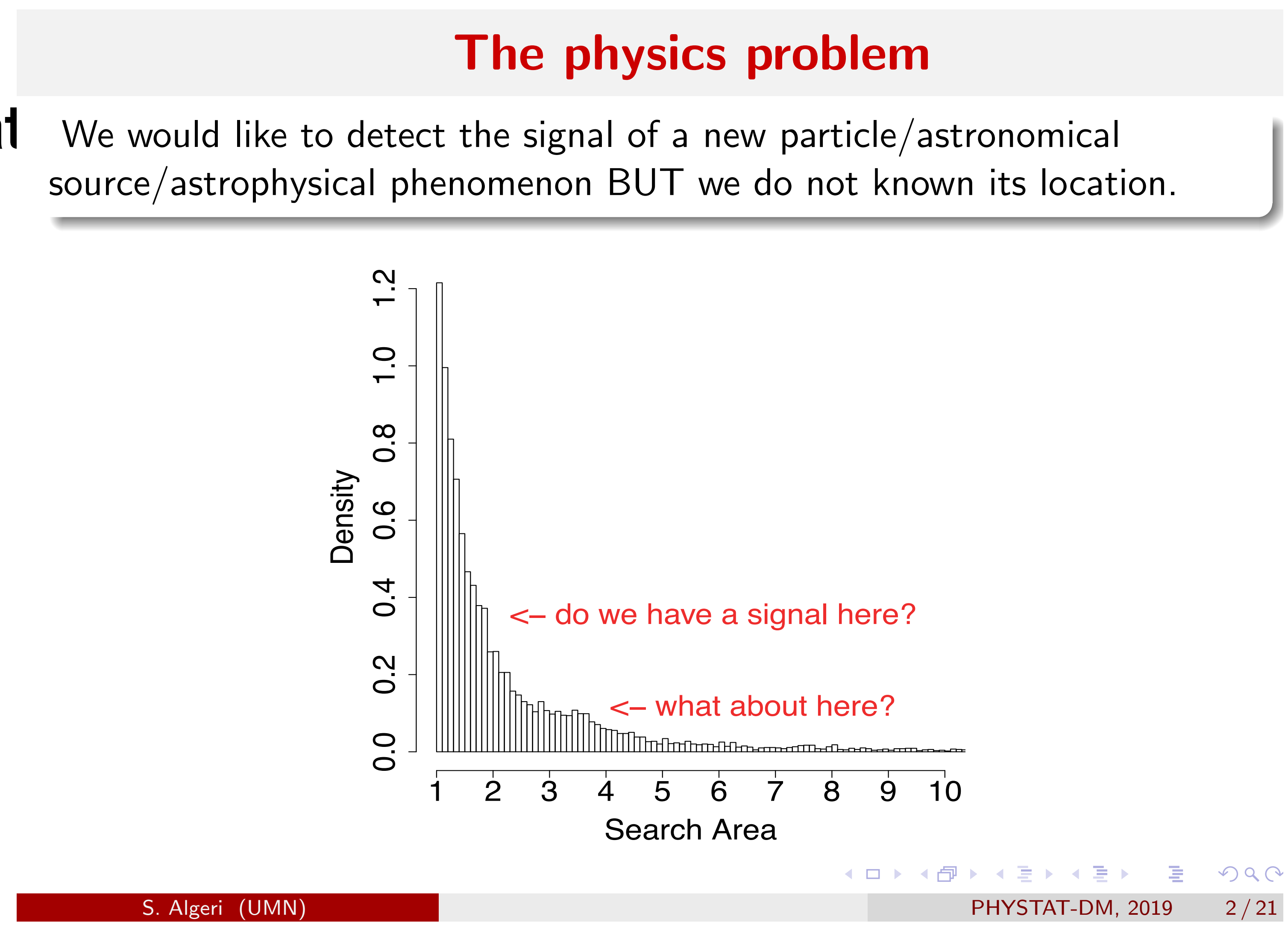
A. Brazzale
and
I. Volobouev

SMALL-SAMPLE THEORY

(higher order asymptotics)

Discovery

- How do we convince ourselves and others that we have seen a real signal?
- Claim discovery based on single event?
  - Similar issues in the presence of backgrounds
- Need background (and signal) model that you trust
- How significant is a result? What significance do you need?
  - Nuisance parameters
  - Asymptotic approximat
  - Look elsewhere effect



S. Algeri

Discovery

How many σ's for discovery?

SEARCH	SURPRISE	IMPACT	LEE	SYSTEMATICS	No. σ
Higgs search	Medium	Very high	M	Medium	5
Single top	No	Low	No	No	3
SUSY	Yes	Very high	Very large	Yes	7
B_s oscillations	Medium/Low	Medium	Δm	No	4
Neutrino osc	Medium	High	$\sin^2 2\vartheta$, Δm^2	No	4
$B_s \rightarrow \mu\, \mu$	No	Low/Medium	No	Medium	3
Pentaquark	Yes	High/V. high	M, decay mode	Medium	7
$(g-2)_\mu$ anom	Yes	High	No	Yes	4
H spin ≠ 0	Yes	High	No	Medium	5
4^{th} gen q, l, ν	Yes	High	M, mode	No	6
Dark energy	Yes	Very high	Strength	Yes	5
Grav Waves	No	High	Enormous	Yes	8

Suggestions to provoke discussion, rather than `delivered on Mt. Sinai'

How would you rate 'Dark Matter'?

Bob Cousins: "2 independent expts each with 3.5σ better than one expt with 5σ"

Final thoughts

- This is fun (in doses)
  - I’ve enjoyed debating power constraints, Bayes v. Frequentist, background modeling techniques, etc at all the breaks
- 50 years from now, no one will remember who had the best limit
- There's an opportunity right now to re-standardize to try to remove some of the unnecessary variation
  - LZ and XENON1T are on board (at least those of us in the room)
  - Low mass searches are still new, but not starting from scratch
    - Already facing background difficulties, intense competition
- Could lead to future combinations to increase sensitivity

Many thanks

Louis
Lyons

Olaf
Behnke

Constanze
Hasterok

Tarek
Saab

Many thanks

- **Heather Battey, Imperial College**
- **Sara Algeri, University of Minnesota**
- **Alessandra Brazzale, University of Padua**

Before re-inventing the wheel, try to see if Statisticians have already found a solution to your statistics analysis problem.

Don't use your square wheel if a circular one already exists.

"Good luck"

Many thanks

Boris
Bart
Cristian
Derek
Jörn
Jelle
Oskar Kleins auditorium


ROMER ABSOLUTE