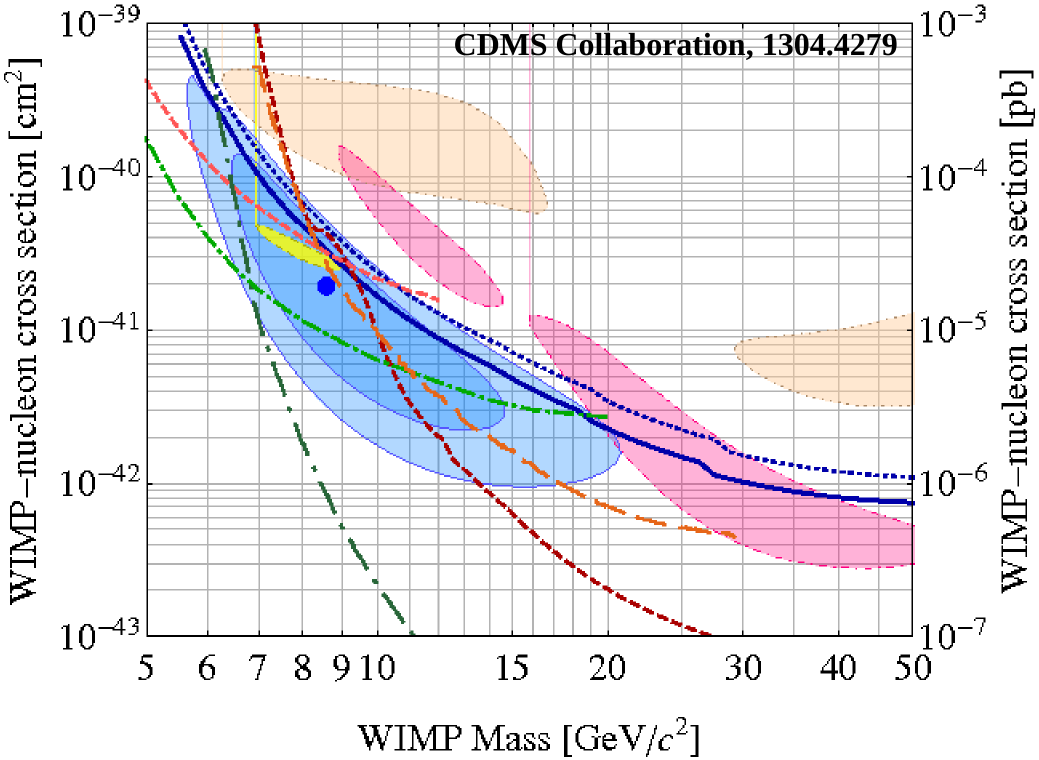
Direct Detection: Statistical Issues and Conservatism

Chris Savage
University of Utah



Are experiments consistent in the light WIMP region?

- Weak statistics
- Flawed/misleading statistics
- Missing/incomplete statistics
- Conservative statistics

Compatibility

- Ideal:
  joint likelihood analysis

- Problem:
  some backgrounds unknown $\Rightarrow$ no likelihood
  Gap/interval methods: p-value (no bkgd subtraction) > p-value (bkgd subtraction)

- Ad-hoc solution:
  plot separate likelihood & exclusion regions,
  look for overlap

Compatibility Regions

- Gap/interval exclusion regions:
  (non-) goodness-of-fit indicator
  - Excludes only parameters with bad fits
  - Can allow parameters with bad fits

- Delta chi-square region (χ^2 - χ^2_{min}):
  not a goodness-of-fit indicator
  - Can contain parameters with bad fits
  - Can exclude parameters with good fits

$\Rightarrow$ g.o.f. given by χ^2_{min}

Care must be taken when interpreting compatibility

Compatibility Regions

- Chi-square regions:
  regions where parameters have good χ^2/dof

  - Generally weaker confidence regions than $\Delta\chi^2$,
    …if reality described by point in this parameter space

  - But actually addresses above "if"

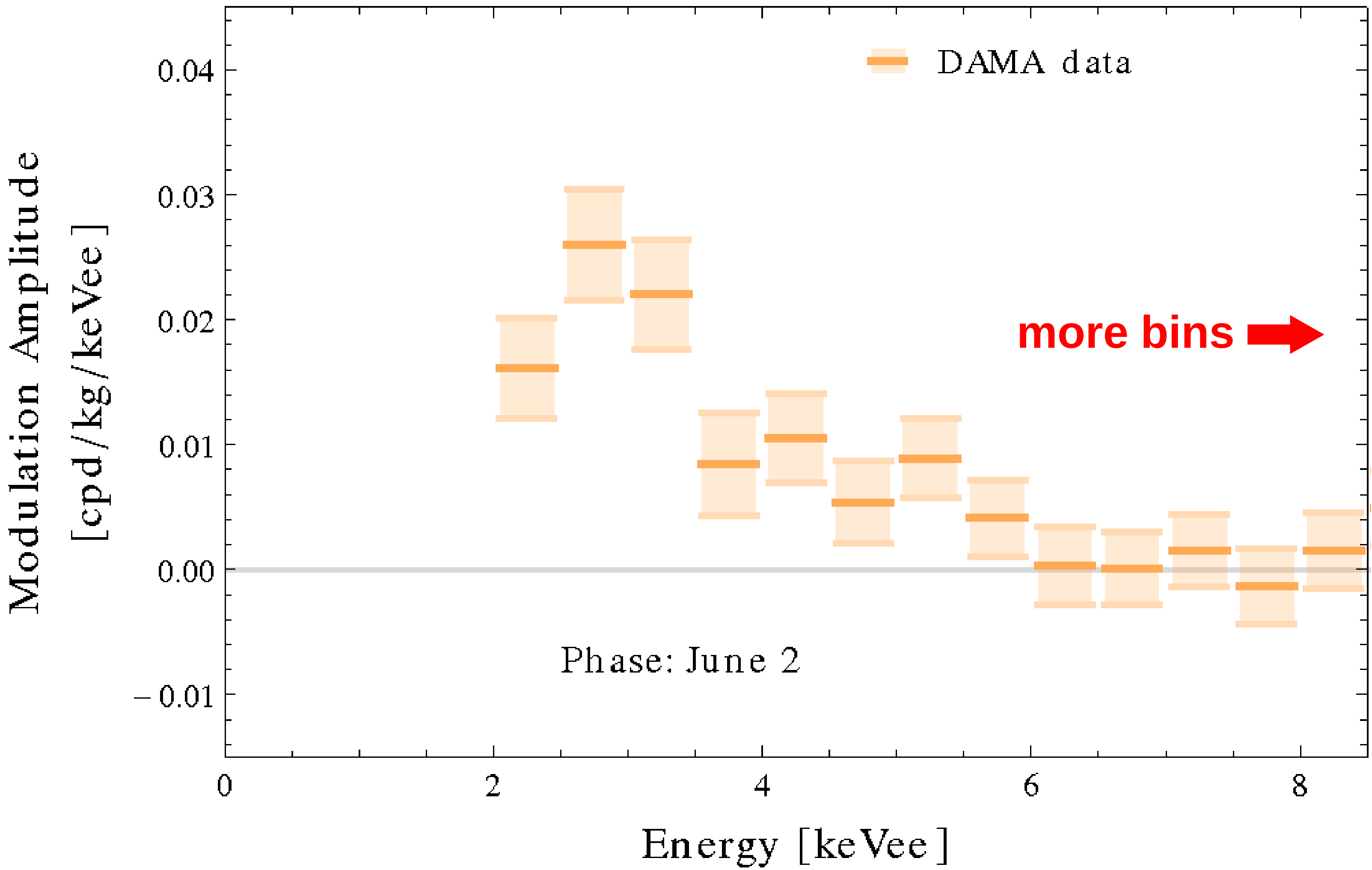
  - Overlap of these regions gives closest indication of what ideal joint analysis g.o.f. would be (i.e. are these results compatible?)

- Problem:
  χ^2 g.o.f. test can be **weak** if poor choice of binning

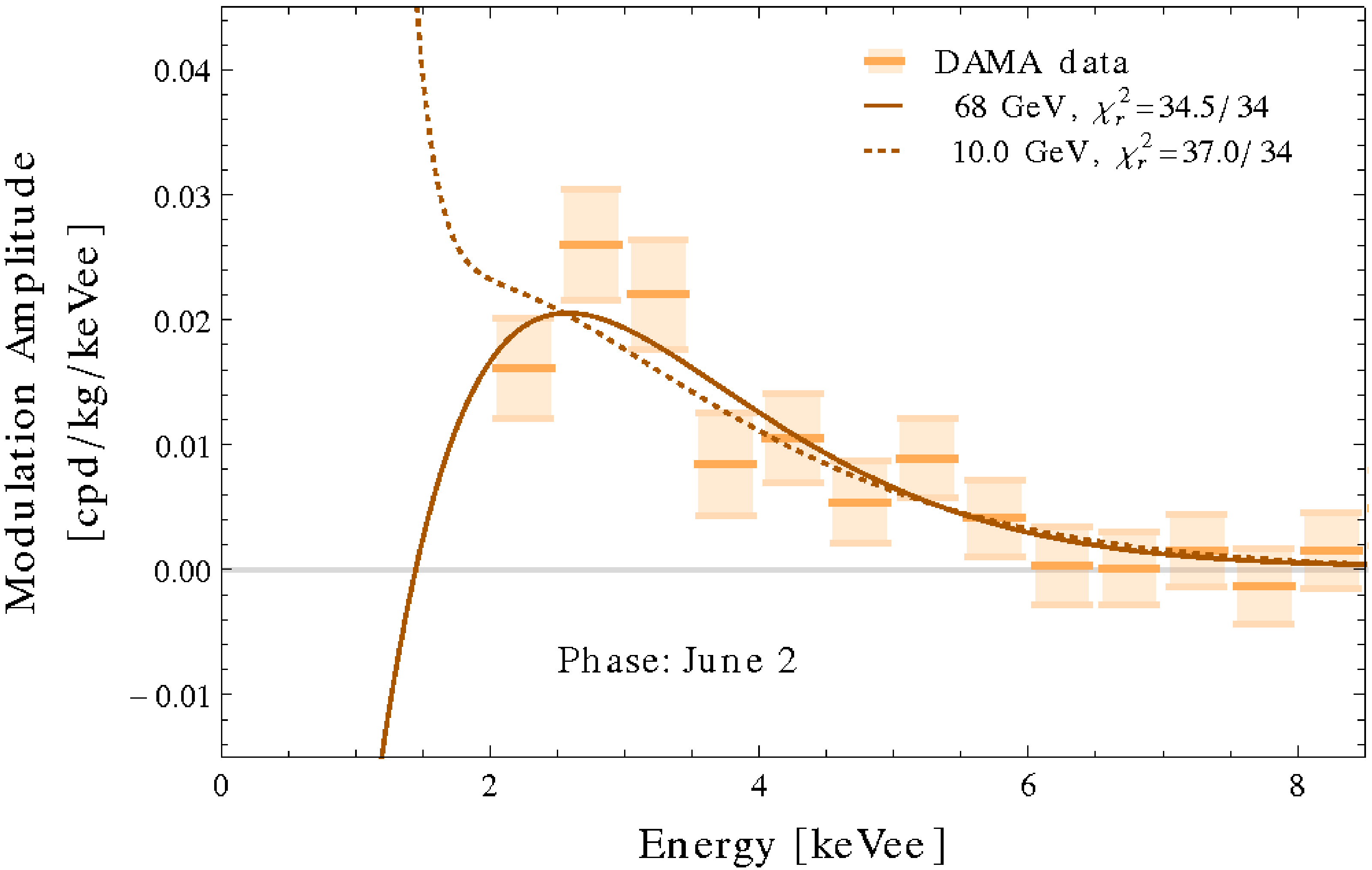
e.g. Savage, Gelmini, Gondolo & Freese (2009)

Weak Statistics

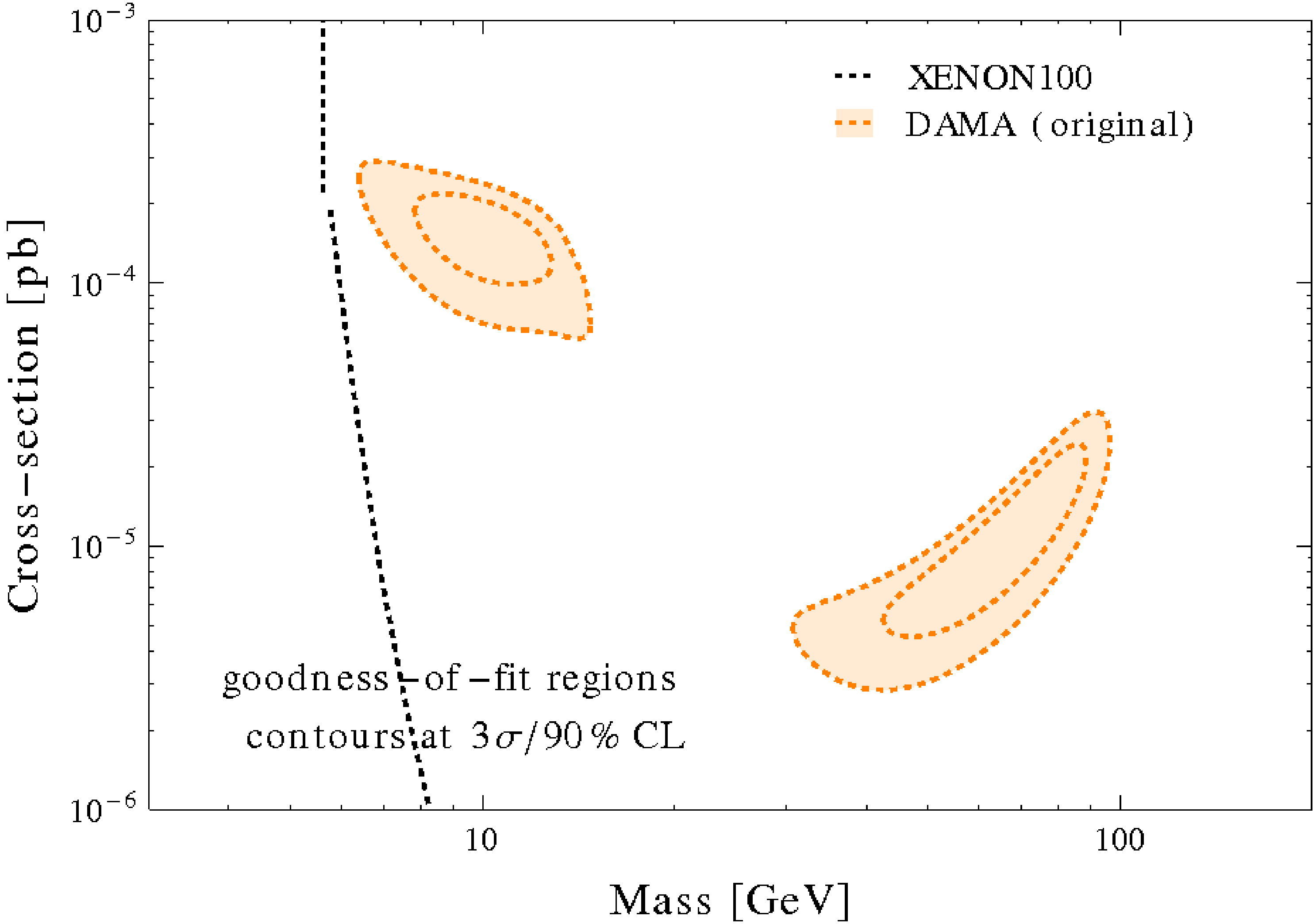
DAMA Results



DAMA Fits


DAMA data
68 GeV, $\chi_r^2 = 34.5/34$
10.0 GeV, $\chi_r^2 = 37.0/34$
Phase: June 2
Modulation Amplitude [cpd/kg/keVee]
Energy [keVee]
0.04
0.03
0.02
0.01
0.00
– 0.01
0
2
4
6
8

DAMA Fits



XENON100
DAMA (original)
goodness-of-fit regions
contours at 3σ/90% CL
Cross-section [pb]
Mass [GeV]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10
100

Binning

Original bins (36):

- Most narrower than energy resolution
- Most expected to have negligible signal

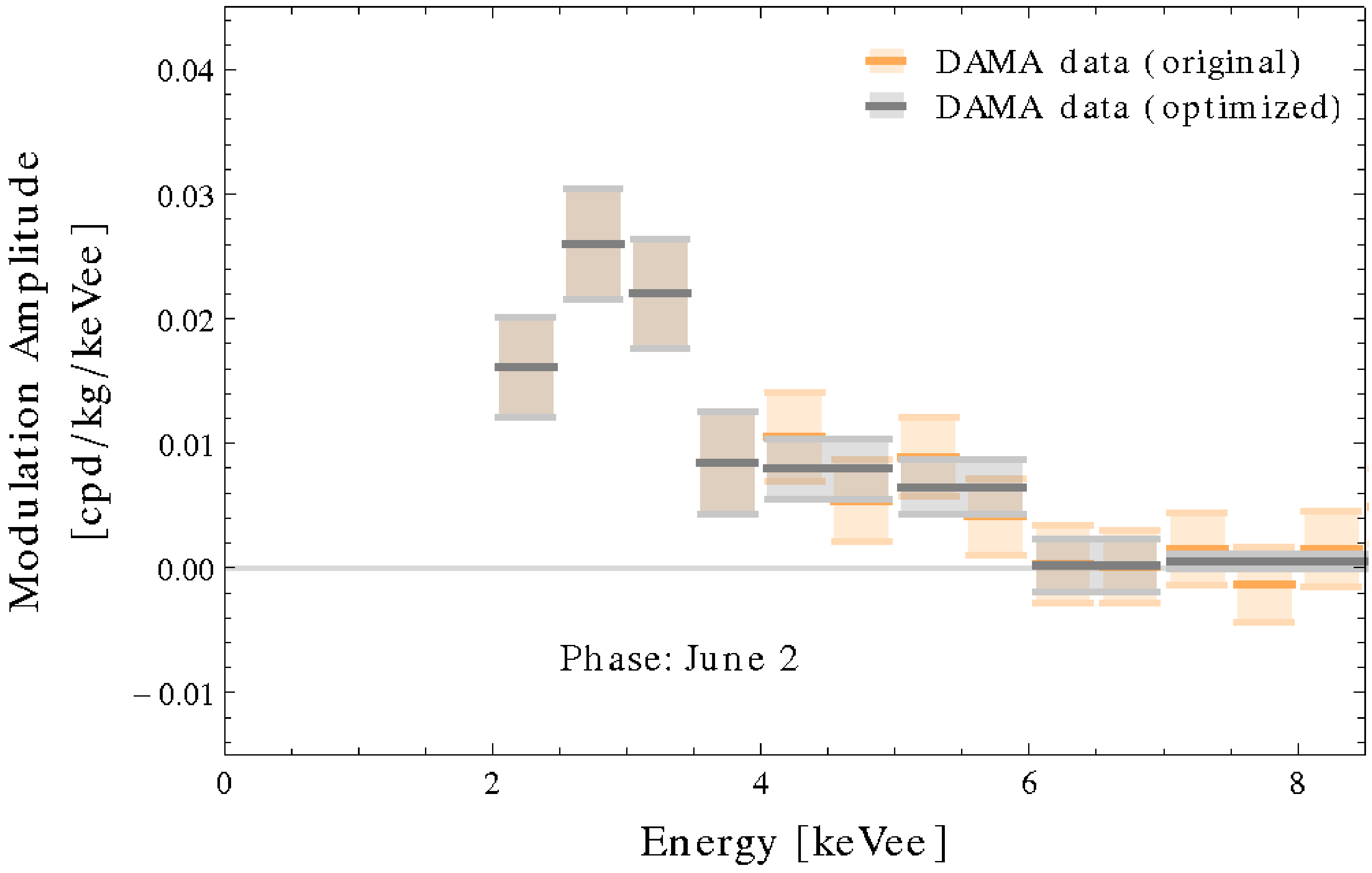
$\Rightarrow$ Sensitivity of goodness-of-fit weakened!

More optimal choice of bins (8):

- Combine bins much smaller than energy resolution
- Combine all bins above 7 keVee

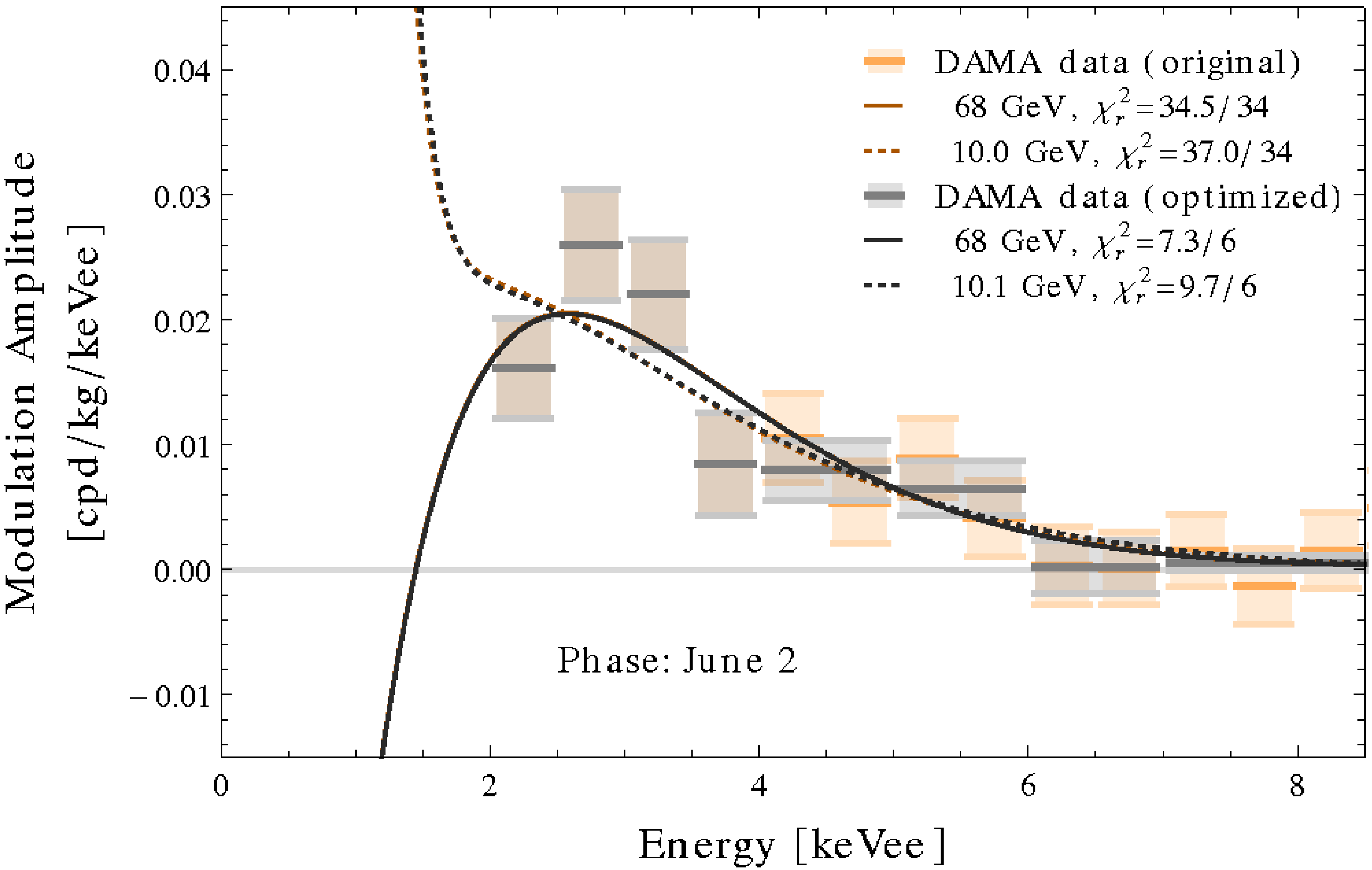
Kelso, Sandick & Savage (in progress)

Binning



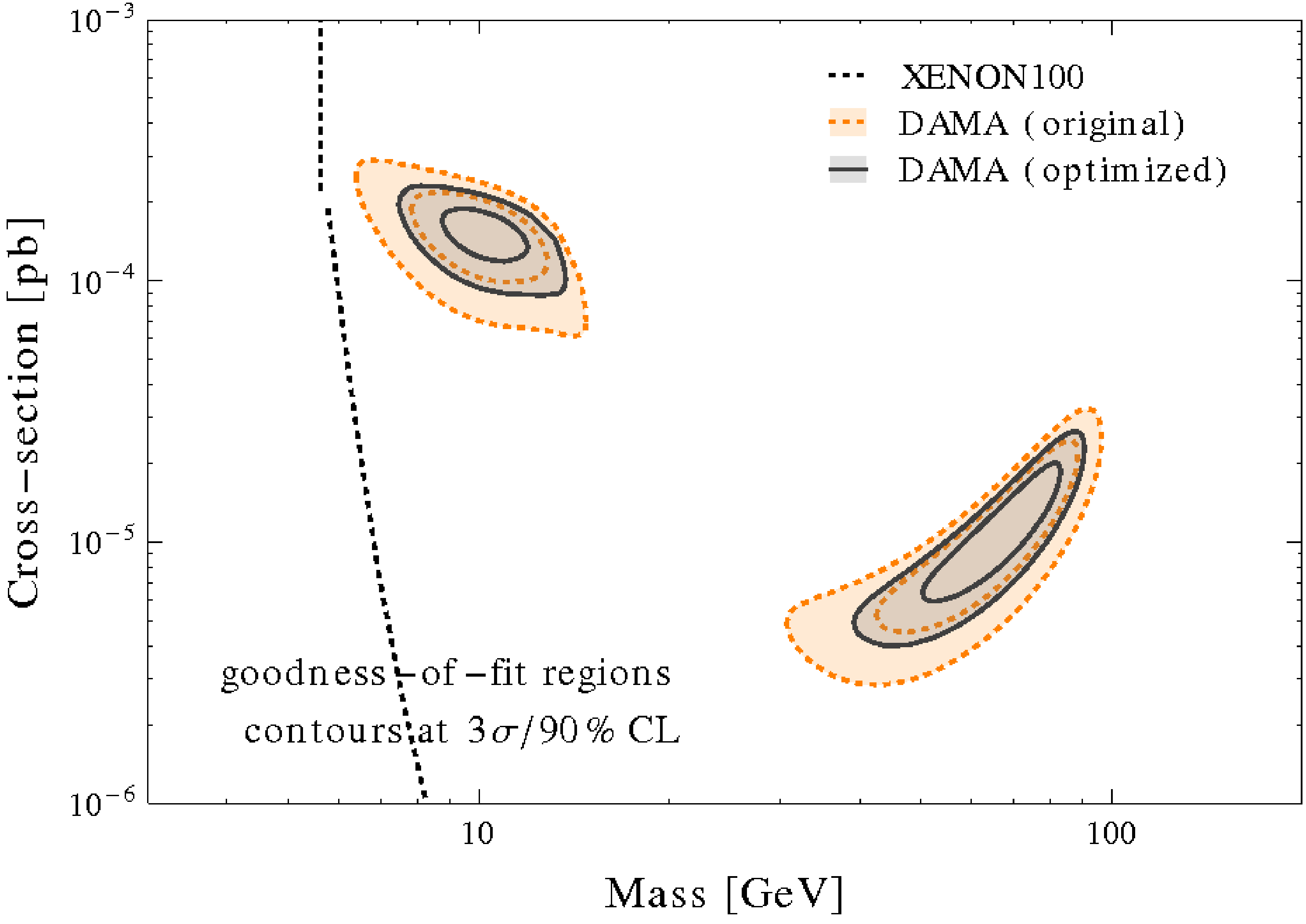
DAMA data (original)
DAMA data (optimized)
Phase: June 2
Modulation Amplitude [cpd/kg/keVee]
Energy [keVee]
0.04
0.03
0.02
0.01
0.00
−0.01
0
2
4
6
8

Binning



DAMA data (original)
68 GeV, $\chi_r^2 = 34.5/34$
10.0 GeV, $\chi_r^2 = 37.0/34$
DAMA data (optimized)
68 GeV, $\chi_r^2 = 7.3/6$
10.1 GeV, $\chi_r^2 = 9.7/6$
Phase: June 2
Modulation Amplitude [cpd/kg/keVee]
Energy [keVee]
0.04
0.03
0.02
0.01
0.00
−0.01
0
2
4
6
8

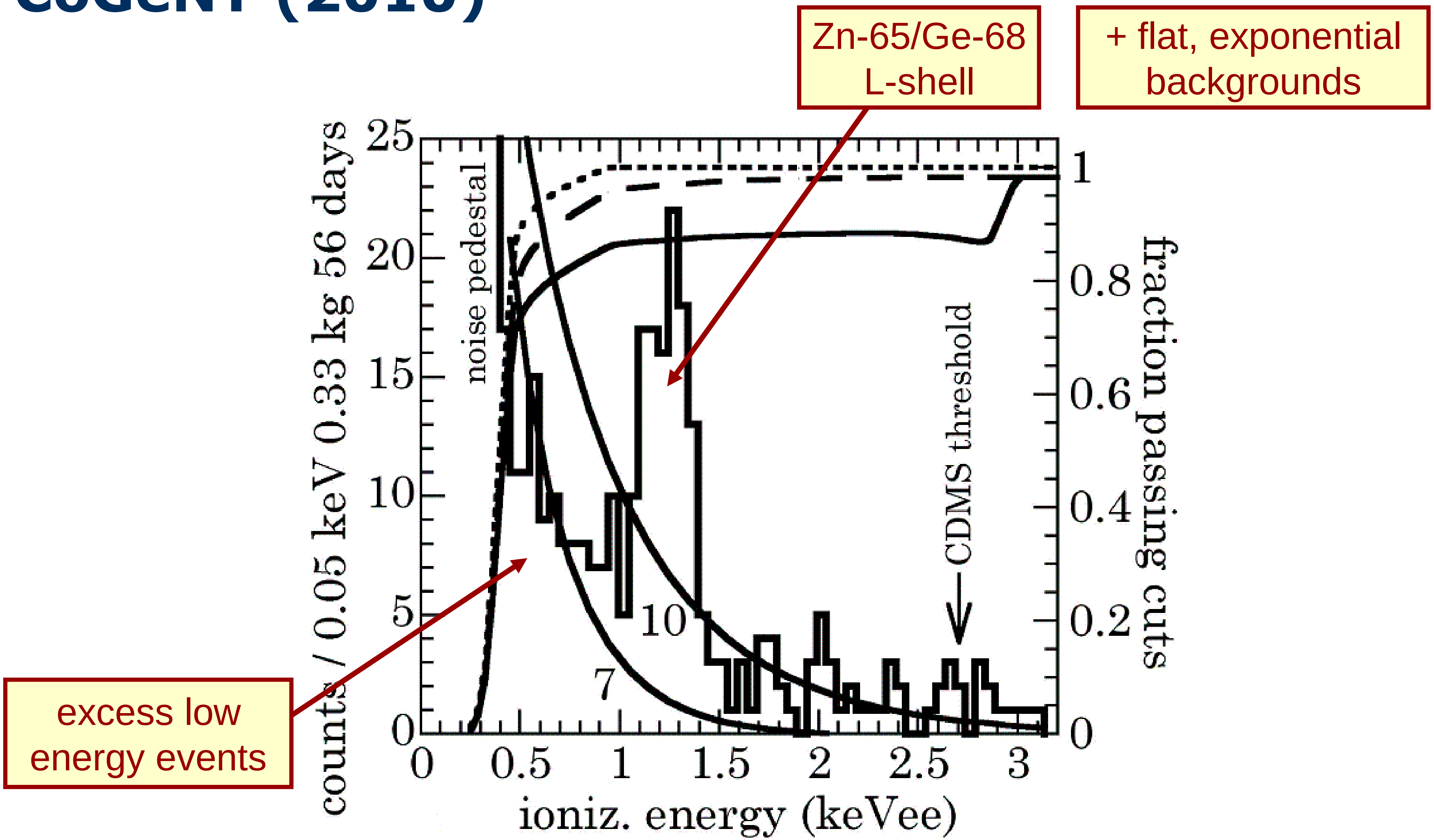
Binning



XENON100
DAMA (original)
DAMA (optimized)
goodness-of-fit regions
contours at 3σ/90% CL
Cross-section [pb]
Mass [GeV]
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10
100

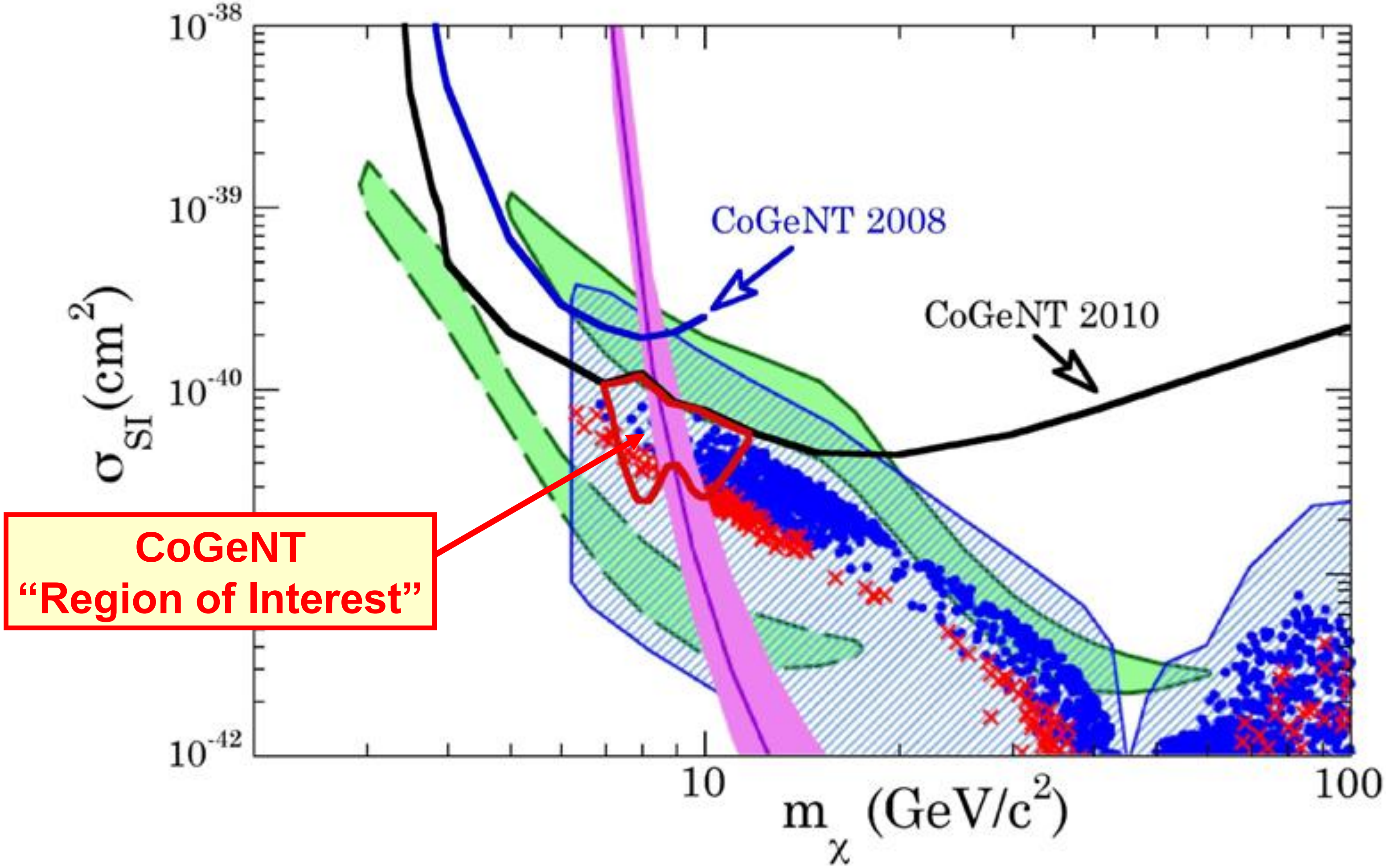
Misleading/Flawed Statistics

CoGeNT (2010)



Aalseth et al. [CoGeNT], PRL **106**, 131301 (2011)

CoGeNT (2010)

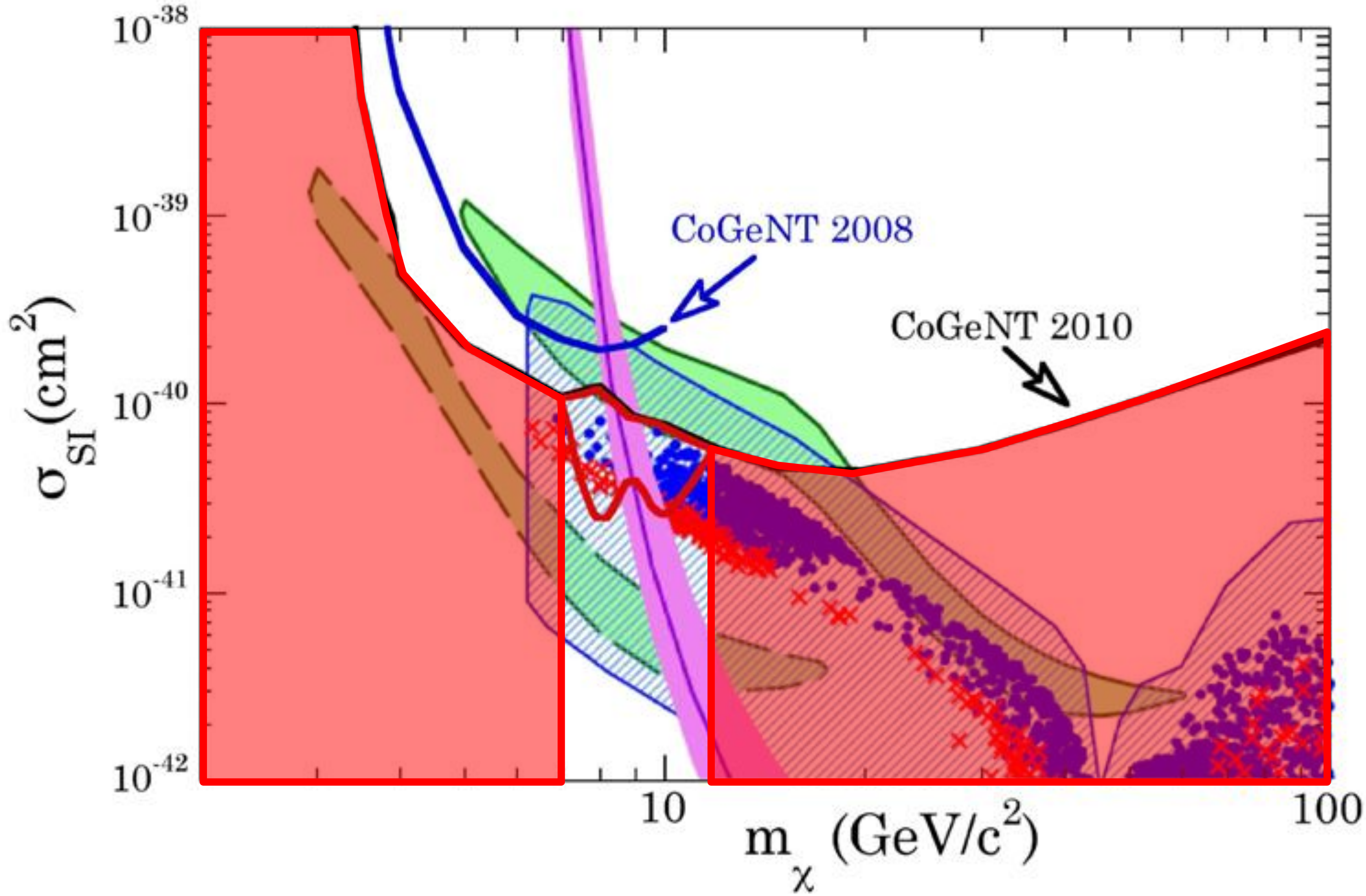


What do regions mean?

CoGeNT (2010): "Region of Interest"

- At each mass, minimize cross-section and determine 90% CL confidence interval
- Raster scan: provides no constraint on mass
- What is 90% CL *region*?

CoGeNT (2010)



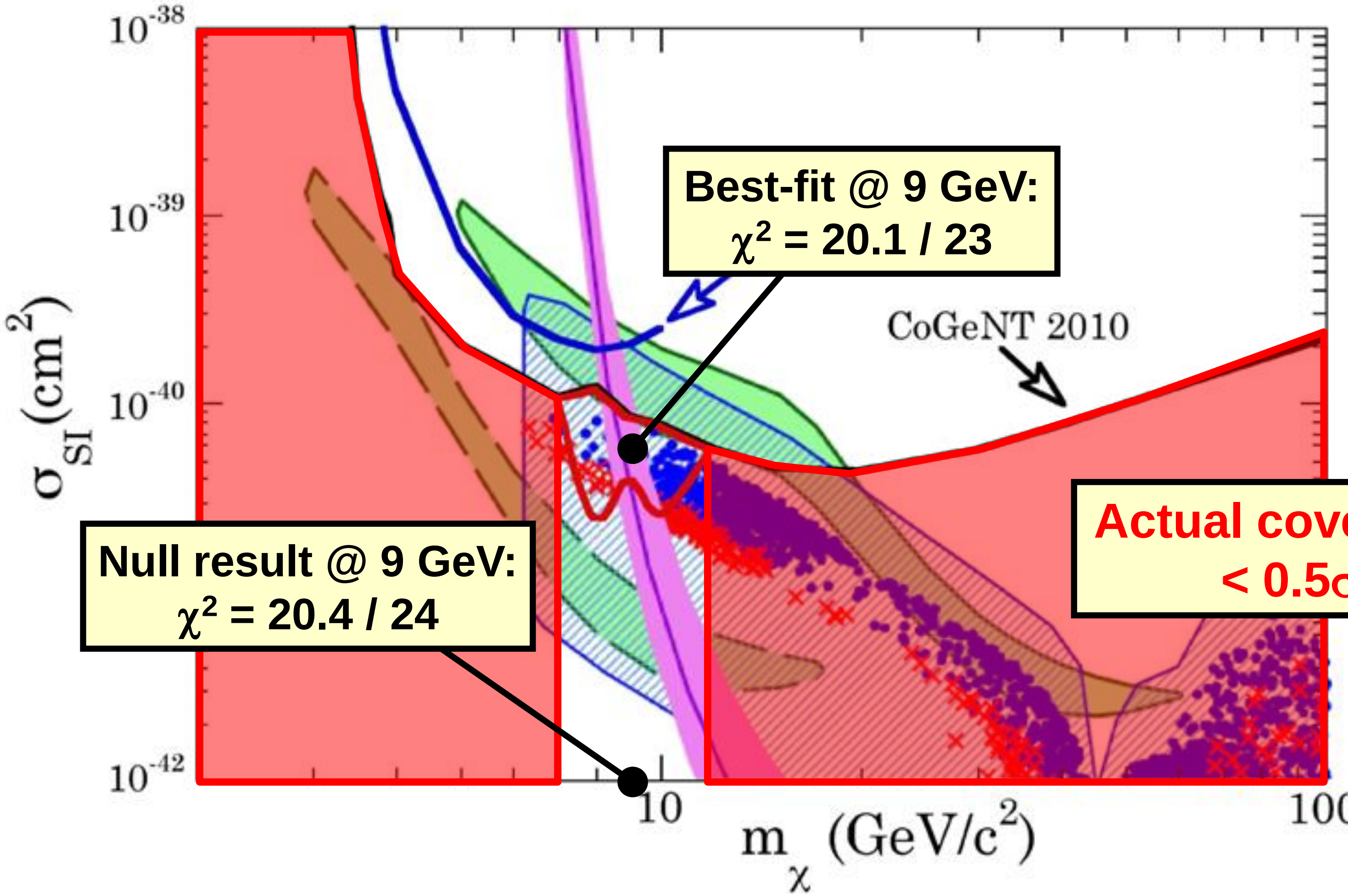
Raster scan 90% CL region

CoGeNT (2010): Confidence Interval

Figure 4 (top) displays the extracted sensitivity in spin-independent coupling (σ_{SI}) vs WIMP mass (m_χ). For m_χ in the range ~7–11 GeV/c^2 the best-fit to the WIMP coupling acquires a value with a lower 90% C.L. interval incompatible with zero. The upper and lower 90% C.L. intervals for the coupling define the red contour in Fig. 4. For example, the best fit for $m_\chi = 9$ GeV/c^2 provides a reduced chi-square $\chi^2/dof = 20.1/18$ at $\sigma_{SI} = 6.7(\pm 1.2) \times 10^{-41}$ cm^2. However, the null hypothesis (same background model minus the WIMP response) yields a similar $\chi^2/dof = 20.4/20$, the result of the WIMP response being nearly exponential in shape. Based on the quality of fits, it is not possible to distinguish between an unknown background and dark matter.

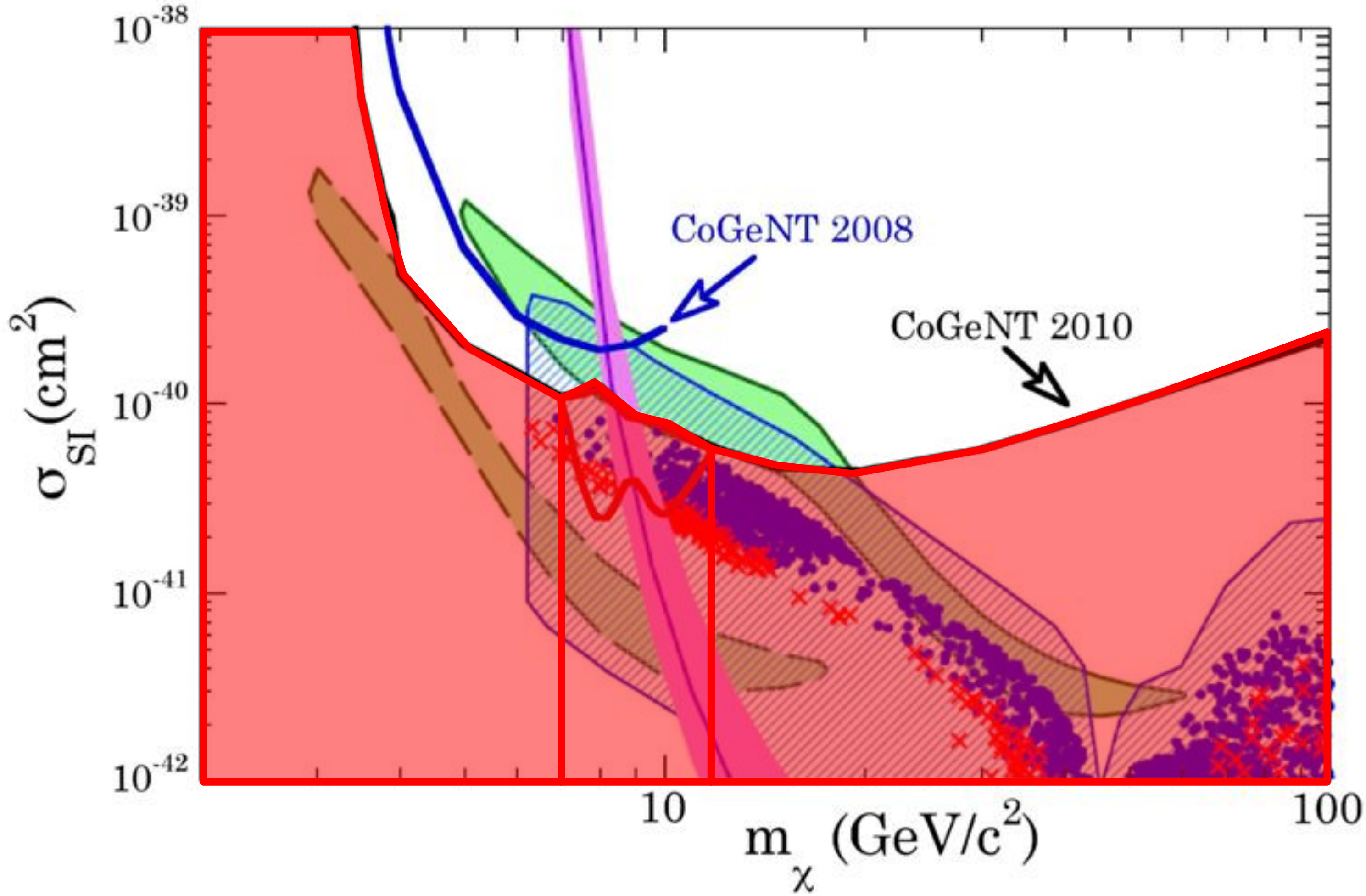
Aalseth et al. [CoGeNT], PRL **106**, 131301 (2011)

CoGeNT (2010)



for 90% CL coverage: $\Delta\chi^2 < 2.7$ (actual: $\Delta\chi^2 = 0.3$)

CoGeNT (2010)



Raster scan 90% CL region (corrected)

CoGeNT (2010)

WIMP minimized over M & σ $\rightarrow \chi^2 = 20.1 / 22$

No WIMP (null hypothesis) $\rightarrow \chi^2 = 20.4 / 24$

Hypothesis ratio test:

$\Delta\chi^2 = 0.3$ on 2 d.o.f. (for 1σ: $\Delta\chi^2 = 2.3$)

$\Rightarrow$ No preference for WIMPs!

"Region of Interest" is not very interesting (statistically speaking)

How to get it back (2011+): drop exponential background contribution

CoGeNT (2010)

What happened?

Chris, the granularity in the plot for that mass region is very coarse, at 6, 7, 8, 9, 10, 12, 15 GeV so do not give much meaning to the narrowing of the shape in the middle (9 GeV), I think that is what you are saying. ***Using the NonlinearModelFit function in Mathematica with ParameterConfidenceIntervalTable option*** and the details we give in the handout we obtain lower 90% CL boundaries that are incompatible with zero, elsewhere in WIMP mass they are. As you say (and we say in the paper), this is not the same as a proper chi-square analysis.

J. Collar (2010)

Mathematica description:

- `NonlinearModelFit` produces a nonlinear model of the form $\hat{y} = f[x_1, \ldots, \beta_1, \ldots]$ under the assumption that the original y_i are independent normally distributed with mean $\hat{y}_i$ and common standard deviation.

SOFTWARE BLACK BOX

CoGeNT (2010)

CoGeNT analysis reference

We mention the reduced chi-square in the paper with the intention to show that any exponential background can fit the data (see transparencies from our DM10 talk, available 3/1/10, for more details on our general attitude towards this "signal"). Hopefully that message is clear, please let us know if it isn't. In the case of a model with no WIMPs and for our particular choice of binning, one has d.o.f.= 28 bins - 4 free parameters - 4 constraints = 20. The four free parameters are in that case the amplitude of the combination of L-shell peaks (see next section), two more for the exponential (amplitude and exponent) and the free constant background. Incidentally, if one drops the 4 constraints (physicality in the sign of the parameters), the fit remains the same, but the quoted reduced chi-square would be slightly relaxed. As for the model when the WIMP component is introduced, d.o.f.= 28 -5- 5 =18 (i.e., one extra parameter, the WIMP coupling, and its corresponding physicality in the sign constraint).

How many degrees of freedom?

- Bins – number of parameters fit – constraints on parameters
- Constraints on *parameters* do not improve fit (can actually *increase* d.o.f.)
- Makes fits look worse

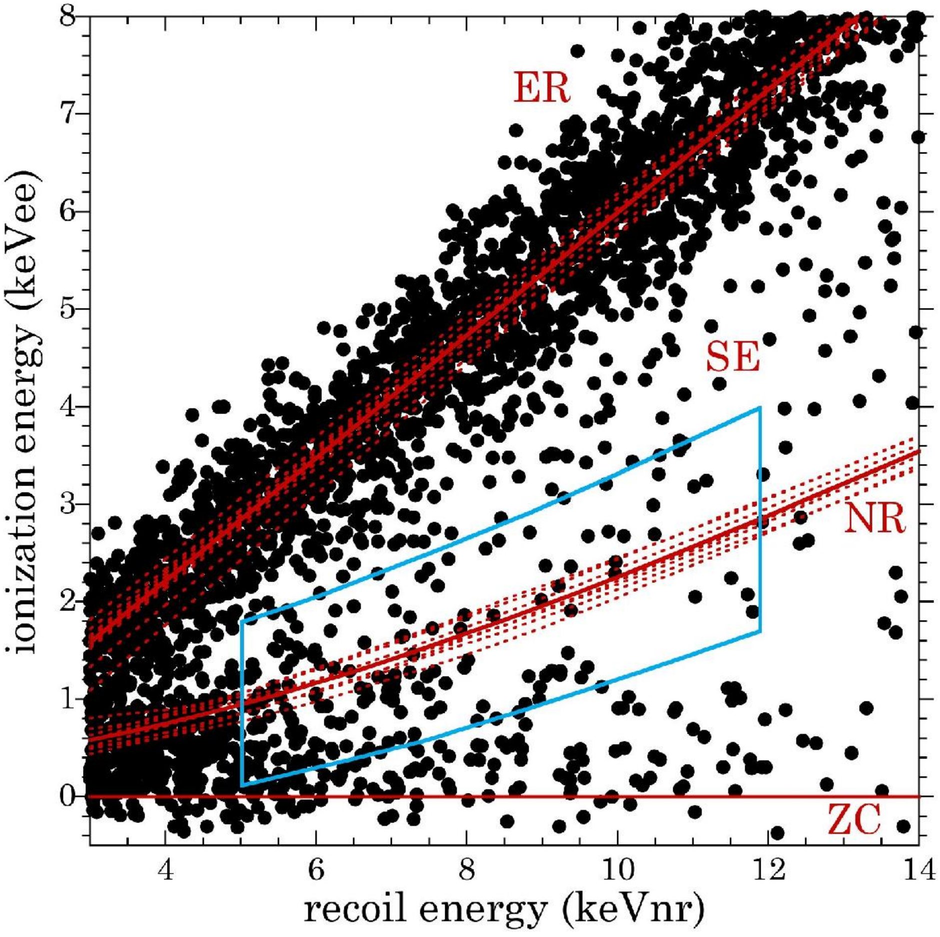
Missing/Incomplete Statistics

CDMS low-energy analysis

A Maximum Likelihood Analysis of Low-Energy CDMS Data, Collar & Fields, arxiv:1204.3559.

- Recoil energy & charge for each event
- Model backgrounds:
  - Electron recoils
  - Surface events
  - Zero charge (edge events)
- +nuclear recoils (signal)

⇒ **Likelihood analysis**



CDMS low-energy analysis

Ahmed et al. [CDMS Collaboration], PRL **106**, 131302 (2011):

- Low-energy events consistent with background estimates
- …but calculate no background-subtraction constraints

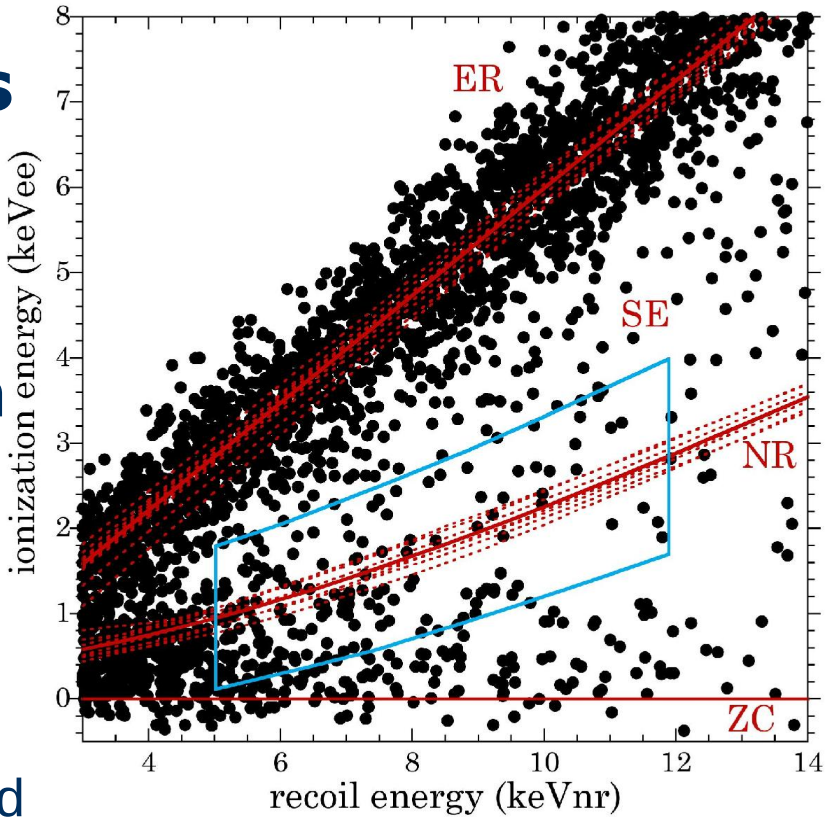
Collar & Fields:

- Fit each detector ($\times$8) and combined fit with and without NR's
- 5.7σ preference for (WIMP-like) nuclear recoil population over null hypothesis in their model
- Estimate NR rate above 3 keVee (each detector and combined)

Is there a 5σ detection in CDMS?

CDMS low-energy analysis

- Hypothesis test
- Parameter/interval estimation
- **…goodness-of-fit?**



Why goodness-of-fit?

- ~ 6000 events, ~ 200 “signal” events
- Signal region: in long tail of background (expect significant contamination)
- If model of tail is off, likelihood analysis is biased towards finding signal *even if it is not there*

Evidence of signal or of just less bad of two poorly fit models?

CDMS low-energy analysis

Collar & Fields results

Detector	Exposure (kg-day)	-2 ln(Λ)/Δd.o.f.	p-value	S_1(eVee)	NR rate > 3 keVnr (c/kg-day)	A_2^{NR} (keVnr^{-1})
all	240.4	37.8 / 5	$4.1 \cdot 10^{-7}$	303	0.93 ± 0.17	0.650 ± 0.081
all−T3Z5	214.6	45.9 / 5	$9.5 \cdot 10^{-9}$	266	0.95 ± 0.14	0.701 ± 0.079
T1Z2	43.4	13.5 / 2	$1.1 \cdot 10^{-3}$	220	0.77 ± 0.20	0.569 ± 0.139
T1Z5	35.0	12.5 / 2	$1.9 \cdot 10^{-3}$	201	0.50 ± 0.17	0.714 ± 0.299
T2Z3	28.0	5.7 / 2	$5.7 \cdot 10^{-2}$	246	1.31 ± 0.68	0.659 ± 0.222
T2Z5	34.7	2.7 / 2	$2.6 \cdot 10^{-1}$	442	0.76 ± 0.41	0.745 ± 0.355
T3Z2	7.8	3.9 / 2	$1.4 \cdot 10^{-1}$	333	3.38 ± 1.24	0.705 ± 0.215
T3Z4	29.6	12.8 / 2	$1.7 \cdot 10^{-3}$	142	0.39 ± 0.15	0.636 ± 0.202
T3Z5	25.8	0.22 / 2	$8.9 \cdot 10^{-1}$	406	0.15 ± 0.38	2.01 ± 2.44
T3Z6	36.1	3.27 / 2	$1.9 \cdot 10^{-1}$	228	0.61 ± 0.31	0.707 ± 0.361

- Global fit:
  signal rate above 3 keV is 0.93 ± 0.17 events/kg/day

- 8 independent measurements (individual detectors):
  fluctuations about global best fit $\chi^2 = 29.7/8$ (> 3σ)

CDMS low-energy analysis

If the backgrounds are correctly modeled:

…the background only case is rejected at > 5σ

…but the excess is inconsistent with WIMPs at > 3σ

Need goodness-of-fit check!

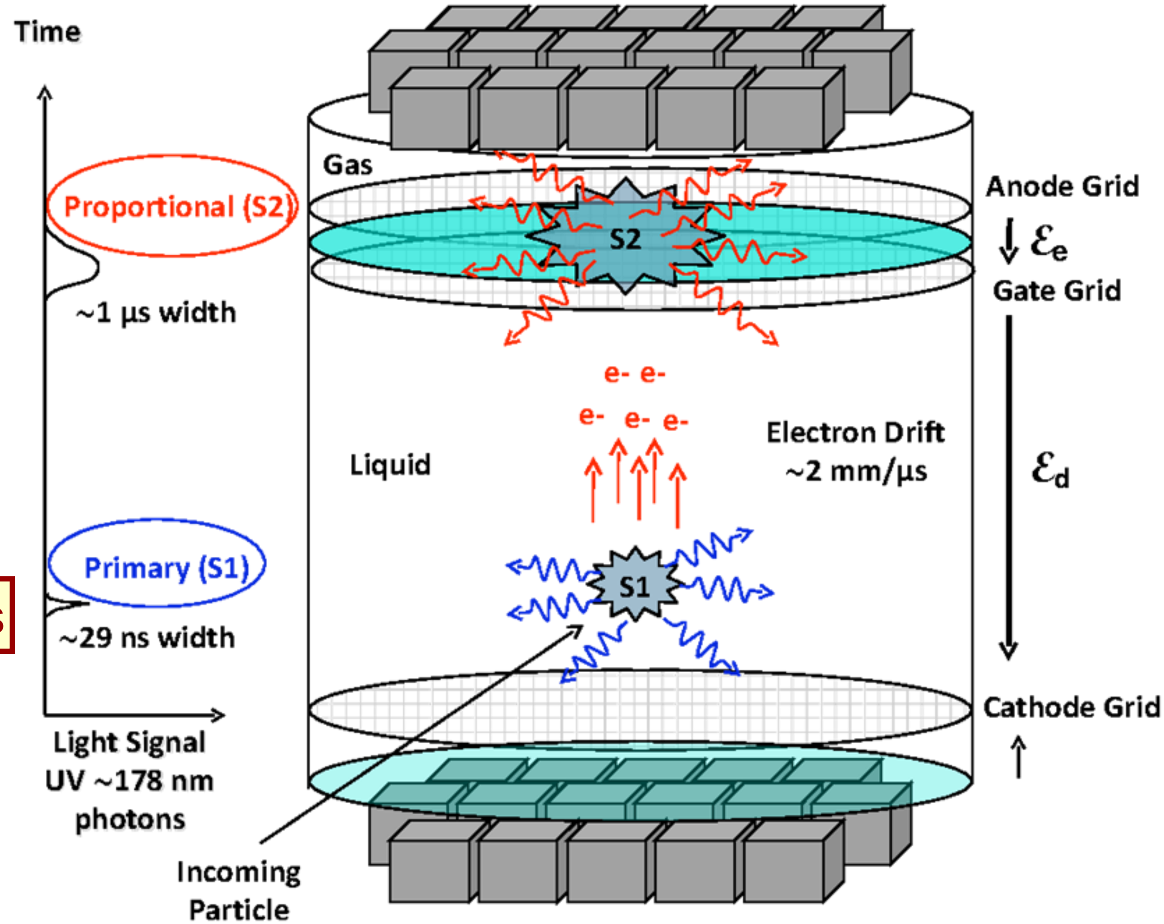
CS & Weniger (in progress)

Conservative Statistics

Liquid Xenon/Argon detectors: principles of operation

S2: ionization electrons

S1: prompt photons



Aprile et al. (2010) [XENON10]

XENON

See talk by Rafael Lang

- Poisson statistics
  - Analysis threshold: 2/3/4 photo-electrons (PEs)
  - Require at least two PMT triggers
  - Recoils produce large number of scintillation photons, but only ~ 6% trigger PMT (absorption by liquid/walls, PMT photocathode quantum efficiency)

    33 photons: average S1 = 2 PE, but sometimes 0, 1, 3, 4,…

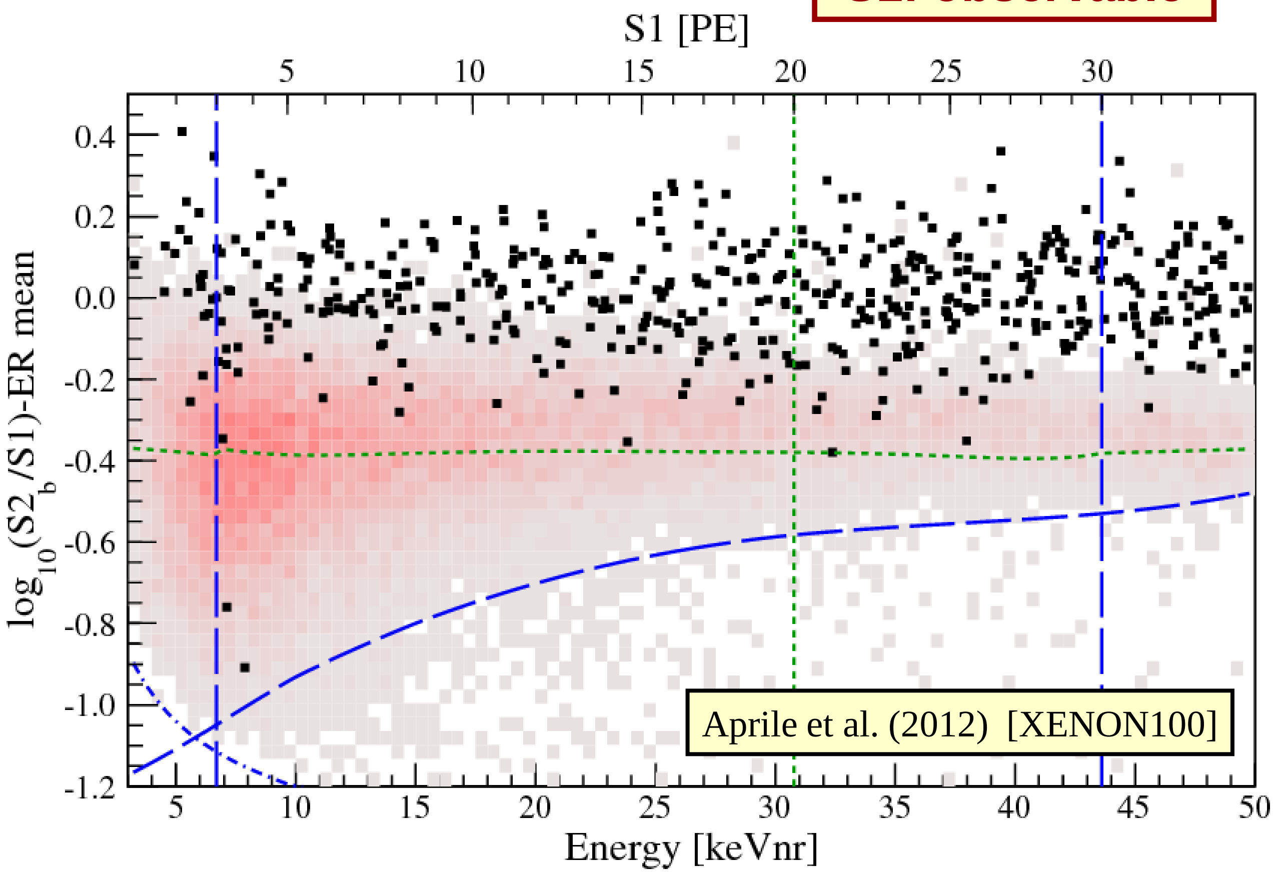
    17 photons: average S1 = 1 PE, but sometimes 0, 2, 3, 4,…

- How much scintillation in event? $\mathbf{L_{eff}}$

- Analysis “energy” range (badly labeled)

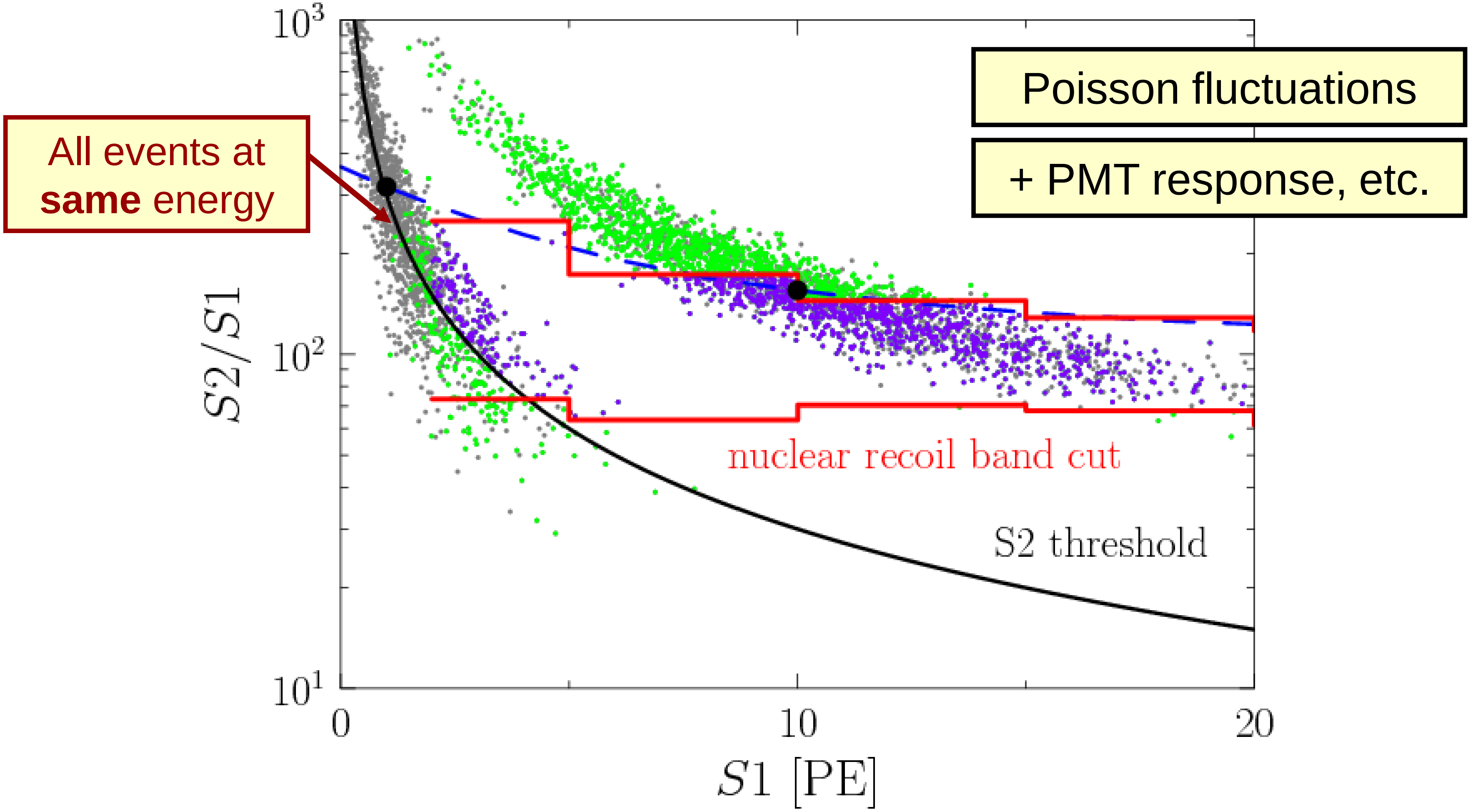
XENON Event Energies

S1: observable

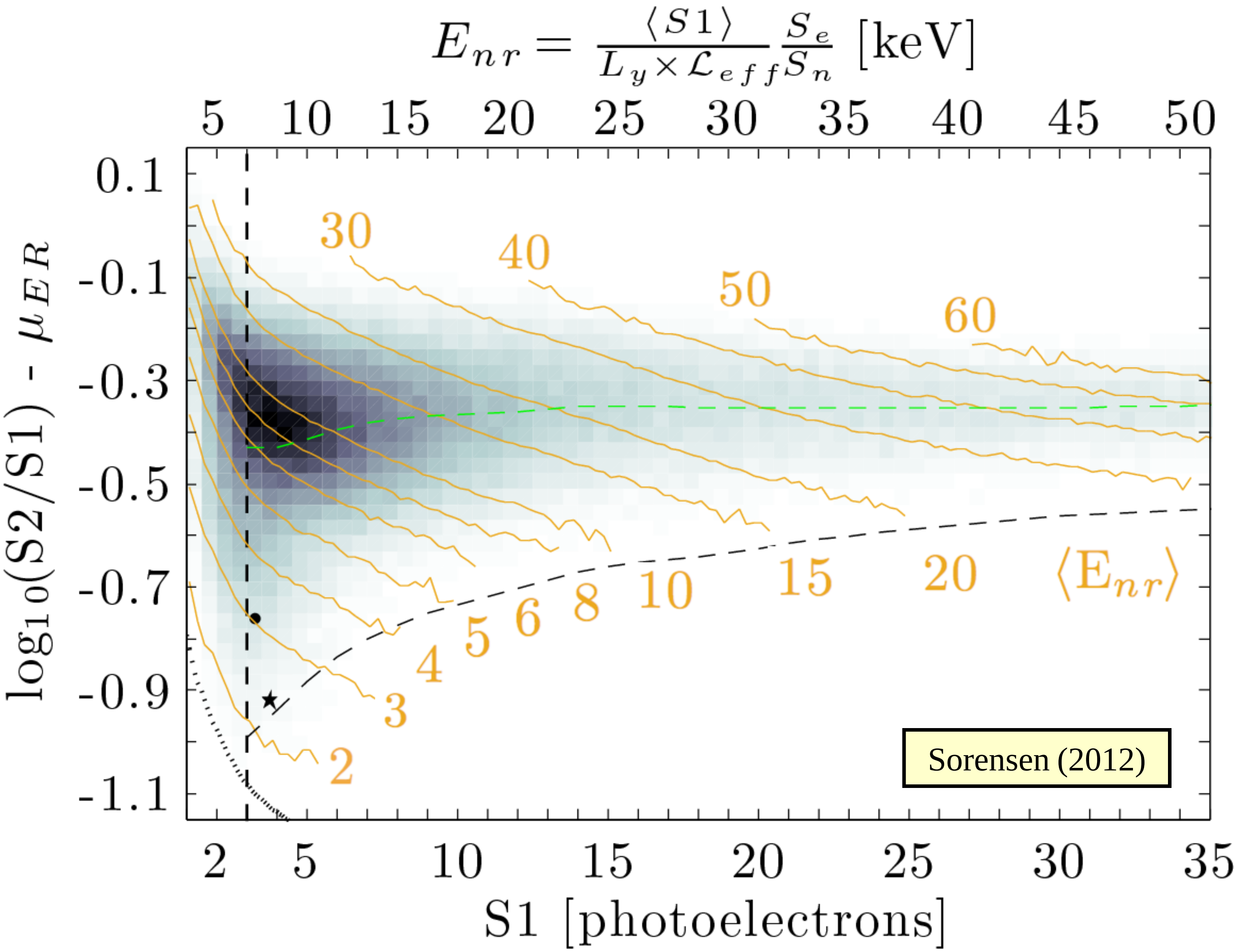


Bad energy estimator

XENON Event Energies

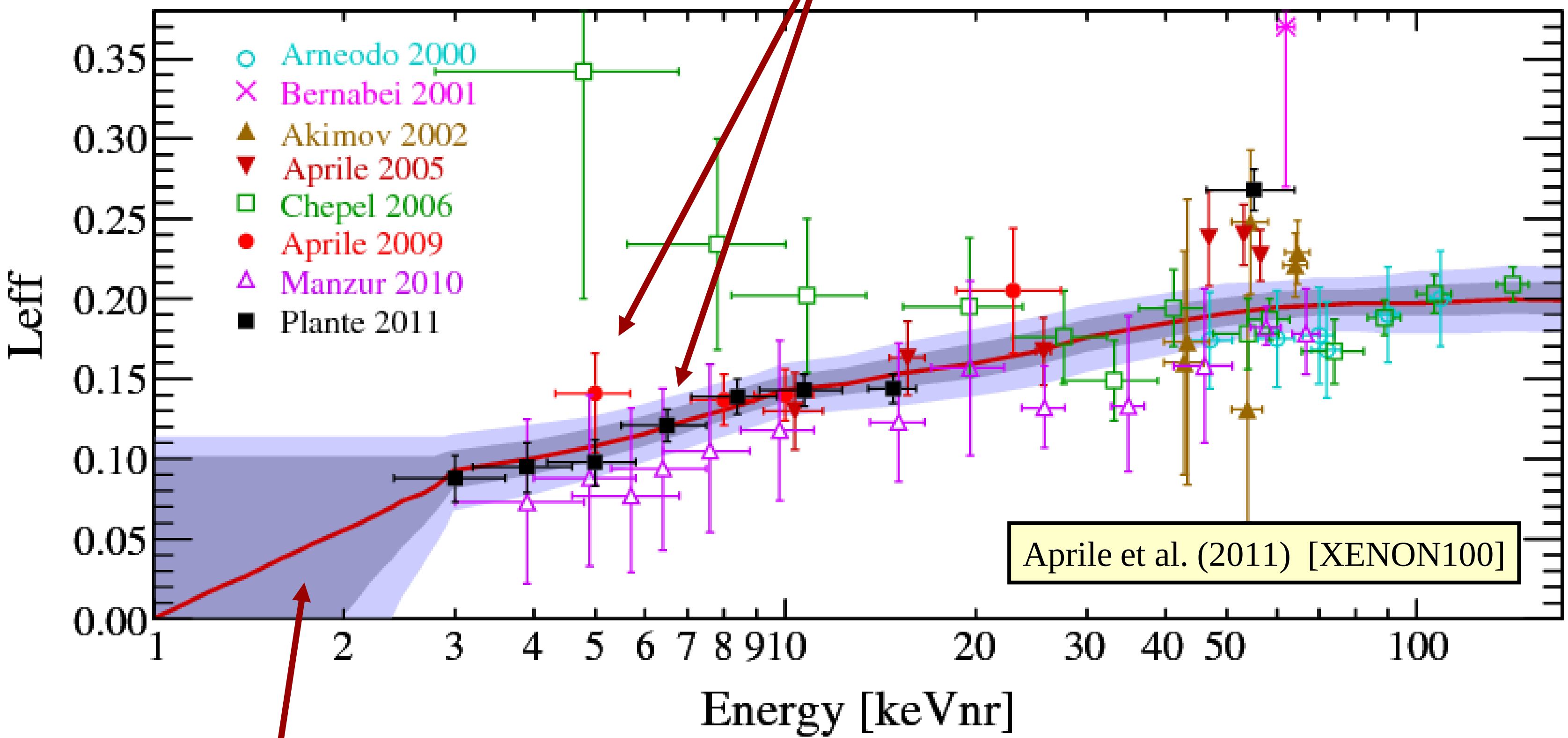


XENON Event Energies



XENON Leff

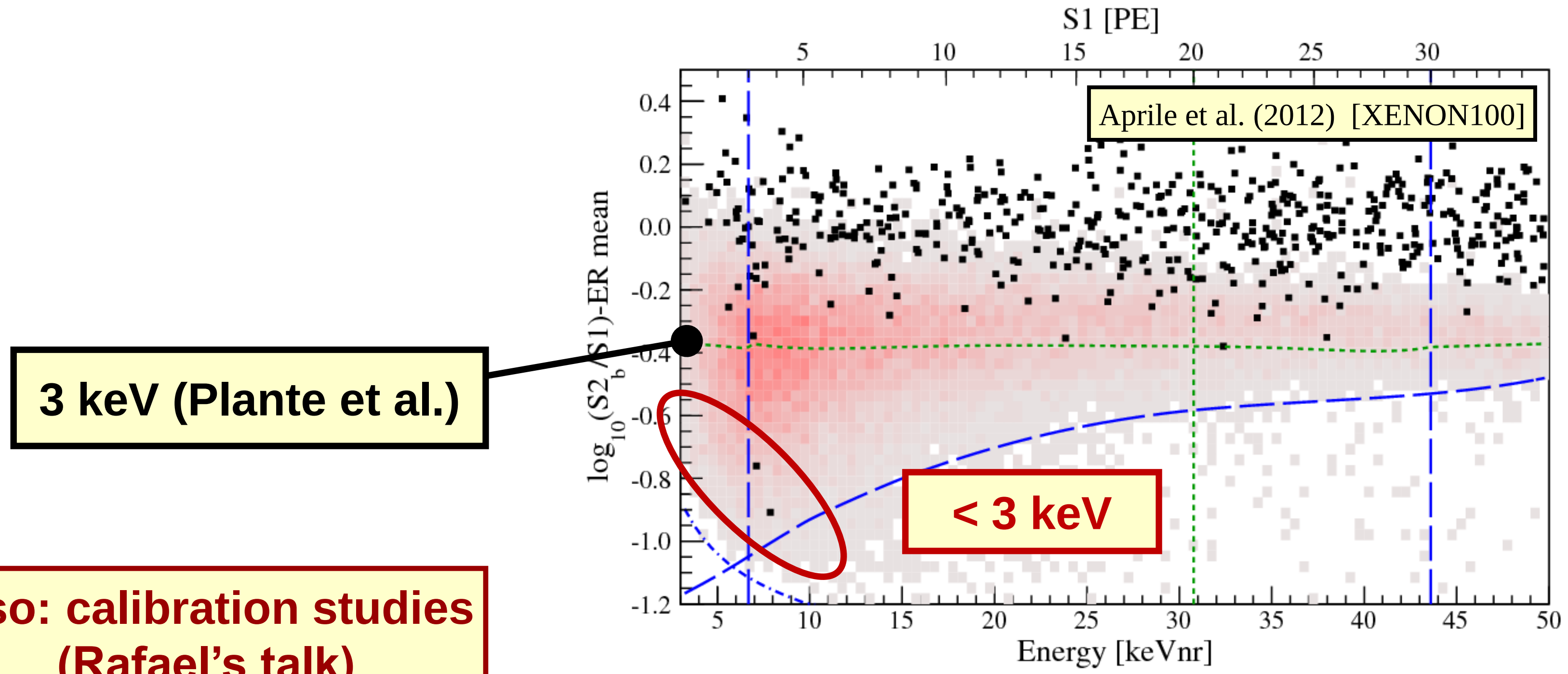
What is happening here?
(stat/systematic uncertainties)



What is happening here?
(unknown behavior)

XENON Leff

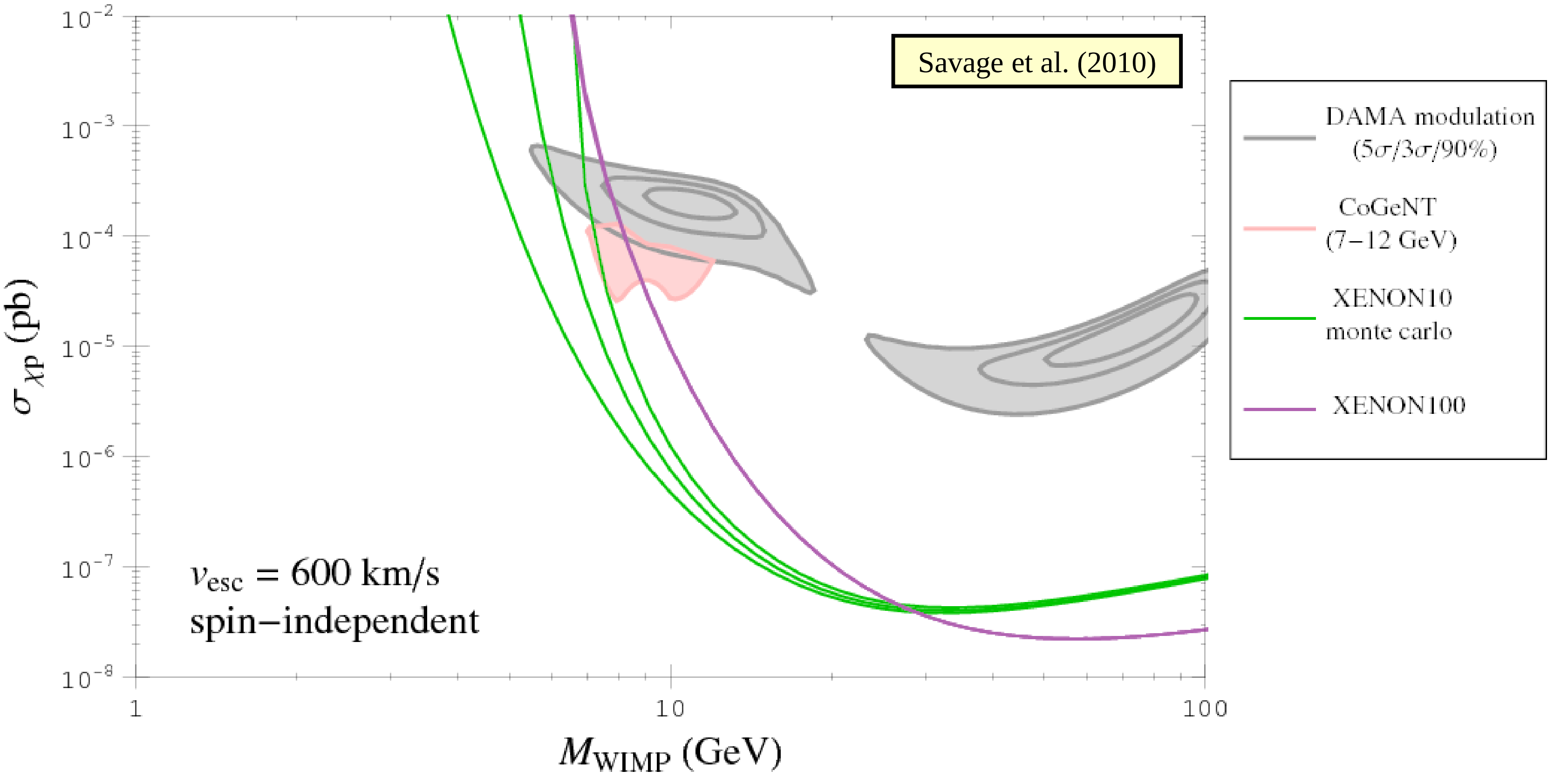
- Unknown Leff behaviour below 3 keV
- Conservative assumption: treat Leff = 0 there
- **Too** conservative?



Also: calibration studies (Rafael's talk)

XENON Leff

Caveats:
- XENON 100 day exposure with higher threshold
- Leff cutoff at 3.9 keV (now non-zero to at least 3 keV)



Left to right: constant, falling, zero $\mathcal{L}_{eff}$ below 3.9 keVnr

XENON Leff

- Can Leff uncertainties be used to reconcile experimental results?
- Assumptions are already very conservative (and *known* to be overly conservative)
  $\Rightarrow$ constraints almost certainly cannot be made weaker
- Constraints are *very probably* significantly better at low masses
- Conservative Leff + no upward Poisson fluctuations: overkill

Summary and Remarks

Weak statistics and overly conservative assumptions do not imply compatibility

Goodness-of-fit

(DO IT)