NEPPSR Project

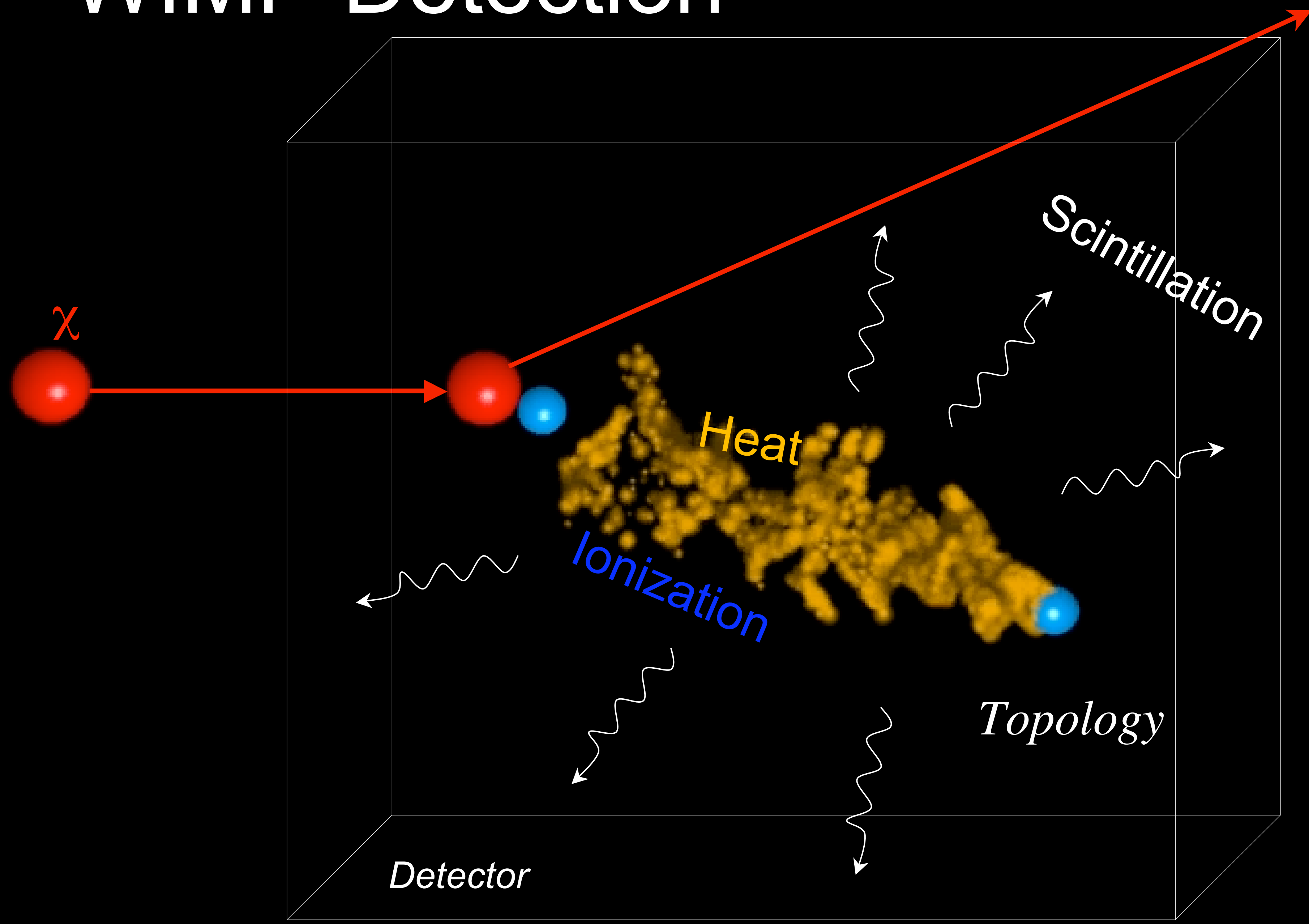
Denis Dujmic (MIT)

Project: Signal Sensitivity in Low Rate Experiments

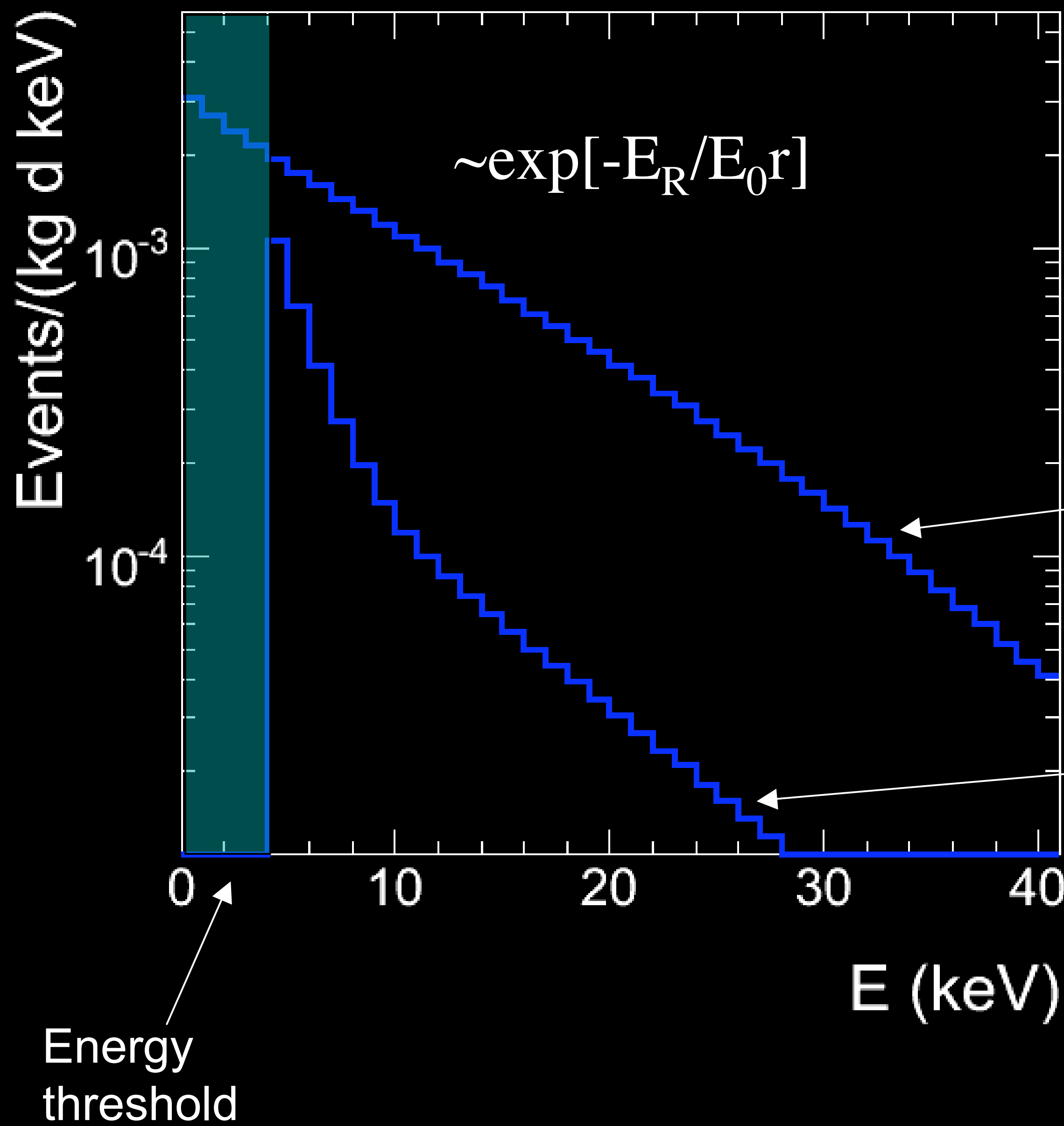
- Analysis of small samples
  - Set a limit on detector sensitivity for low-background experiments
- ROOT
  – Create, analyze, display, store datasets
  – Root tutorial by:
    - o Michael Betancourt
    - o Jeremy Lopez
    - o Wei Wang

Project Motivation

WIMP Detection



Recoil Rate



WIMP flux

$\boxed{\sim 3\ 10^{11}\ \text{WIMPs/kg/day}}$

(LXe, M_χ=100GeV, ρ=0.3GeV/cm^3)

Recoil rate v.s. E_R energy

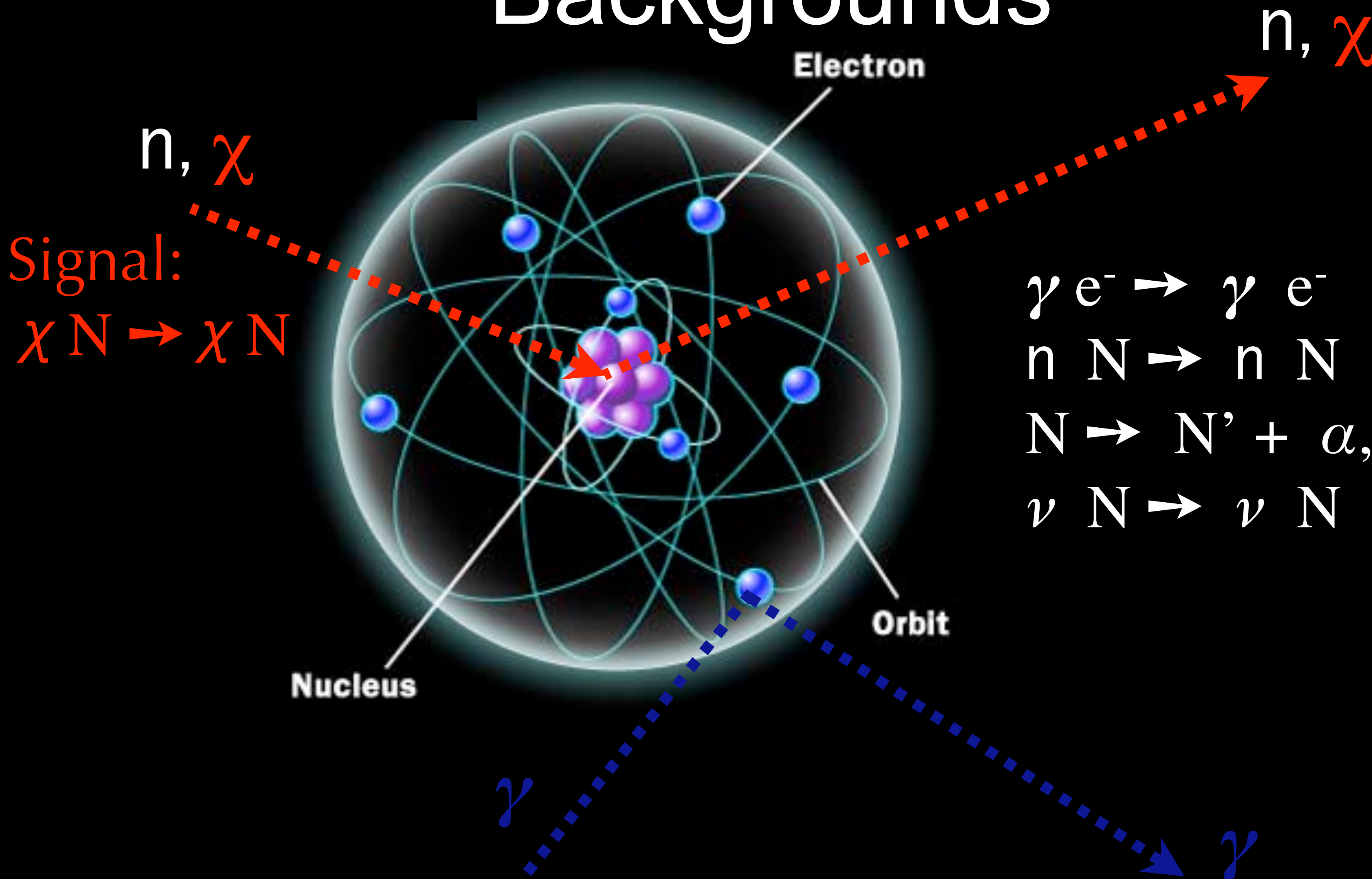
$\boxed{\sim 0.1\ \text{events/kg/day}}$

Detected rate v.s. E_{eee} electron-equivalent energy

('quenching' = E_{eee}/E_R)

$\boxed{\sim 0.01\ \text{events/kg/day}}$

Backgrounds



$\gamma\, e^- \rightarrow \gamma\, e^-$

$n\ N \rightarrow n\ N$

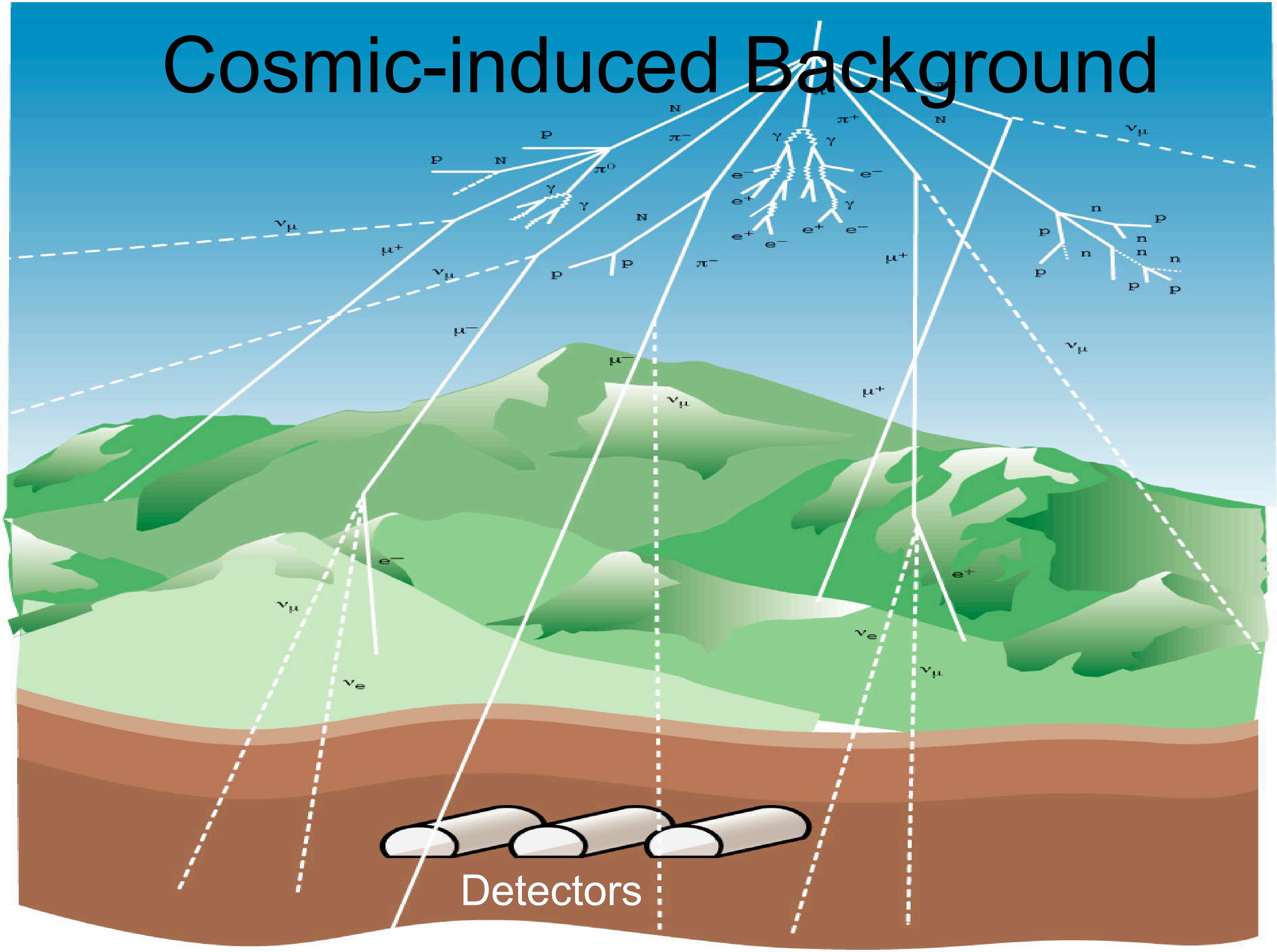
$N \rightarrow N' + \alpha, e^-$

$\nu\ N \rightarrow \nu\ N$

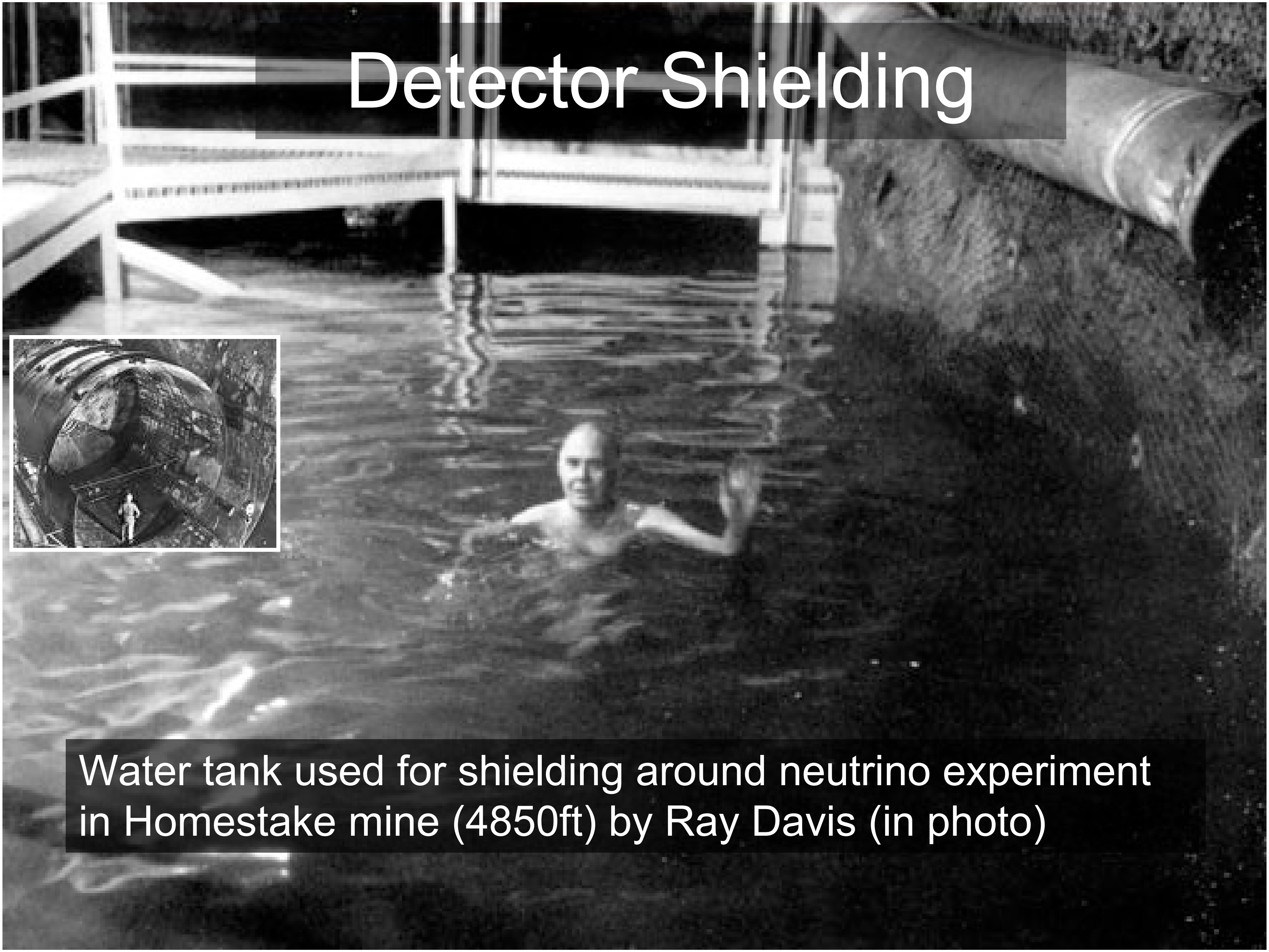
Untamed background rate ~10^6 events/(kg day)

~10^8 larger than signal !!

Cosmic-induced Background



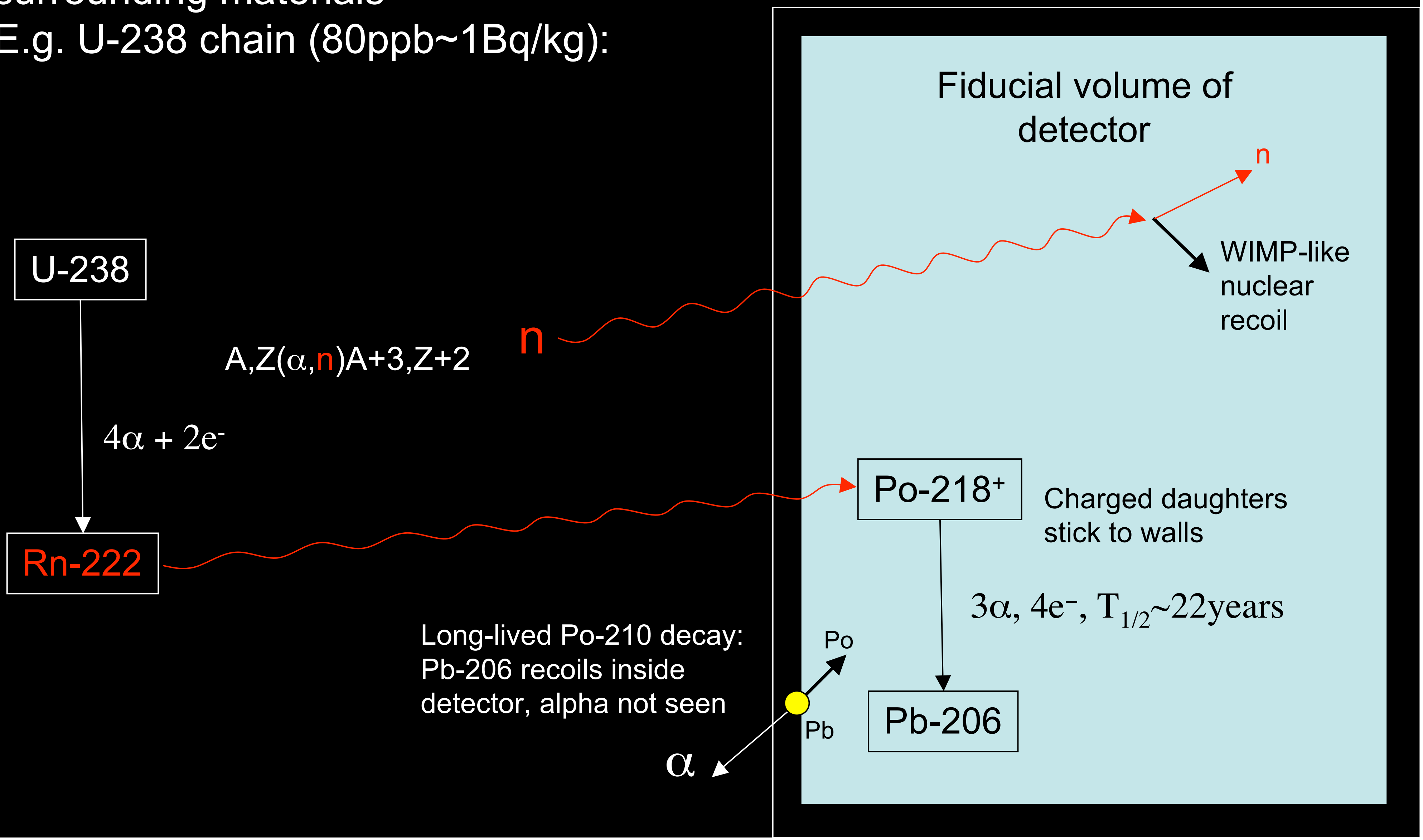
Detector Shielding



Water tank used for shielding around neutrino experiment in Homestake mine (4850ft) by Ray Davis (in photo)

Detector Radioactivity

Traces of U (~ppb), Th (~ppb), K (~ppm) contaminate detector and surrounding materials
E.g. U-238 chain (80ppb~1Bq/kg):

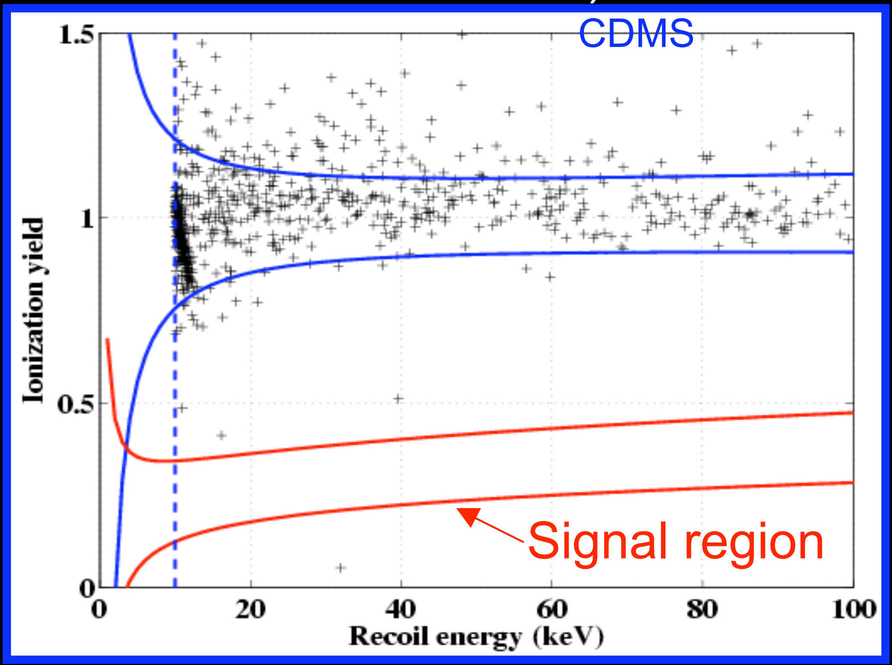


Event Signature

Further suppress background based on event signature

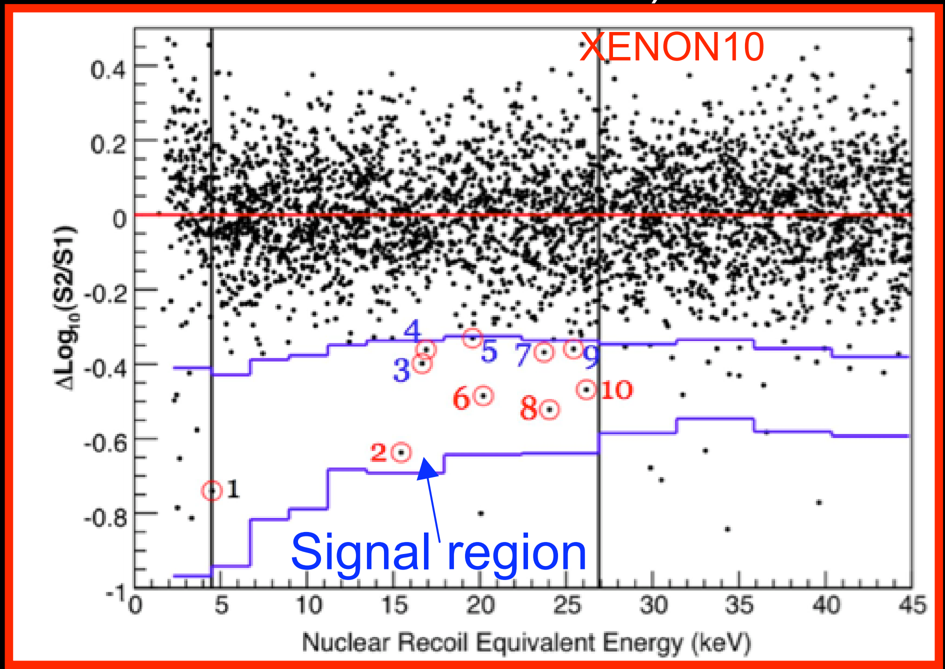
E.g. gamma suppression based on ionization/heat/scintillation signature

Ionization vs heat
(gammas with more ionization for same amount of heat)



0 observed events in signal region

Ionization vs scintillation
(gammas with less scintillation for same amount of ionization)



10 events observed in signal region

Low Statistics at NEPPSR VI

NEPPSR 2009 PROGRAM

	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
7:45		Breakfast	Breakfast	Breakfast	Breakfast
9:00	travel	Low Background Methods Pocar - UMass	Searches for New Physics Brau - UMass	free time	Astrophysics of Dark Matter Finkbeiner - Harvard
10:30		Statistics Blocker - Brandeis	Neutrino Beam Experiments Wascko - ICL		Beyond the SM Weiner - NYU
11:45	lunch	lunch	lunch		adjourn
1:00	Dark Matter Gaitskell - Brown	Low Energy Probes of New Physics Miller - Boston U.	Natural Neutrinos Formaggio - MIT		
2:15	Project Dujmic - MIT	project time	Energy Loss Fisher - MIT	QCD Surrow - MIT	
3:45	SUSY Nelson - Northeastern		GUTs and Proton Decay Kearns - BU	LHC Status Brandenburg - Harvard	
5:00	Social hour	Social hour	Social hour	Social hour	
5:45	Dinner	Dinner	Dinner	Dinner	
8:00	Root Tutorials	Student Seminars	Round Table	Student Seminars	

NEPPSR 09 Project

Analysis of Small Samples

• Assume we see few surviving events after all detector and analysis cuts

• The statement on WIMP rate depends on what we know about *background(s)*:

Rate known? (b)	Distribution known? (db/dE)	Analysis method
no	no	Poisson
no	no	Maximum Gap (Yellin)
yes	yes/no	Feldman-Cousins
no	yes	Maximum likelihood

Poisson Limit

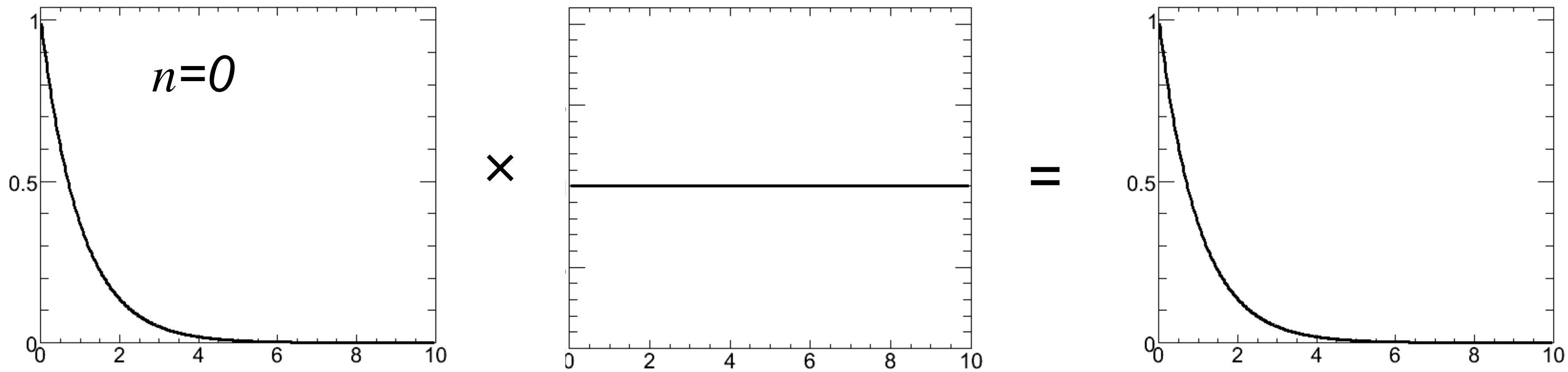
Suppose background rate and distribution not known
Allowing possibility that all events can be signal - obtain a Bayesian upper limit on number of signal events:

$$CL = \int p(theory \mid experiment) = \frac{\int \overbrace{p(experiment \mid theory)}^{\textit{Poisson}} \, \overbrace{\pi(theory)}^{\textit{Prior}}}{\pi(data)}$$

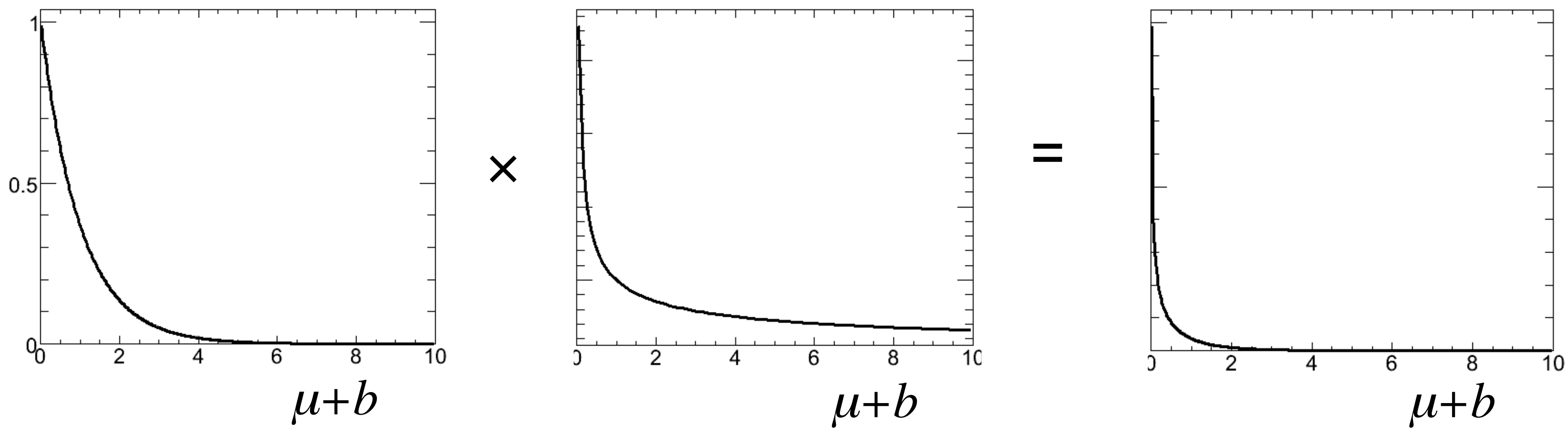
$$CL(90\%) = \int_0^{(\mu+b)_{90}} \underbrace{\frac{(\mu+b)^n e^{-(\mu+b)}}{n!}}_{\text{Probability for observing n events if } \mu+b \text{ expected}} \cdot \underbrace{\pi(\mu+b)}_{\text{Prior?}} \, d(\mu+b)$$

Poisson Limit

Prior flat in μ+b:

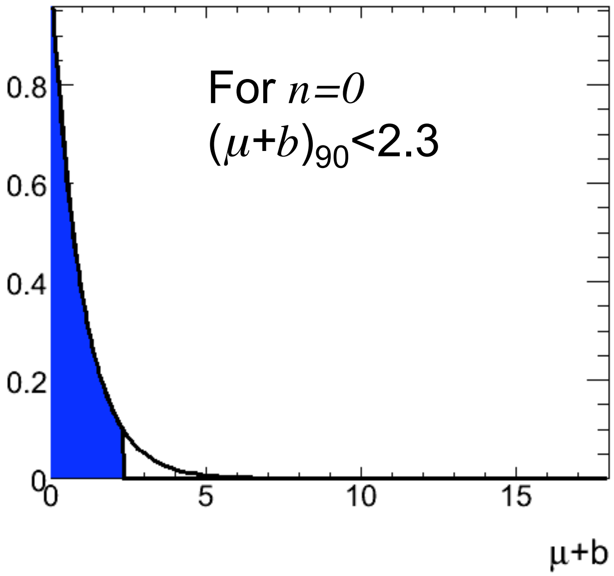


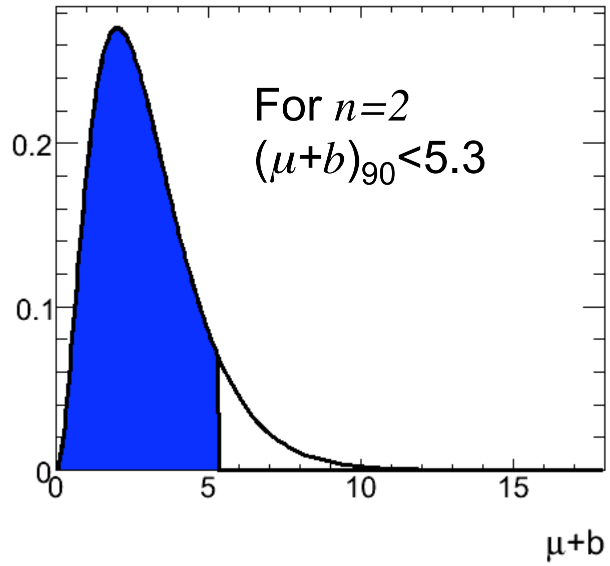
Prior flat in $(\mu\text{+b})^{0.5}$:

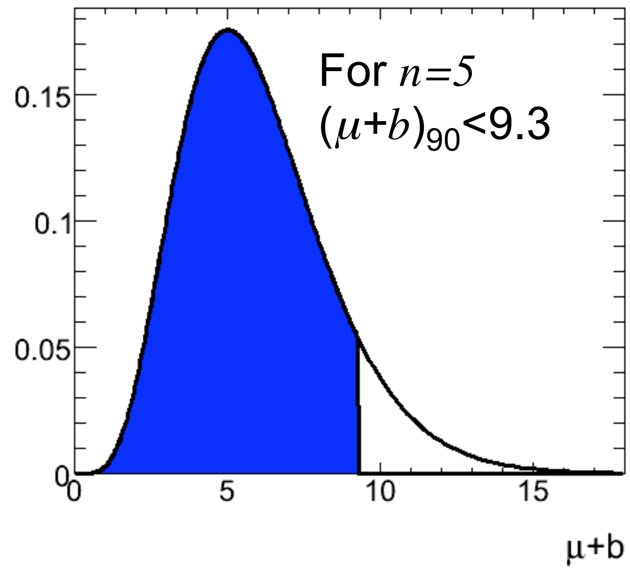


Poisson Limit

Prior flat in μ+b:







Prior flat in $(\mu+b)^{0.5}$:

$(\mu+b)_{90}<1.4$ $(\mu+b)_{90}<4.6$ $(\mu+b)_{90}<8.6$

We only assumed that count rates follow Poisson statistics
We can improve this limit with additional information on signal and background properties

Maximum Gap

What if instead of counting events that we observed, we count signal events that didn't happen ?
E.g. find the biggest gap between data points in some variable.

Here is the logic:
-If we assume too large event rate, then such gap is very unlikely
-If we assume too low event rate, then there must exist even large gap

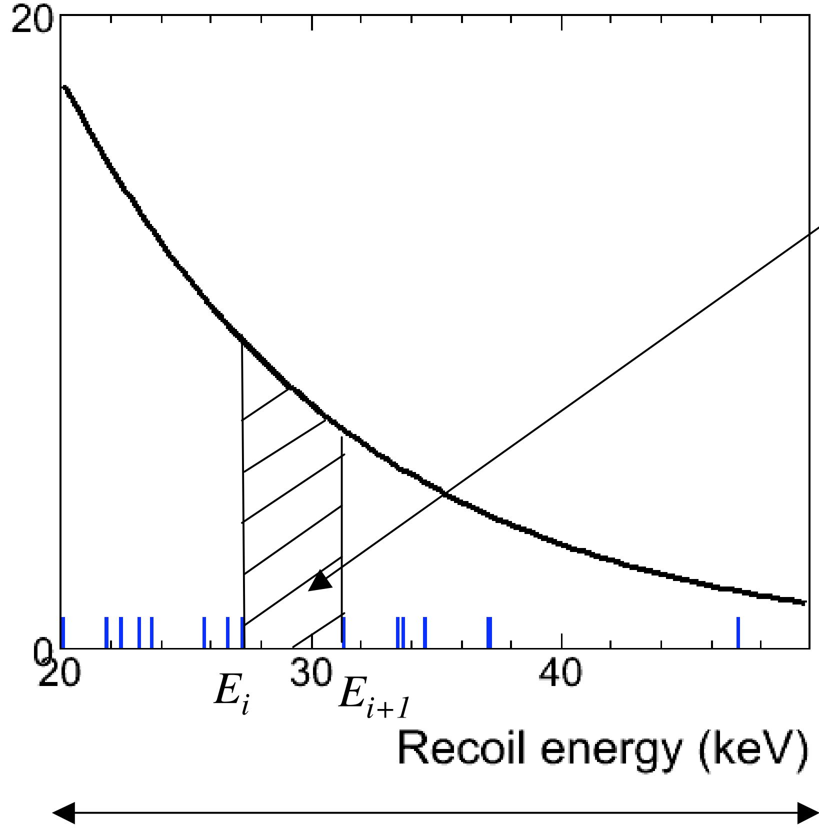
This approach is described in:

Yellin, Phys Rev D66 (2002)

(used by WIMP search experiments)

Maximum Gap

Example with recoil energies:



Find number of expected events in each energy gap

$$x_i = \int_{E_i}^{E_{i+1}} dE_R \frac{dN}{dE_R}$$

Choose gap with maximum number of expected events ('maximum gap')

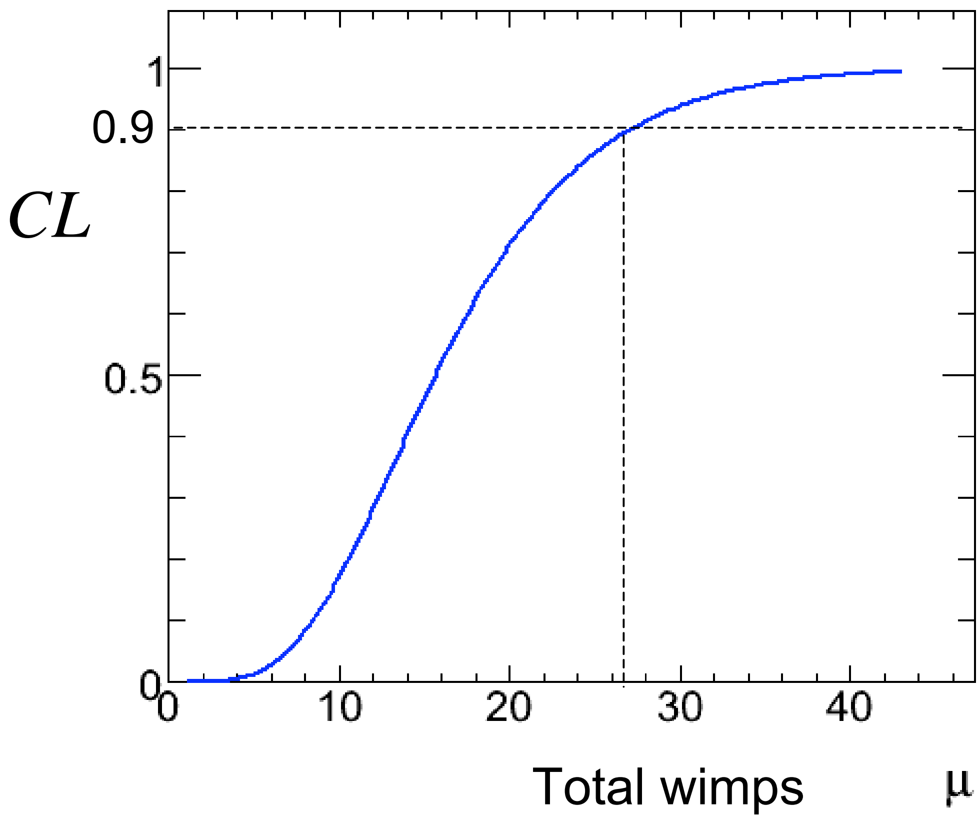
Total number of expected events

$$\mu = \int_{E_{min}}^{E_{max}} dE_R \frac{dN}{dE_R}$$

Maximum Gap

Probability of maximum gap being smaller than x (i.e. signal rate higher than expected):

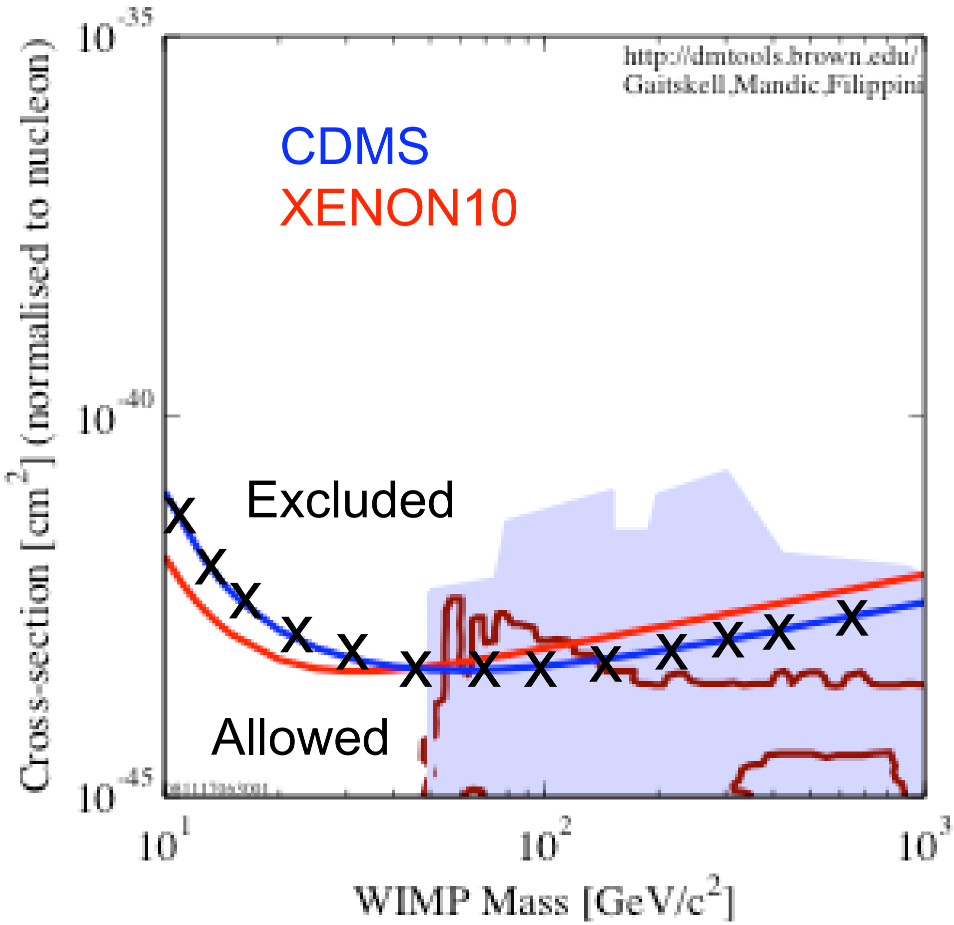
$$CL(x,\mu) = \sum_{k=0}^{m} \frac{(kx-\mu)^k e^{-kx}}{k!}\left(1+\frac{k}{\mu-kx}\right) \qquad \text{with} \qquad m \le \mu / x$$



Note, this method can only give upper limit

Maximum Gap

Setting a WIMP limit:



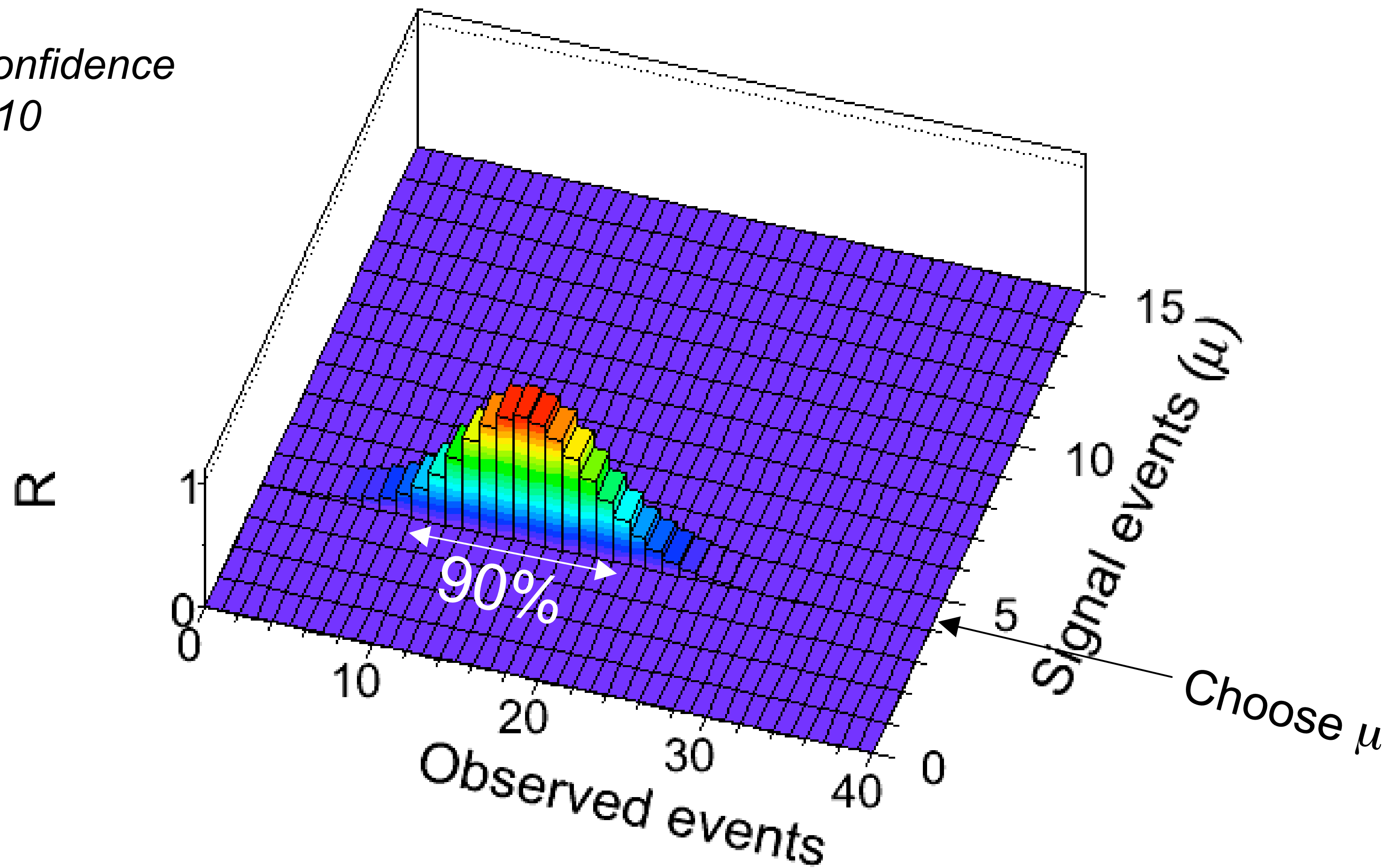
- dN(E)/dE depends on WIMP mass
- for each WIMP mass find upper limit on # of signal events using Maximum Gap method
- convert to cross section (per nucleon)

$$N_\chi = \rho_T V_T \sigma v_\chi \rho_\chi$$

Feldman-Cousins

A frequentist approach based on construction Neyman's confidence belts:
- for each physical μ, select a set that includes 90% of observed events

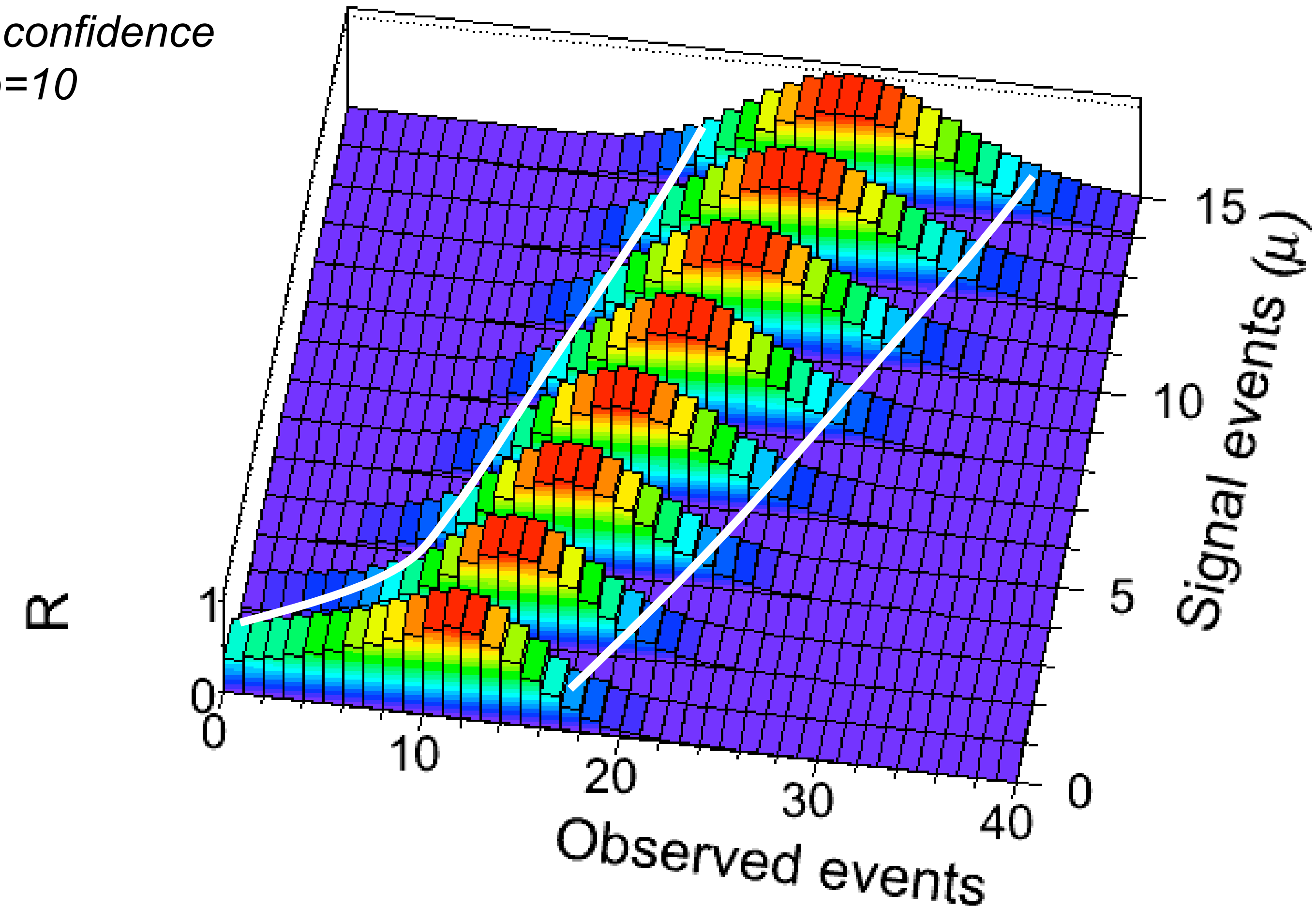
FC 90% confidence belt for b=10



Feldman-Cousins

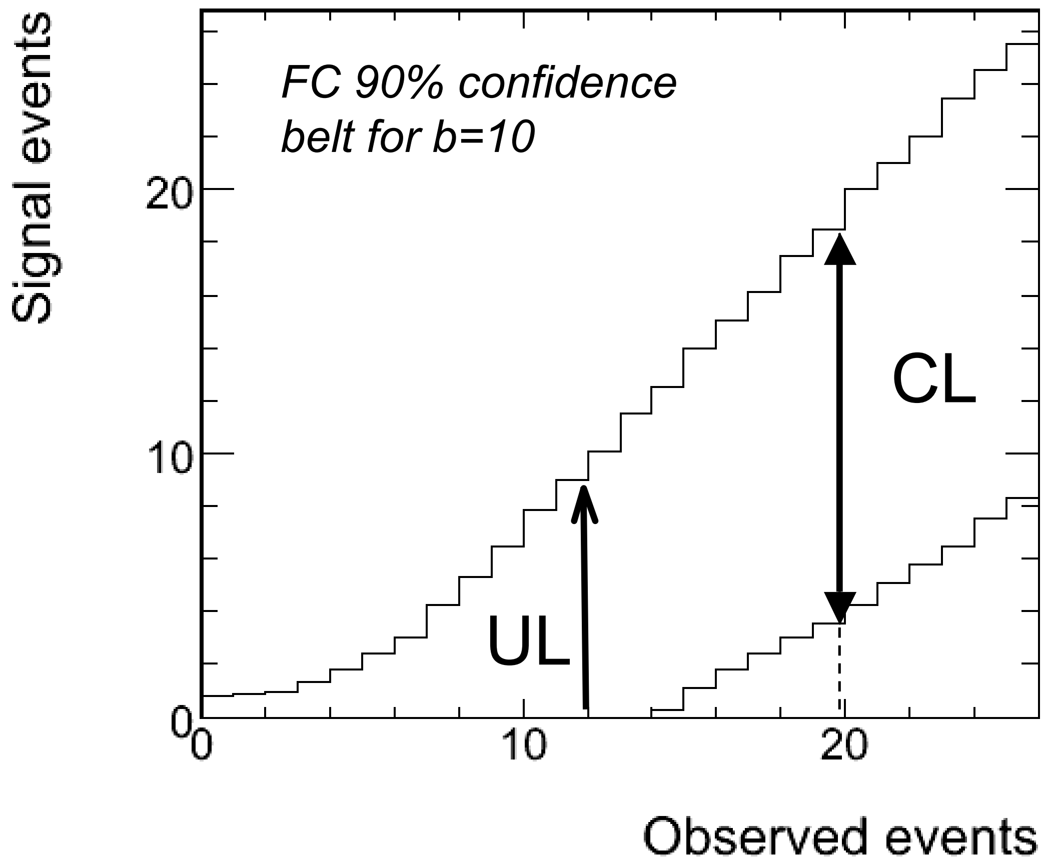
A frequentist approach based on construction Neyman's confidence belts:
- for each physical μ, select a set that includes 90% of observed events

FC 90% confidence belt for b=10



Feldman-Cousins

- find intersection of measured value with 90%CL line(s)



Feldman-Cousins

A trick is in deciding which events to include.

- Order by probability ratio:

$$R = \frac{p(n, \mu + b)}{p(n, \mu^{*} + b)}$$

Poisson probability for observing n events for given b and μ

Most likely *physical* value of $\mu = \mu *$ to observe n events

This approach described in:

- Feldman, Cousins, Phys Rev D57 (1988)
- Feldman, NEPPSR 2005

Feldman-Cousins

Computational example for μ=0.5 and b=3 from F-C paper

n	$P(n\|\mu)$	μ_{best}	$P(n\|\mu_{\text{best}})$	R	rank	U.L.	central
0	0.030	0.0	0.050	0.607	6		
1	0.106	0.0	0.149	0.708	5	√	√
2	0.185	0.0	0.224	0.826	3	√	√
3	0.216	0.0	0.224	0.963	2	√	√
4	0.189	1.0	0.195	0.966	1	√	√
5	0.132	2.0	0.175	0.753	4	√	√
6	0.077	3.0	0.161	0.480	7	√	√
7	0.039	4.0	0.149	0.259		√	√
8	0.017	5.0	0.140	0.121		√	
9	0.007	6.0	0.132	0.050		√	
10	0.002	7.0	0.125	0.018		√	
11	0.001	8.0	0.119	0.006		√	

FC ordering uses same events for upper limit and central limit

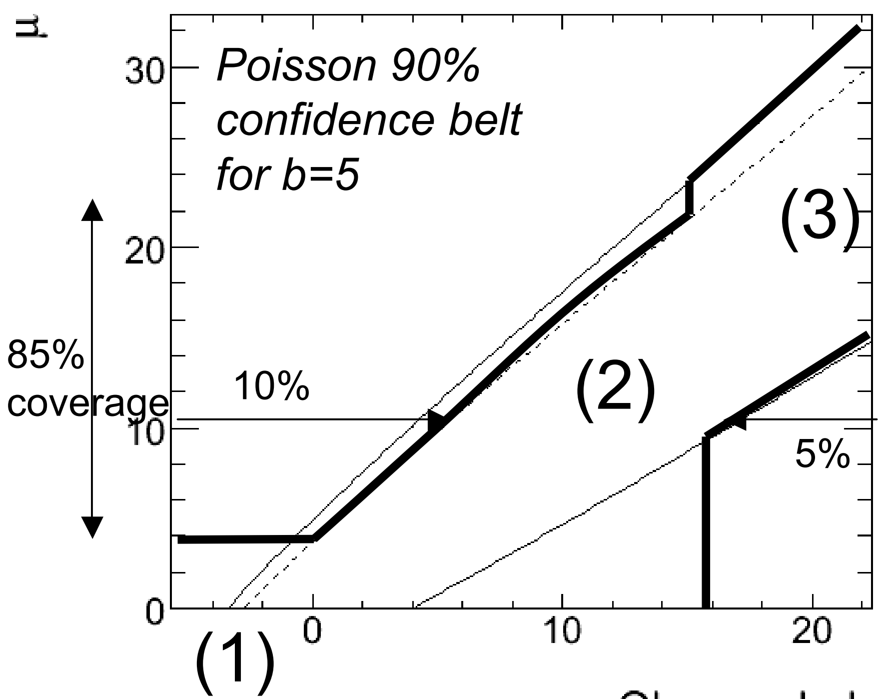
Poisson ordering uses different events for upper limit and central limit
Undercoverage, flip-flopping

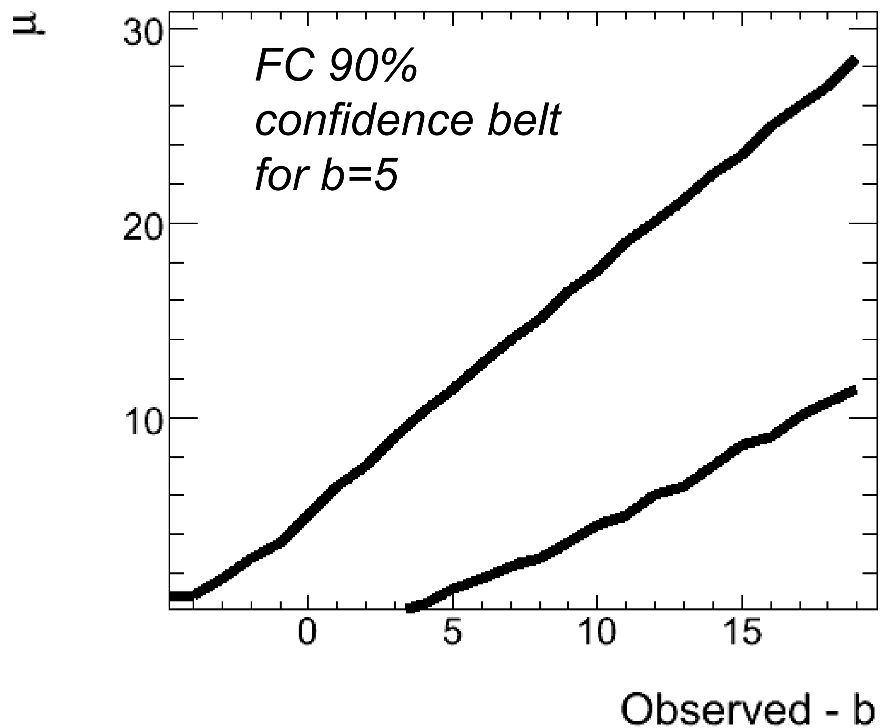
Feldman-Cousins

Example: under-coverage and flip-flopping

(1) If signal<0, pretend it's zero
(2) If signal<3σ, use upper limit
(3) If signal>3σ, use central limit

(1)=(2)=(3)





Likelihood Model

Assume background probability distribution function (PDF) is known ($P_{BG}(E_R)$)
Likelihood for an event i:

$$L(i) = P_{WIMP}(E_R;i)\cdot\mu + P_{BG}(E_R;i)\cdot b$$

Total likelihood for a sample

$$\int dE_R P(E_R) = 1$$

$$L = \frac{1}{N!}\cdot e^{-(\mu+b)}\cdot\prod_{i=1}^{N} L(i)$$

Note

$$\int L dE_R = \frac{1}{N!}\cdot e^{-(\mu+b)}\cdot(\mu+b)^N$$

Poisson distribution for N observed events when $\mu+b$ expected ('extended ML')

Maximum Likelihood Fit

- Vary model parameters p_i to maximize likelihood function
- For technical reasons, minimize -log(L):

$$\frac{\partial}{\partial p_i}\left(-\log L\left(p_i^0\right)\right) = 0$$

... in Gaussian approximation:

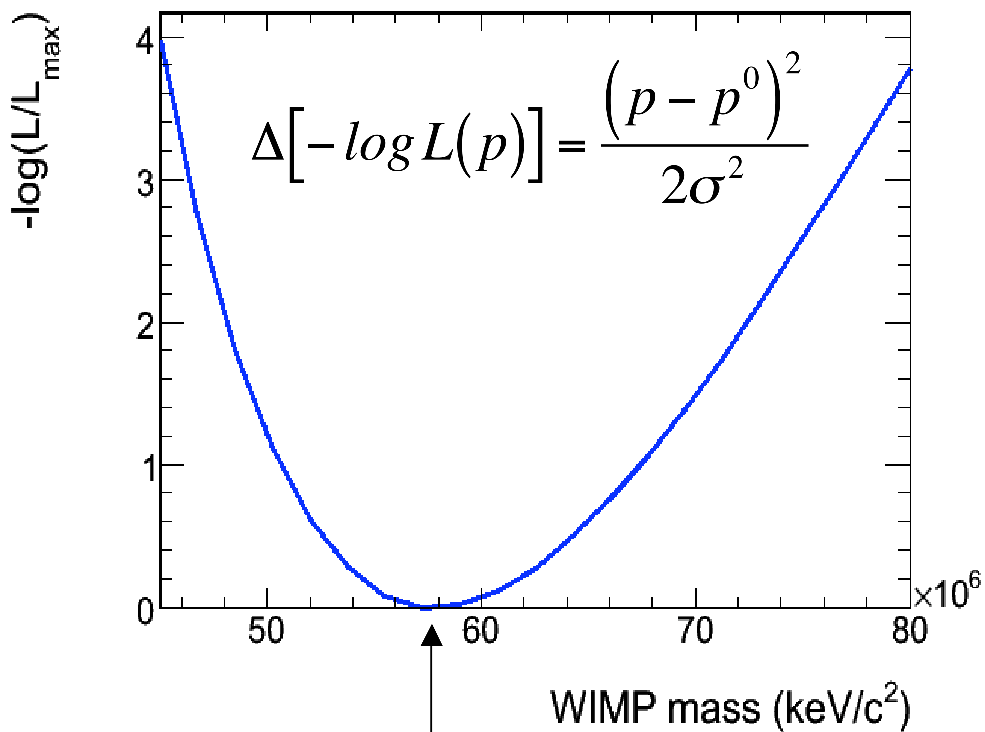
$$\log L(p_i) = \log L\left(p_i^0\right) - \frac{1}{2}\cdot\frac{\partial^2 \log L\left(p_i^0\right)}{\partial p_i^2}\cdot\left(p_i - p_i^0\right)^2 - \sum_{i\neq j}\left(p_i - p_i^0\right)\frac{\partial^2 \log L}{\partial p_i \partial p_j}\left(p_j - p_j^0\right)$$

$$\frac{1}{\sigma^2} = \frac{\partial^2 \log L}{\partial p_i \partial p_j}$$

$$\Delta\left[-\log L(p)\right] = \frac{\left(p - p^0\right)^2}{2\sigma^2}$$

Maximum Likelihood Fit

Suppose we have *many* events - Gaussian approximation ok
Minimum is estimator for true value of parameter

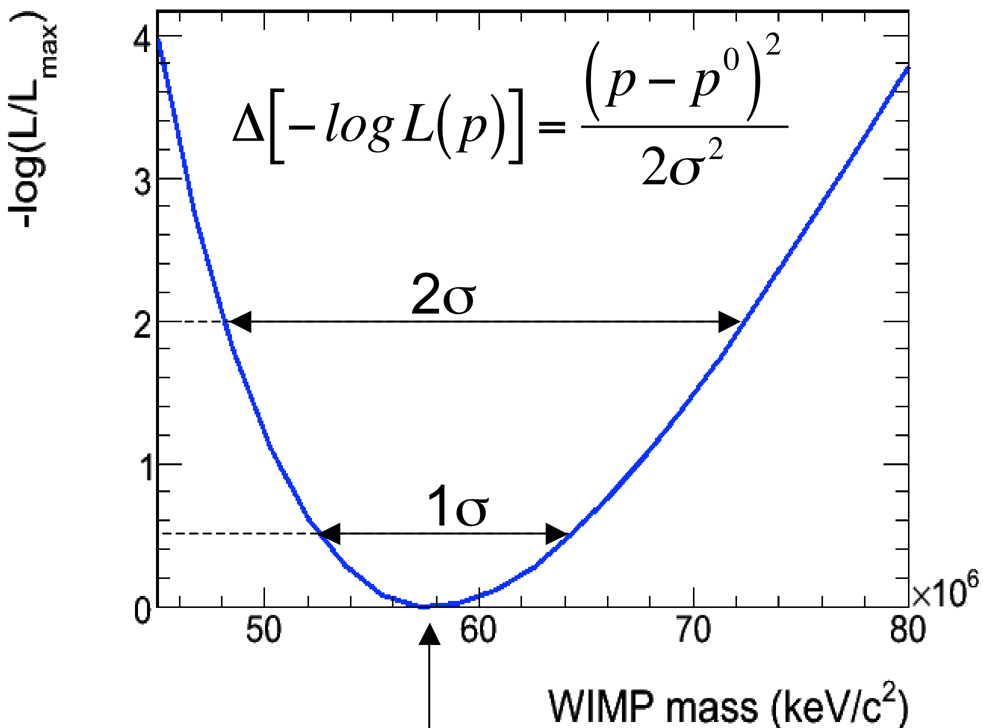


$$\Delta\left[-\log L(p)\right] = \frac{\left(p - p^0\right)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial p_i}\left(-\log L\left(p_i^0\right)\right) = 0$$

Maximum Likelihood Fit

Error estimate, assuming Gaussian distribution around p^0 - symmetric errors



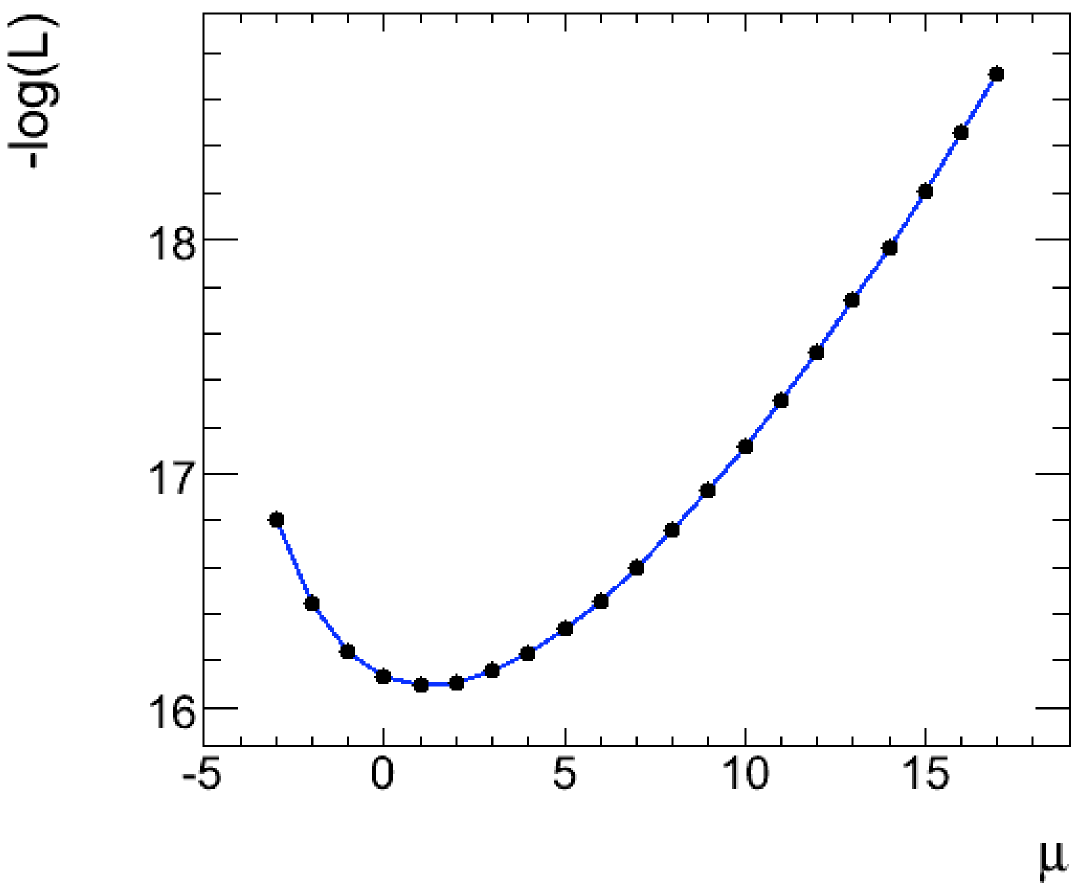
$$\Delta\left[-\log L(p)\right] = \frac{\left(p - p^0\right)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial p_i}\left(-\log L\left(p_i^0\right)\right) = 0$$

Change in log(L)	Change in sigma	% of values
1/2	1	68
2	2	95
$n^2/2$	n	Gamma function Prob(χ^2,dof)

Maximum Likelihood Fit

Small sample - expect non-Gaussian, asymmetric errors



- Likelihood function scan: fix μ, refit while floating other parameters (b)

- Use likelihood function to set upper limit

Upper Limit

Bayesian limit

$$CL = \frac{\int_{-\infty}^{\mu_{90}} d\mu \cdot L(\mu,b)\pi(\mu)}{\int_{-\infty}^{\infty} d\mu \cdot L(\mu,b)\pi(\mu)}$$

Find μ_{90} such that CL=90%
Use a flat prior: $\pi(\mu)=1,\ \mu>0$
$\pi(\mu)=0$, otherwise

Note:experts are picky about priors - everyone has its own best choice
=> In addition to 90%CL, experiments publish *full likelihood function* - later combined with likelihoods from other experiments

Bias in Fitted Parameters?

Most difficult step is to confirm that a fit makes sense.

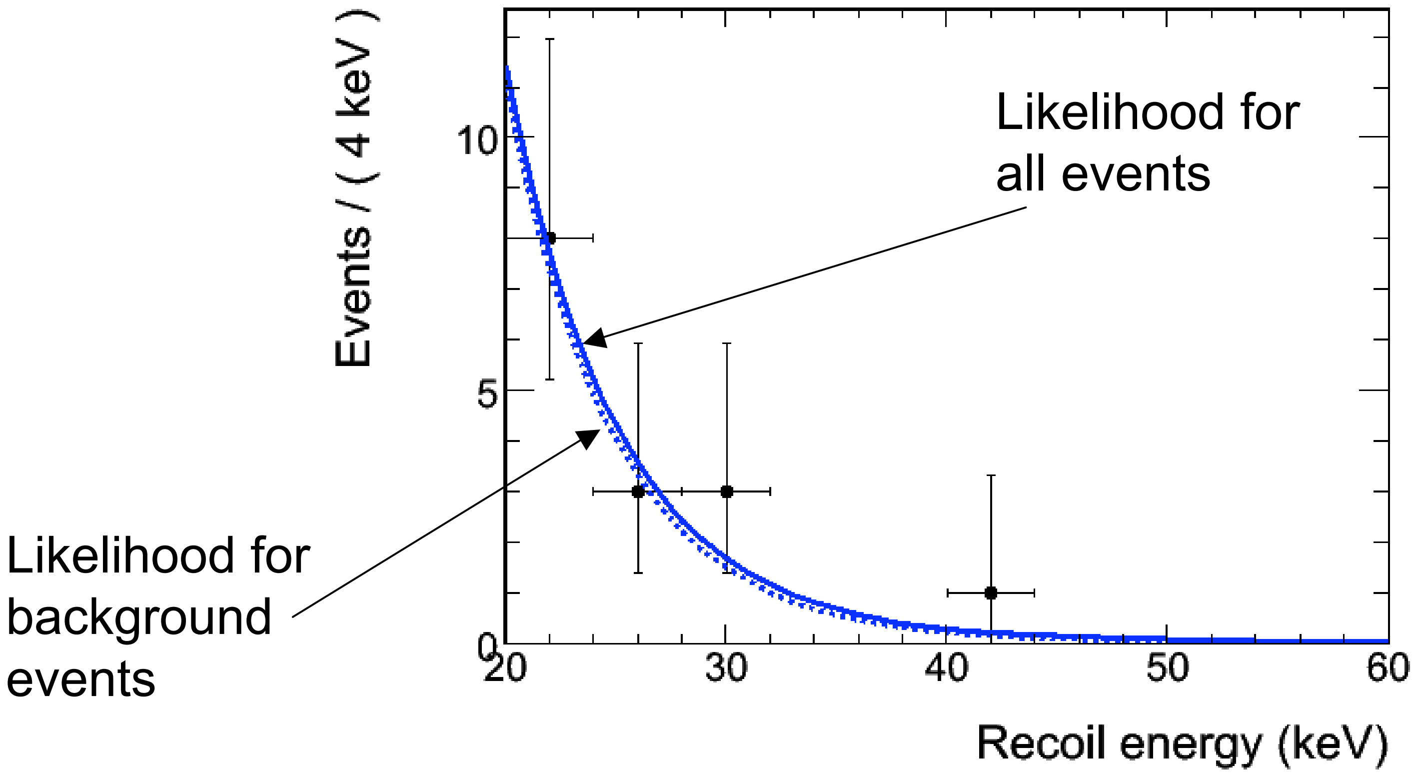
Bias in fitted parameters can originate from

• incorrect PDF’s (ignored correlation between observables, wrong shapes, etc.)
  - verify with simulated dataset, control samples

• minimization problems (e.g. parameters close to edge, convergence to local minima, bugs …)
  - check plots with likelihood projections, ‘pulls’
  - many can be checked with blinded fit parameters.

Visual Check

Overlay data with likelihood function
An obvious, but very useful test - likelihood shape should follow data



NEPPSR Problem Set

Take the following set of measured recoil energies in 20-100keV acceptance interval: 22.4, 27.7, 31.3, 23.6, 31.6, 23.4, 27.7, 24.6, 40.5, 29.7, 22.2, 20.0

Compute the following bounds on the number of signal events.

1) Compute upper limit on signal events with 90%CL using Poisson statistics.

   Discussion: Does the result change with different prior? (e.g. flat in $(\mu+b)^{0.5}$, $\log(\mu+b)$)

2) For 60keV WIMP and 19GeV target mass, and ignoring detector effects, the distribution of recoils energies for signal events is given as

$$P_{WIMP}\left(E_R\right) \propto exp\left[-\frac{E_R}{11.8keV}\right]$$

   Use Maximum Gap method to find upper limit on signal events with 90% C.L.

   Discussion: How would you include detector effects (efficiency, resolution)?

NEPPSR Problem Set

3) Assume the expected background rate is b=10 events. Calculate Feldman-Cousins ordering ratios and construct Neyman 90% confidence bends for signal μ=0,1…12 events. Find an upper limit on the number of signal events for the given sample.
Discussion: How does the upper limit change if b=13? Comment.

4) Take background distribution

$$P_{BG}\left(E_R, cos\gamma\right) \propto exp\left[-\frac{E_R}{5keV}\right]$$

Construct an extended likelihood function and minimize -log(L) to find signal and background events.
Make a likelihood scan and find upper limit on signal events by integrating the likelihood function using a flat prior for μ>0.
Discussion: What if fit gives negative μ?