Dark Matters - The trouble with searching for rare events

Ken Clark
March, 2016

Looking for Rare Events

- First of all, what is considered rare?
- Certainly anything that happens once in the span of a lifetime. Maybe once in a year?
- “You’ll know one when you see one … if you see one”
- More formal:

> binary dependent variables with dozens to thousands of times fewer ones (events, such as wars, vetoes, cases of political activism, or epidemiological infections) than zeros (“nonevents") - Logistic Regression in Rare Events Data

Dealing with Rare Events

- So what can we do? Seem to be three solutions:
  - Wait
  - Predict from other information
  - Interpret the results we have

Rare Events

- Let's assume that anything that happens < 1 time per month is rare
- Could be many things… meteor strikes, government upheavals, floods, birthdays…
- Can we do anything predictive with this class of events?

Answer: Maybe?

- Discrete physical phenomena tend to follow a power law distribution
- Can be used with earthquake magnitude
- Gutenberg-Richter law

$$N = 10^{a-bM}$$

N - number of earthquakes
M - magnitude of the earthquake
a,b - constants

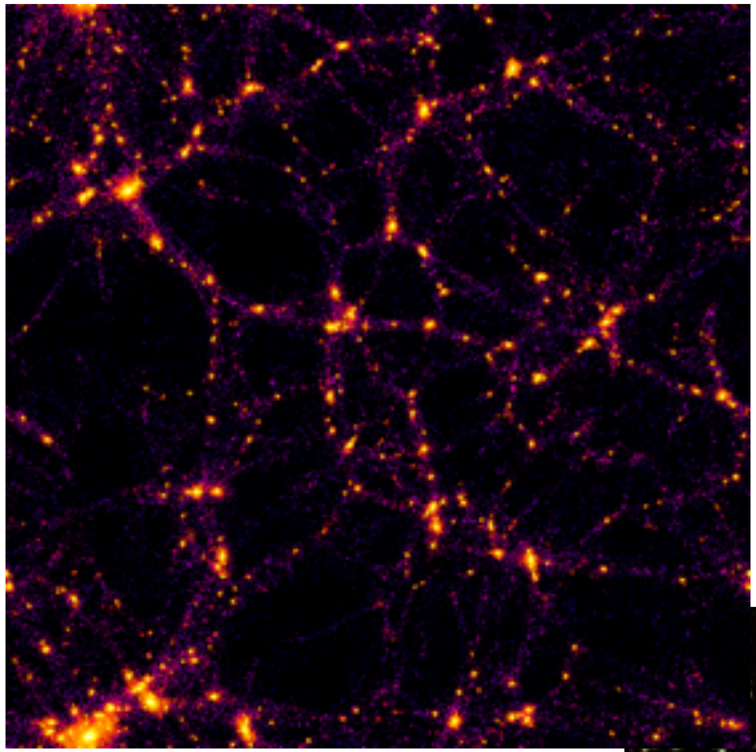
UNIVERSITY OF TORONTO

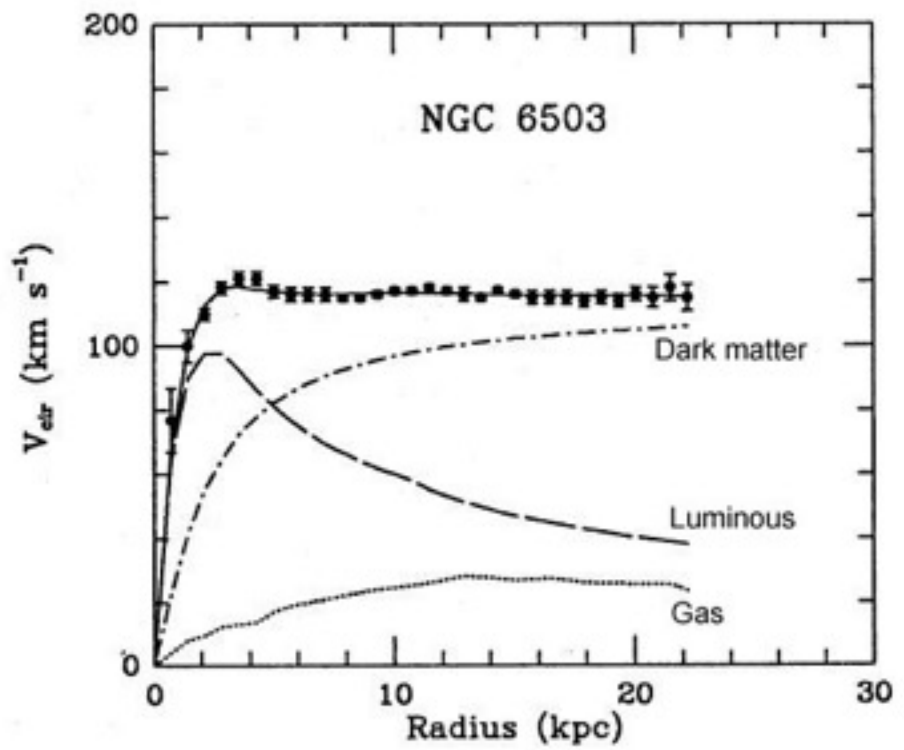
Earthquake Data

- Data on recent earthquakes is available here: http://earthquake.usgs.gov/earthquakes/feed/v1.0/csv.php
- Can you predict the number of occurrences of magnitude 5? 6?
- How accurate are the predictions?

Dark Matter Introduction

Structure formation with simulation



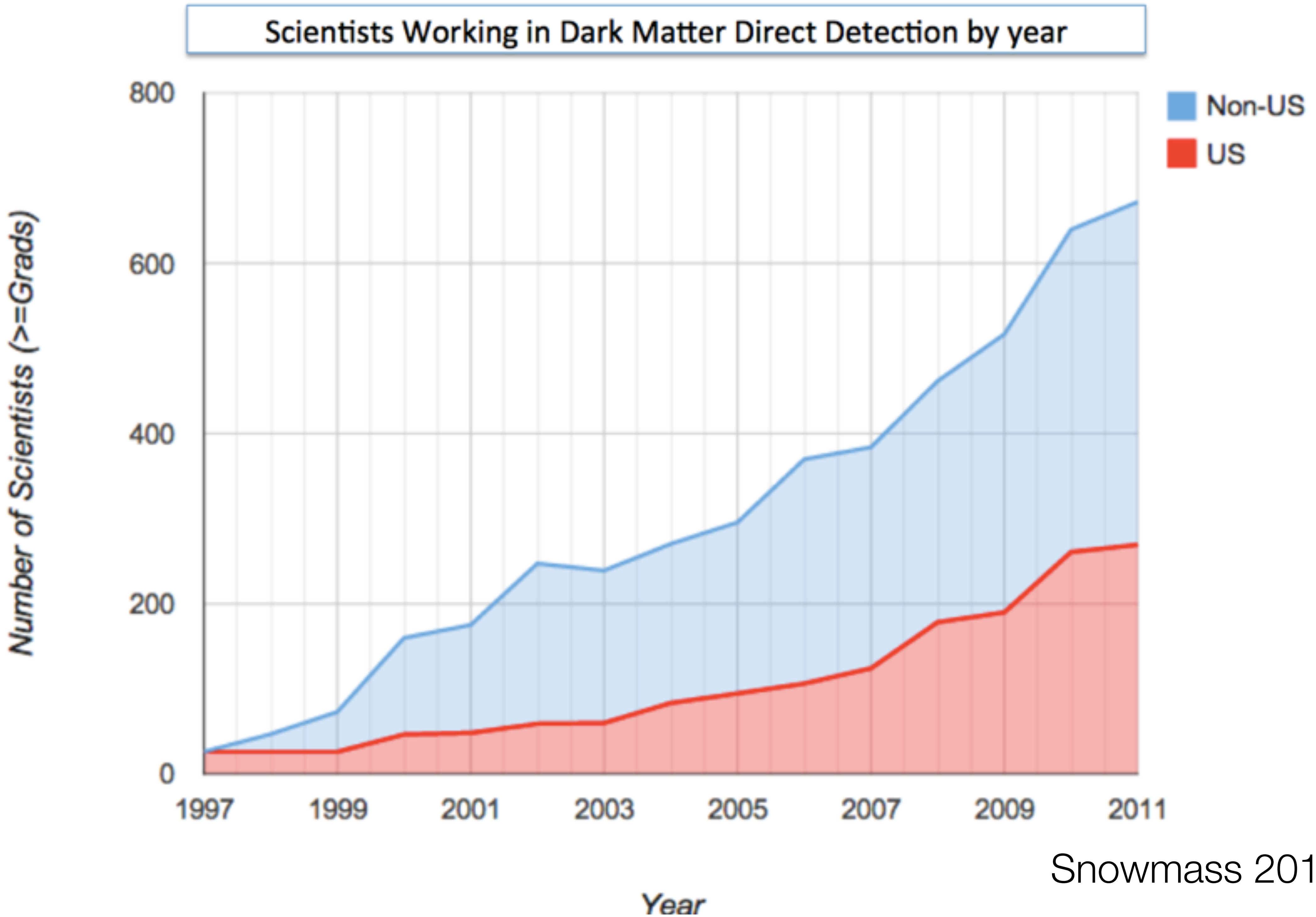


Galaxy rotation curves

Separation of luminous and non-luminous mass

UNIVERSITY OF TORONTO

A Popular Field…



Snowmass 2013

UNIVERSITY OF TORONTO

Dark Matter

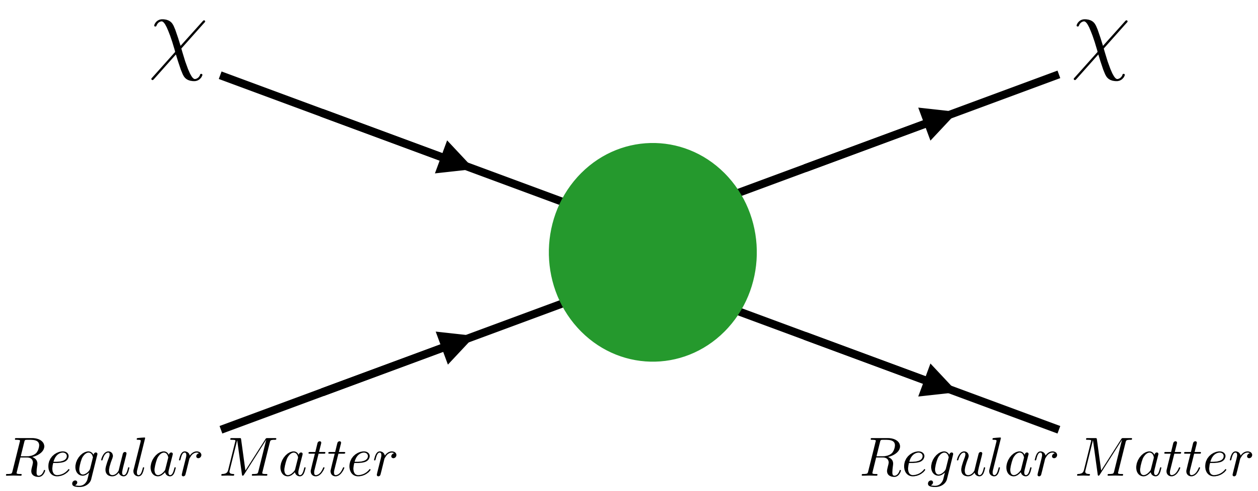
What do we know?

1. Long lived (survived until current day)
2. Non-baryonic (Hydrogen:Deuterium Ratio)
3. No EM interactions (haven't seen it)
4. 80% of all matter (rotation curves, CMB)
5. Non-relativistic (galactic structure formation, rotation curves)

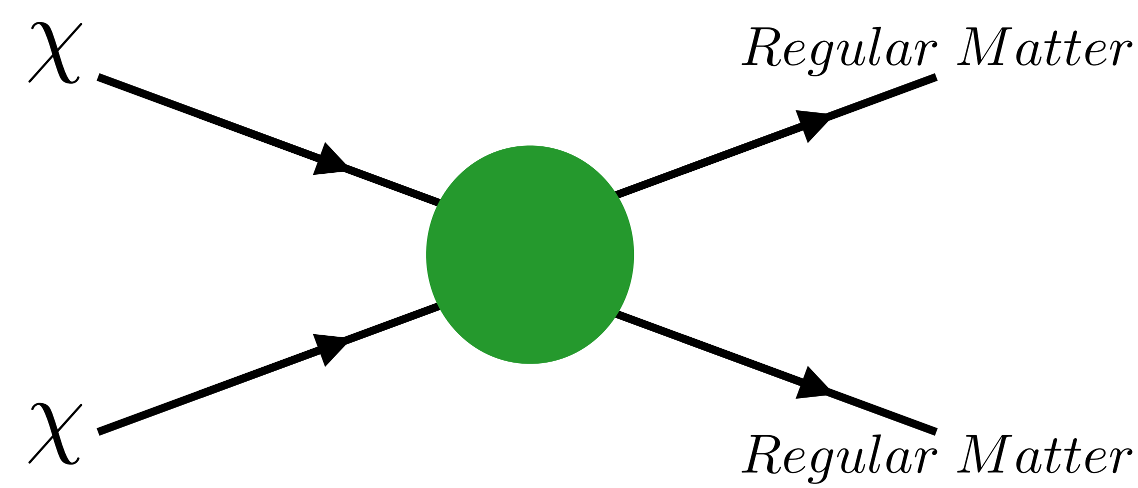
Dark Matter Detection

Two primary types: direct and indirect

Interaction of particle inside the detector



Products of self-interaction outside the detector



Dark Matter Math

$$\frac{dR}{dE_\nu} = \sigma_p \sum_A \overbrace{f(A)}^{\text{mass fraction of target}} \times \underbrace{S(A, E_R)}_{\text{phase space distribution of trapped DM}} \times \overbrace{I(A)}^{\text{coupling enhancement}} \times \underbrace{F^2(A, E_R)}_{\text{nuclear form factor}} \times \overbrace{g(A)}^{\text{nuclear recoil quenching factor}} \times \underbrace{\epsilon(E_\nu)}_{\text{detector response function}}$$

cross-section (→ σ_p)

A - Atomic Number

E_R - nuclear recoil energy

UNIVERSITY OF TORONTO

Direct Detection Methods

- All experiments search for **small** energy deposits
- Make some assumptions to find MAXIMUM energy deposited
  - M_X = 100 GeV = 1.8×10^{-25} kg
  - v = 220 km/s = 2.2×10^{5} m/s

$$T = \frac{1}{2}mv^2 = \frac{1}{2} * 1.8 \times 10^{-25} * (2.2 \times 10^5)^2 = 4.4 \times 10^{-15} J$$

  - Equivalent to a mosquito flying at 0.00015 kph

UNIVERSITY OF TORONTO

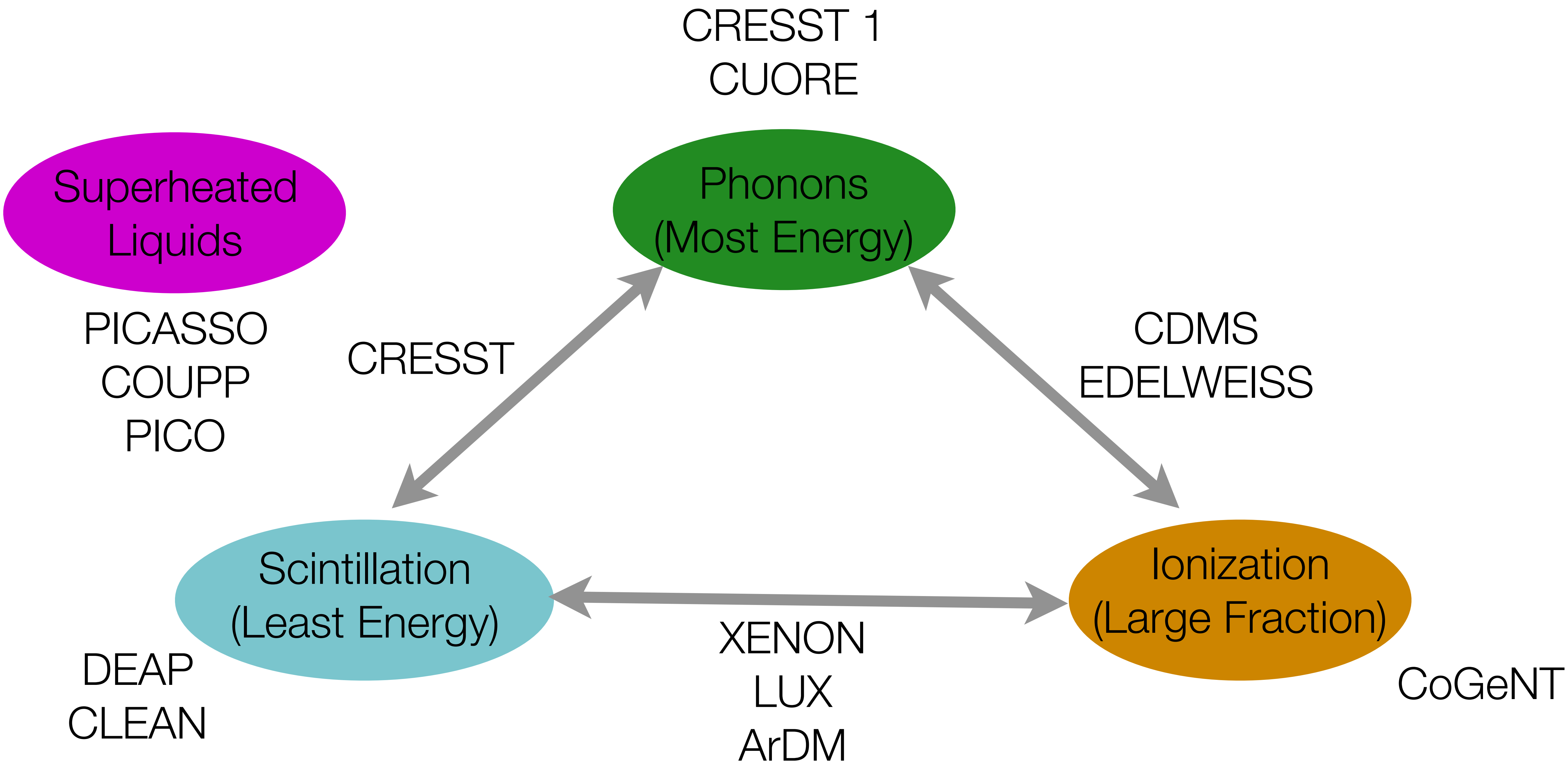
Background Removal

- Because of this small energy deposit, DM experiments are all about dealing with backgrounds
- Done in two ways

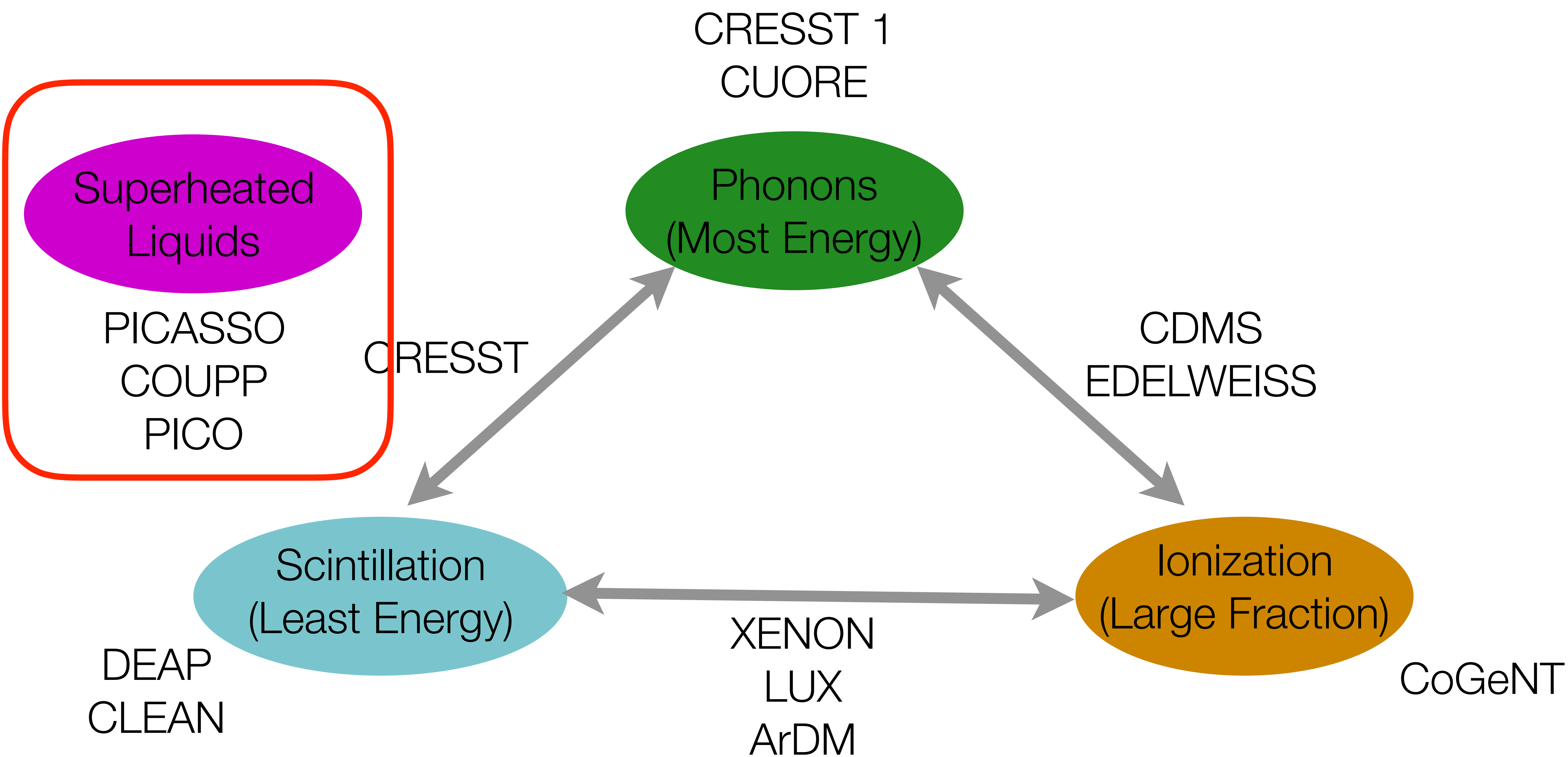
1. Don't have interactions you don't want
2. Tell the "bad" interactions from the "good"

UNIVERSITY OF TORONTO

Background Removal



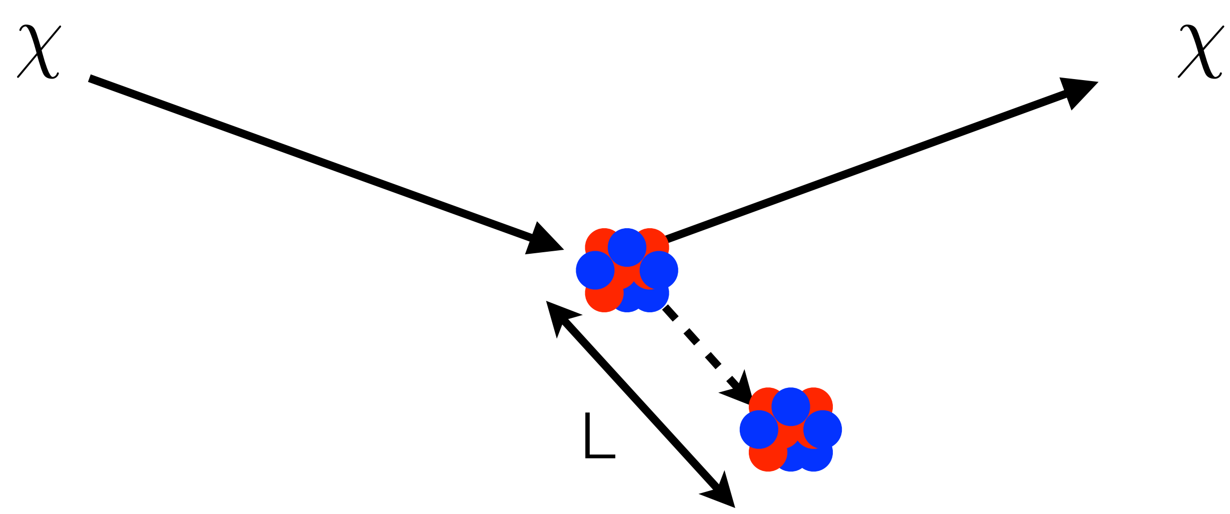
Background Removal



UNIVERSITY OF TORONTO

PICASSO + COUPP = PICO

- PICO is a novel detector using superheated liquid to amplify the energy deposit
- Small deposit of energy triggers the formation of large bubbles, detectable using acoustic or visual methods



Seitz Model

- The currently used model says that the energy for the formation of the bubble must come from the interaction, not the surrounding fluid
- This requires a threshold energy deposit in a critical radius

$$E_{threshold} = 4\pi r_c^2 \left(\sigma - T\frac{d\sigma}{dT}\right) + \frac{4\pi}{3} r_c^3 \rho_b \left(h_b - h_l\right) - \frac{4\pi}{3} r_c^3 \left(P_b - P_l\right)$$

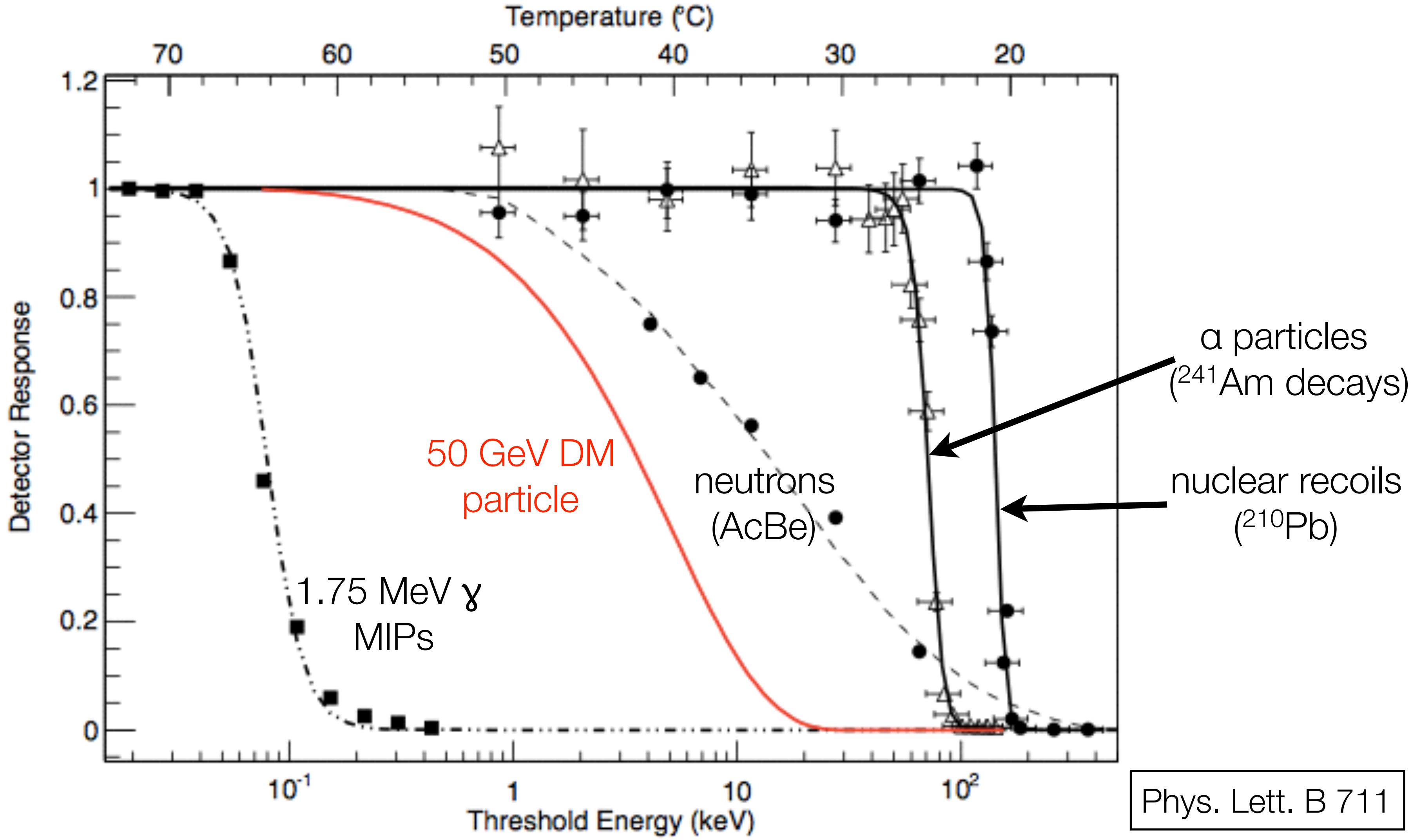
Seitz Model

- The currently used model says that the energy for the formation of the bubble must come from the interaction, not the surrounding fluid
- This requires a threshold energy deposit in a critical radius

$$E_{threshold} = \underbrace{4\pi r_c^2\left(\sigma - T\frac{d\sigma}{dT}\right)}_{\text{Overcoming surface tension}} + \underbrace{\frac{4\pi}{3}r_c^3\rho_b\left(h_b - h_l\right)}_{\text{Vaporization of fluid}} - \underbrace{\frac{4\pi}{3}r_c^3\left(P_b - P_l\right)}_{\text{Double counting}}$$

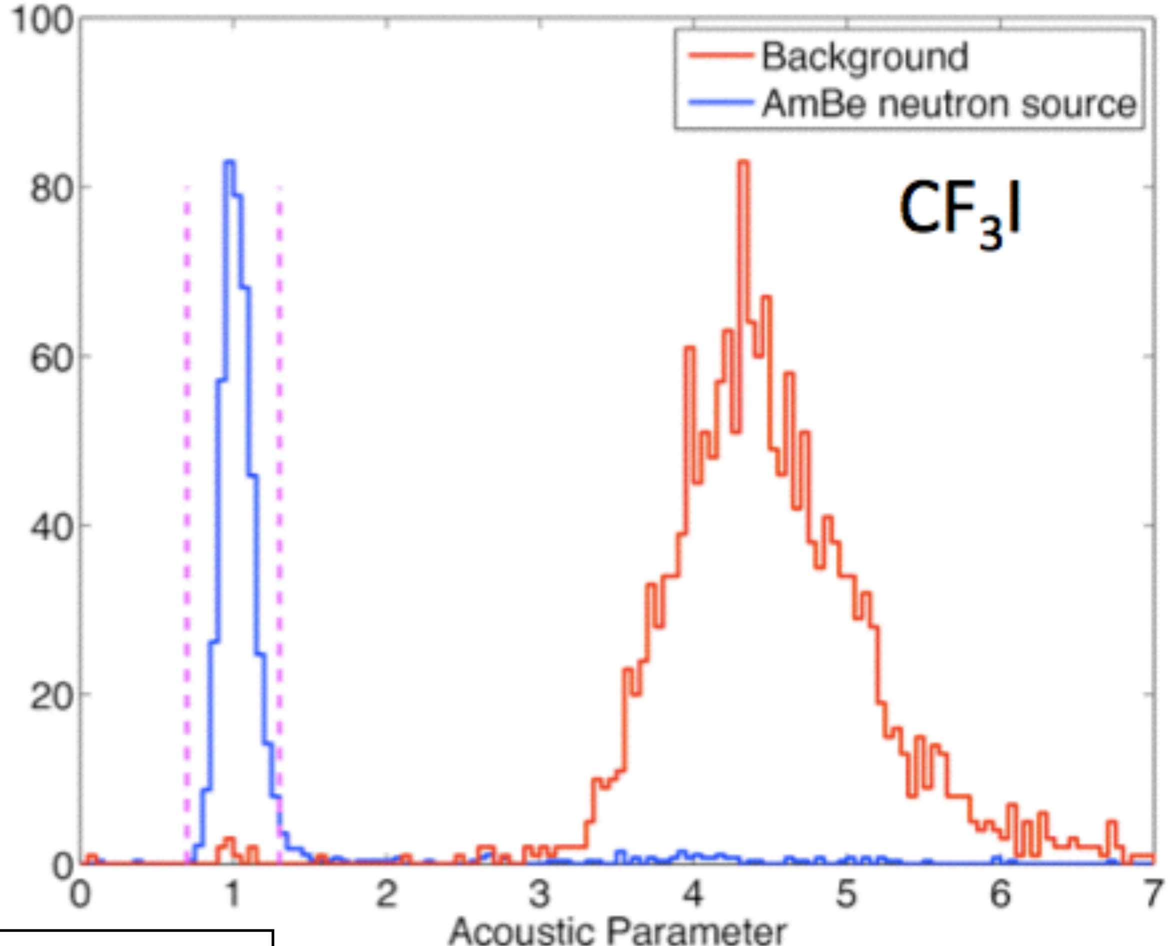
Backgrounds Method 1

- Gammas and betas are effectively not detected by the detector as they do not meet the $E_{threshold}$ in r_c requirement.



Phys. Lett. B 711

Backgrounds Method 2



PRD 86:052001

- Distinguish the bubbles formed by backgrounds from those formed by recoil events
- In this case, alphas vs neutrons

Temperature Correction, Transducer gain, Position correction, Frequency, Power Spectral Density

$$AP = A(T) \sum_j G_j \sum_n C_n(\vec{x}) \sum_{f^n_{min}}^{f^n_{max}} f \times psd^j_f$$

The Ultimate Goal (?)

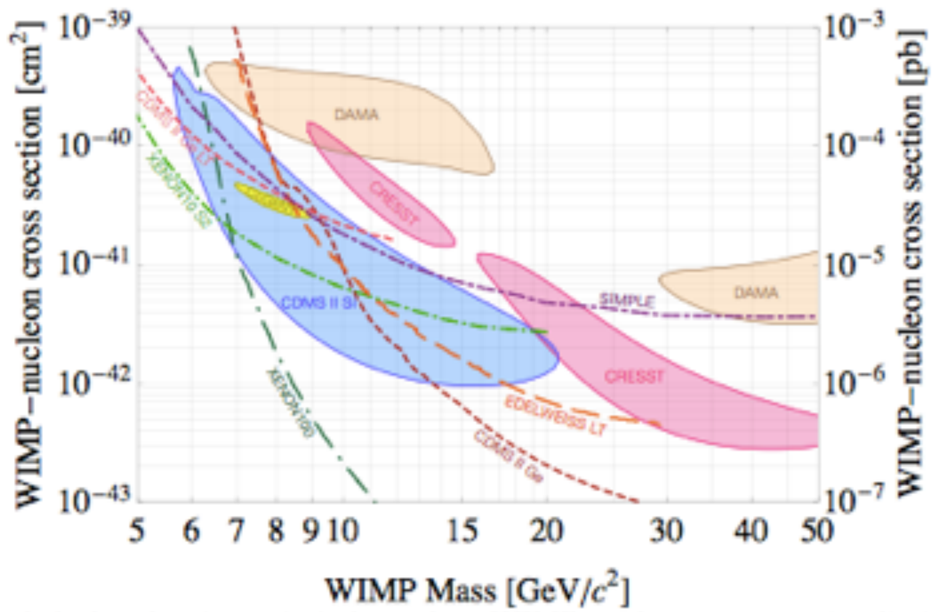


Figure 8. *Expanded plot showing spin-independent WIMP-nucleon cross section limits, including closed contours showing hints for low-mass WIMP signals.*

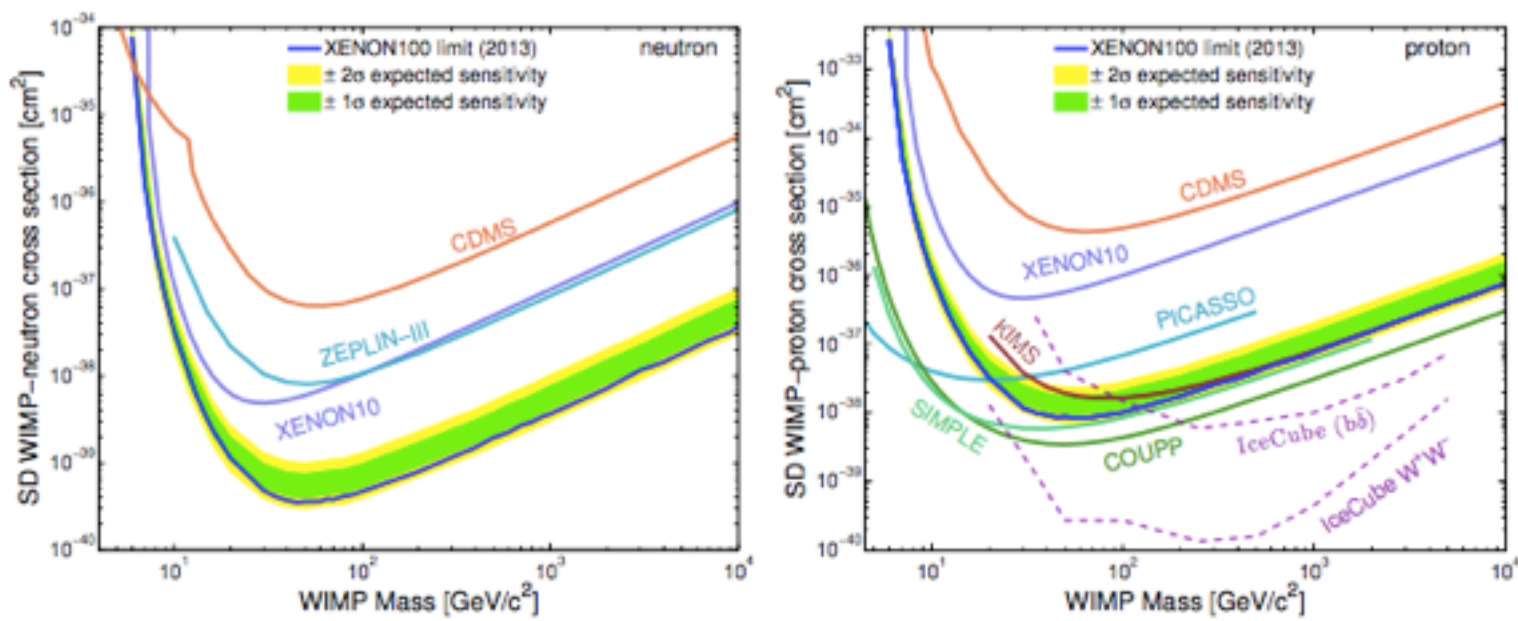


Figure 9. *Spin-dependent WIMP-neutron (left) and WIMP-proton (right) cross section limits versus WIMP mass for direct detection experiments[27, 28, 33, 38, 39, 40, 41], compared with the model-dependent Ice Cube results (model-dependent) as of summer 2013 [42].*

Snowmass 2013

- Goal should be discovery…

UNIVERSITY OF TORONTO

Calculation of Limit

- Calculation based on 4 parameters
  - Background rate B
  - Background misidentification β
  - Signal acceptance α
  - exposure MT (mass x time)

Calculation of Limit

- With NO discrimination (β=1) all interactions are potentially DM
- If there are no events, the 90% CL

$$P_{n(obs)}(n_s, 90) = \frac{(n_s, 90)^{n_{obs}}}{n_{obs}!} e^{-n_s, 90}$$

$$= \frac{(n_s, 90)^0}{0!} e^{-n_s, 90} = 0.1$$

$$\rightarrow n_{s,90} = -ln(0.1) = 2.3$$

Calculation of Limit

- With NO discrimination (β=1) all interactions are potentially DM
- If there are no events, the 90% CL

$$S_{90} = \frac{2.3}{\alpha M T}$$

- Obviously scales with the exposure time

Calculation of Limit

- With far more events than expected signal, assume all events are background

$$S_{90} = \frac{N_{BG} + 1.28\sqrt{N_{BG}}}{\alpha MT}$$

- This can be expressed as

$$S_{90} = \frac{\beta}{\alpha} + \frac{1.28}{\alpha}\sqrt{\frac{\beta B}{MT}}$$

Calculation of Limit

Zero Events

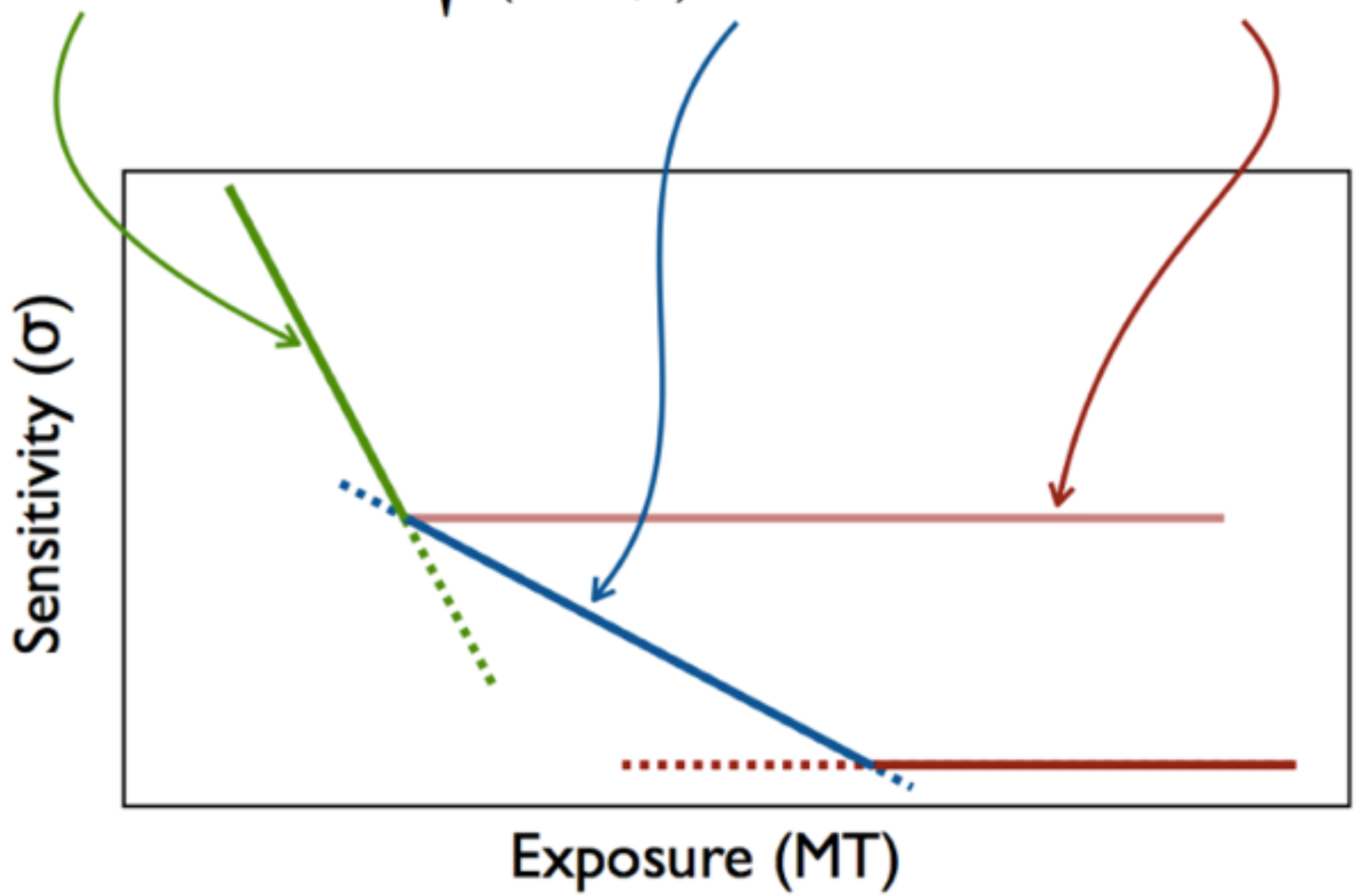
$$S_{90} \propto \frac{2.3}{\alpha MT}$$

Discriminating Detector

$$\sqrt{\frac{\beta(1-\beta)}{(\alpha-\beta)^2}}\sqrt{\frac{B}{MT}}$$

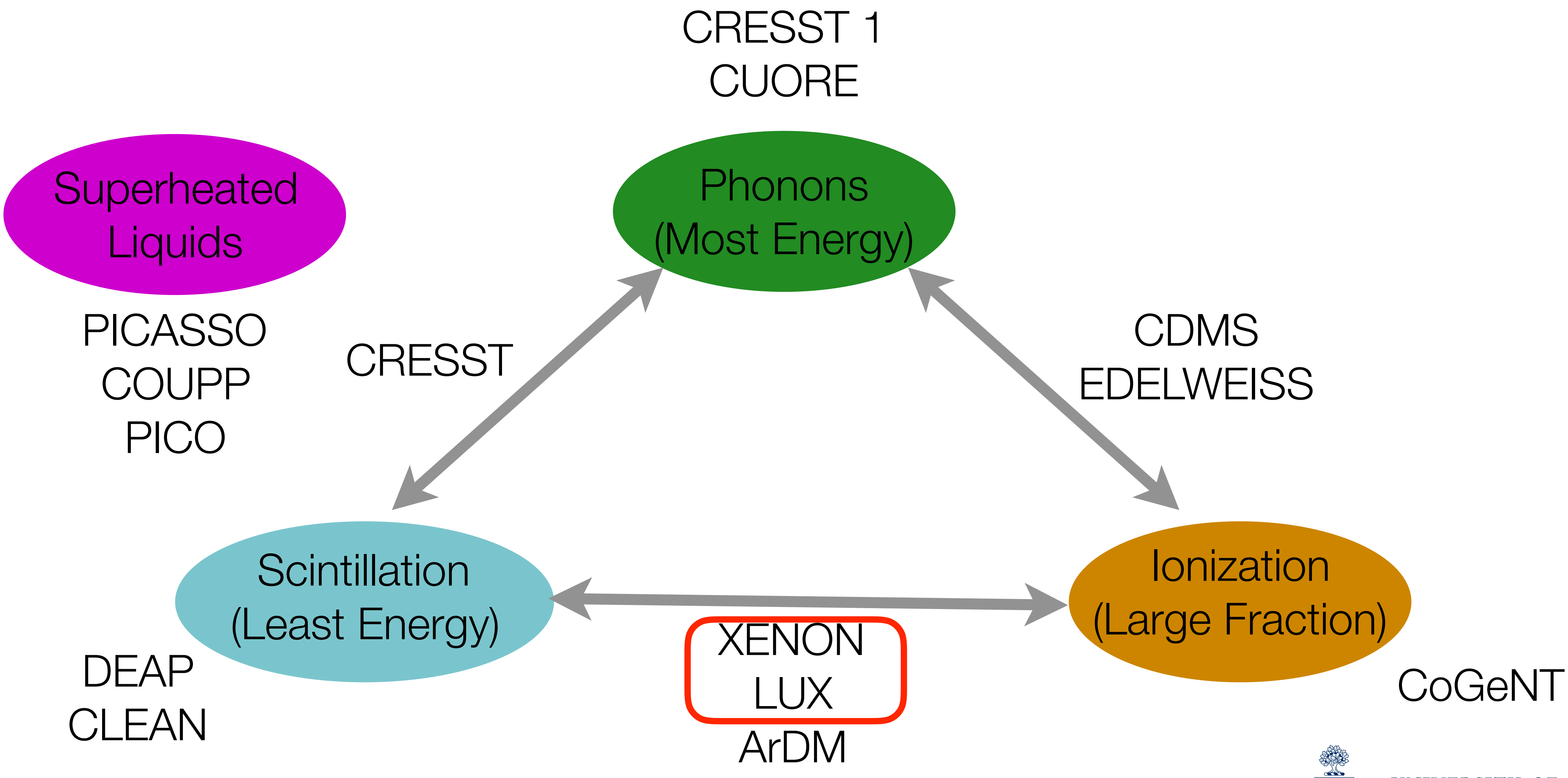
Non-Discriminating

$$\frac{B}{\alpha} + \frac{1.28}{\alpha}\sqrt{\frac{\beta B}{MT}}$$



A Better Way?

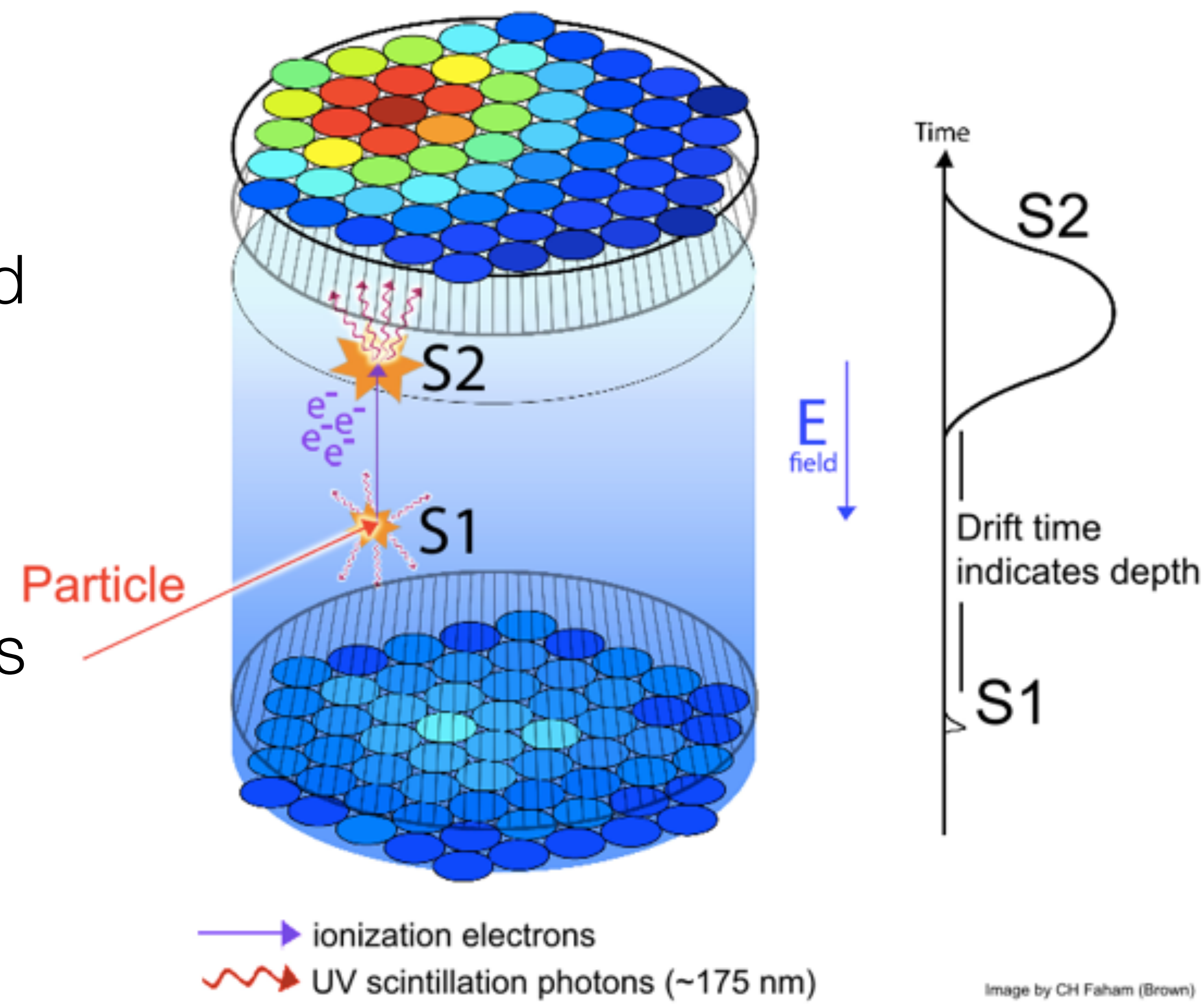
- Let's study a bit more complex example



UNIVERSITY OF TORONTO

Scintillation and Ionization

- Example used here is LUX
- Xenon used in both liquid and gas state within an electric field to amplify small deposit
- Two different scintillations detected, the ratio of which discriminates signal from background



Scintillation and Ionization

- Example used here is LUX
- Xenon used in both liquid and gas state within an electric field to amplify small deposit
- Two different scintillations detected, the ratio of which discriminates signal from background

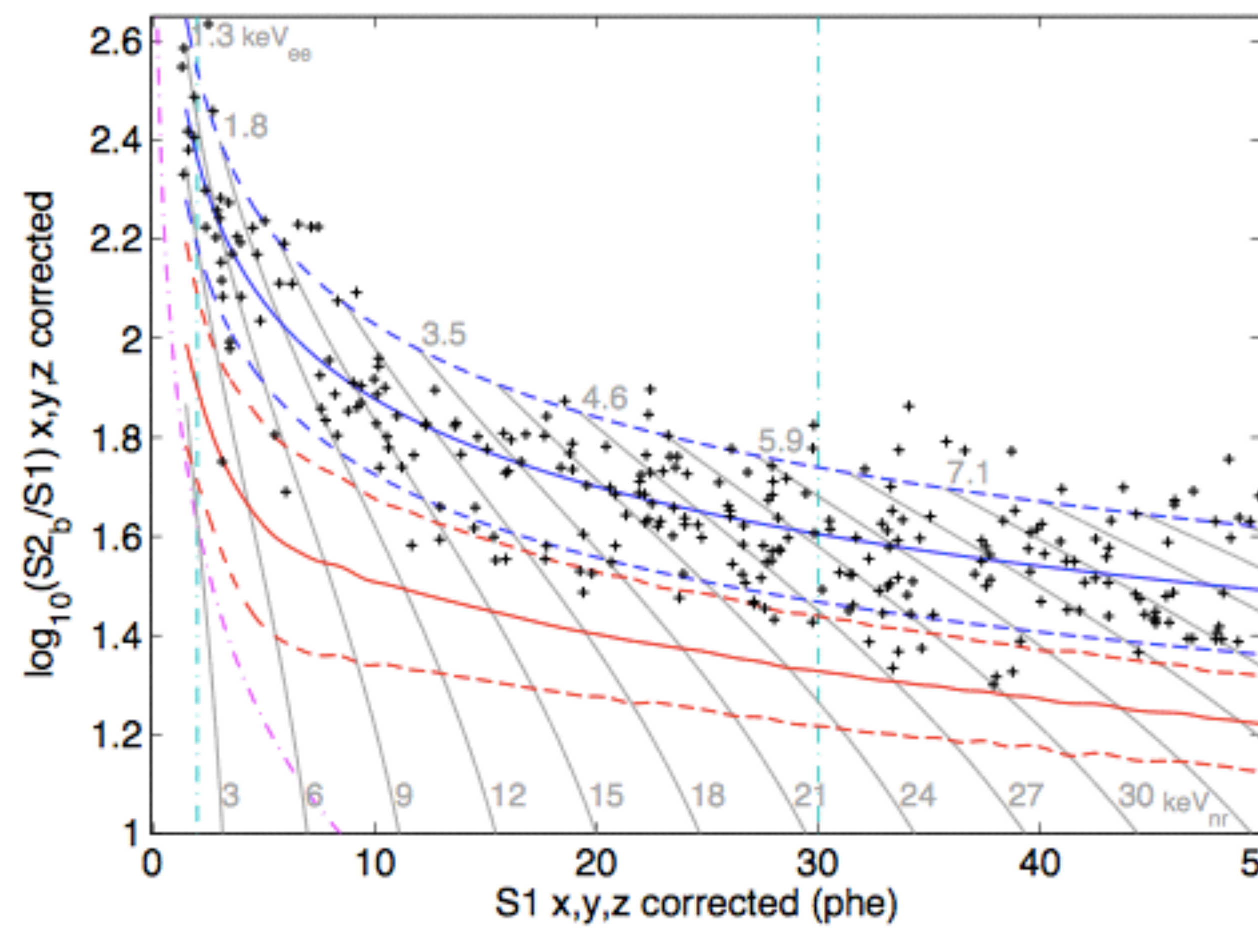
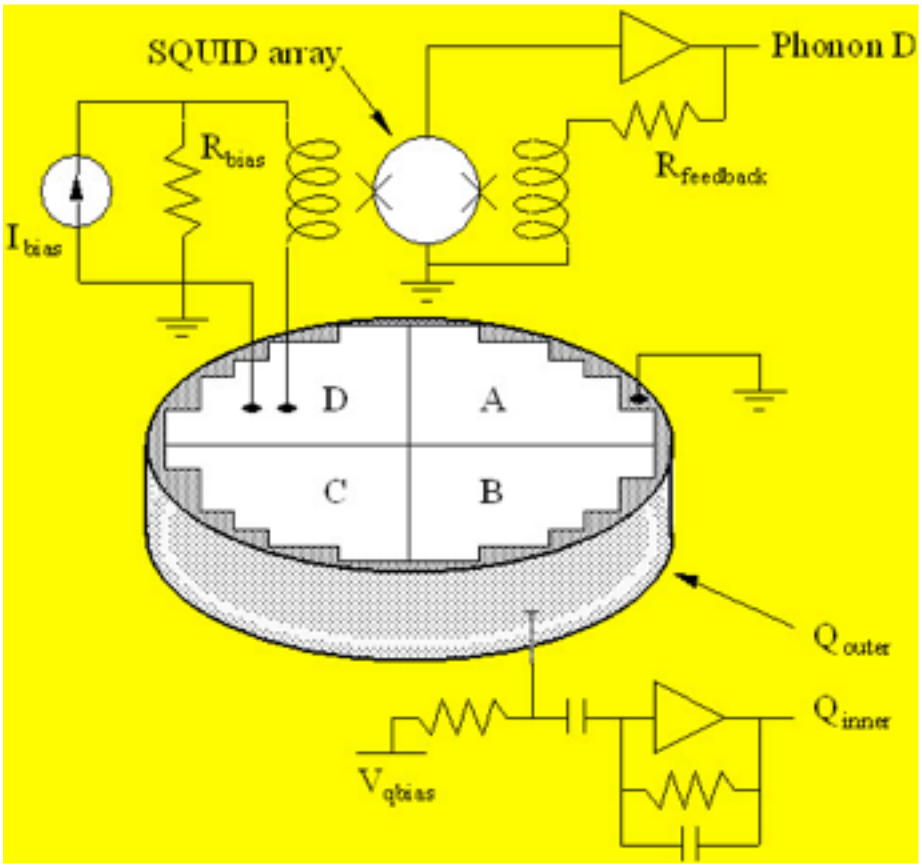
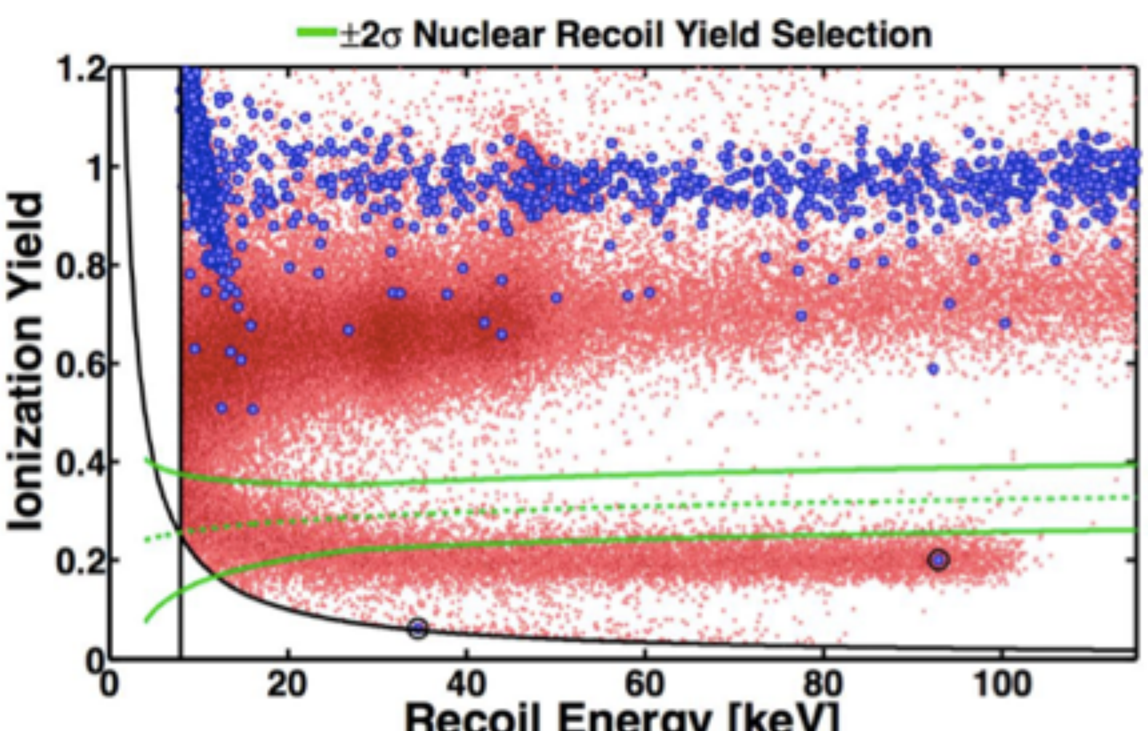


FIG. 4. The LUX WIMP signal region. Events in the 118 kg fiducial volume during the 85.3 live-day exposure are shown. Lines as shown in Fig. 3, with vertical dashed cyan lines showing the 2-30 phe range used for the signal estimation analysis.

arXiv:1310.8214

Phonons and Ionization





- CDMS similarly uses two channels to distinguish background from signal
- Collecting both ionization and phonons allows for discrimination

Risks

- The main (and obvious) risk to removing backgrounds is to the livetime of the experiment
- Aggressive cleaning puts you back into the region discussed previously

Setting a Limit

- Need to know the number of counts and the distribution of the probability function
- Also need to know the expected number of counts seen

Expected Counts

- The simplest expectation is:

$$\frac{dR}{dE_R} = \frac{R_0}{E_0 r} e^{-E_R/E_0 r}$$

- E_R is the recoil energy
- R_0 is the total event rate
- E_0 is the most probable dark matter energy

$$r = \frac{4M_D M_T}{(M_D + M_T)^2}$$

Trick is in the R_0

- Define the R0 so that it can be calculated
- It's the event rate per unit mass for the earth velocity is 0 and the escape velocity is infinite

$$R_0 = \frac{2}{\pi^{1/2}} \frac{N_0}{A} \frac{\rho_D}{M_D} \sigma_0 v_0$$

UNIVERSITY OF TORONTO

Defining The Bounds

- Really need to define the bounds on the observed number of events
- This has to take into account the backgrounds and the uncertainty associated with those backgrounds
- Three methods here:
    1. Feldman Cousins
    2. Yellin
    3. Binned Likelihood

Feldman Cousins

- Want to know the confidence region for the number of signal events given the number of observed and background

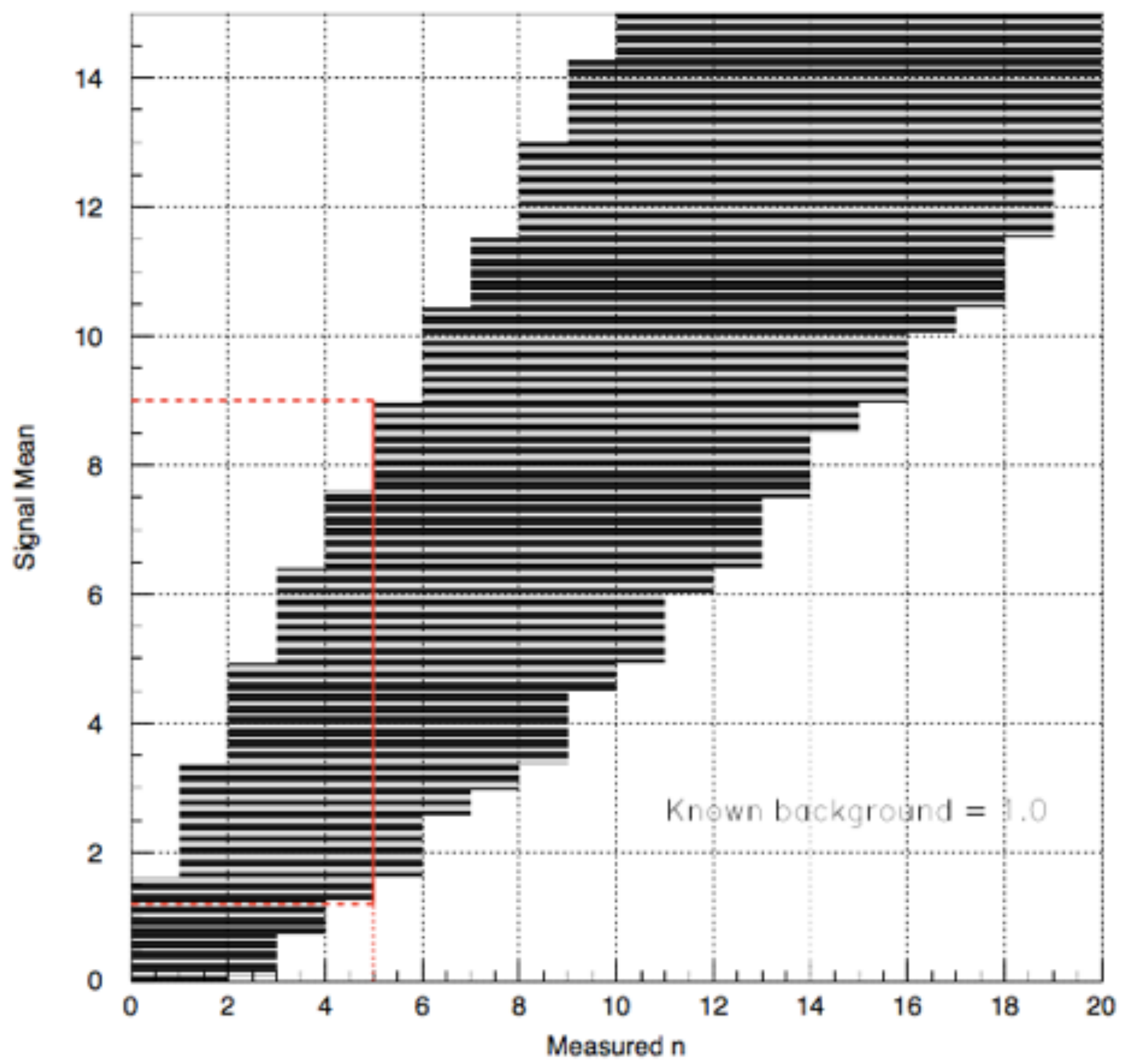
$$P(n|\mu) = \frac{(\mu + b)^n e^{-(\mu+b)}}{n!}$$

- Then maximize the probability, changing μ_{best} and the ratio is the parameter used to define 90%

$$R = \frac{P(n|\mu)}{P(n|\mu_{best})}$$

Feldman Cousins

- Plot this and read the results for your experiment!

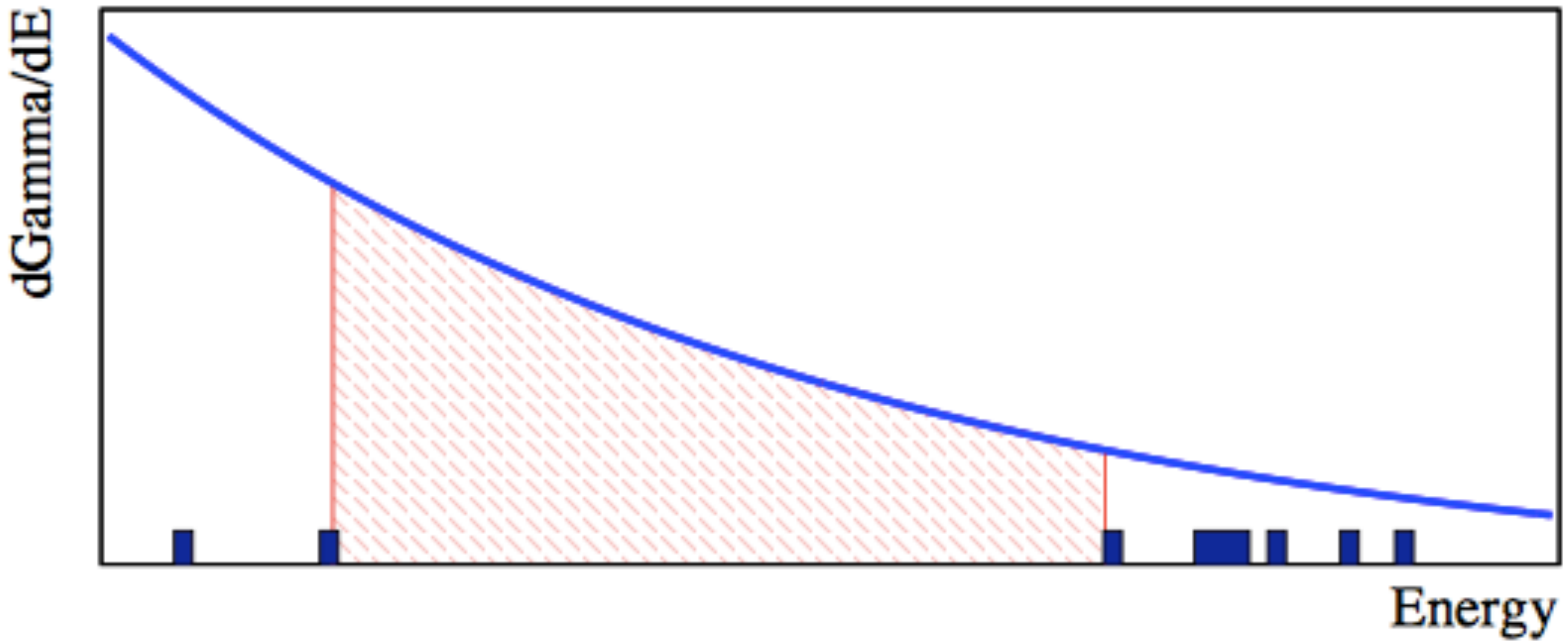


UNIVERSITY OF TORONTO

But… I'm not sure about my Backgrounds

- All DM experiments are in a new region of detector physics
- The backgrounds are not completely understood
- The best measurement of the backgrounds is the dark matter data itself…

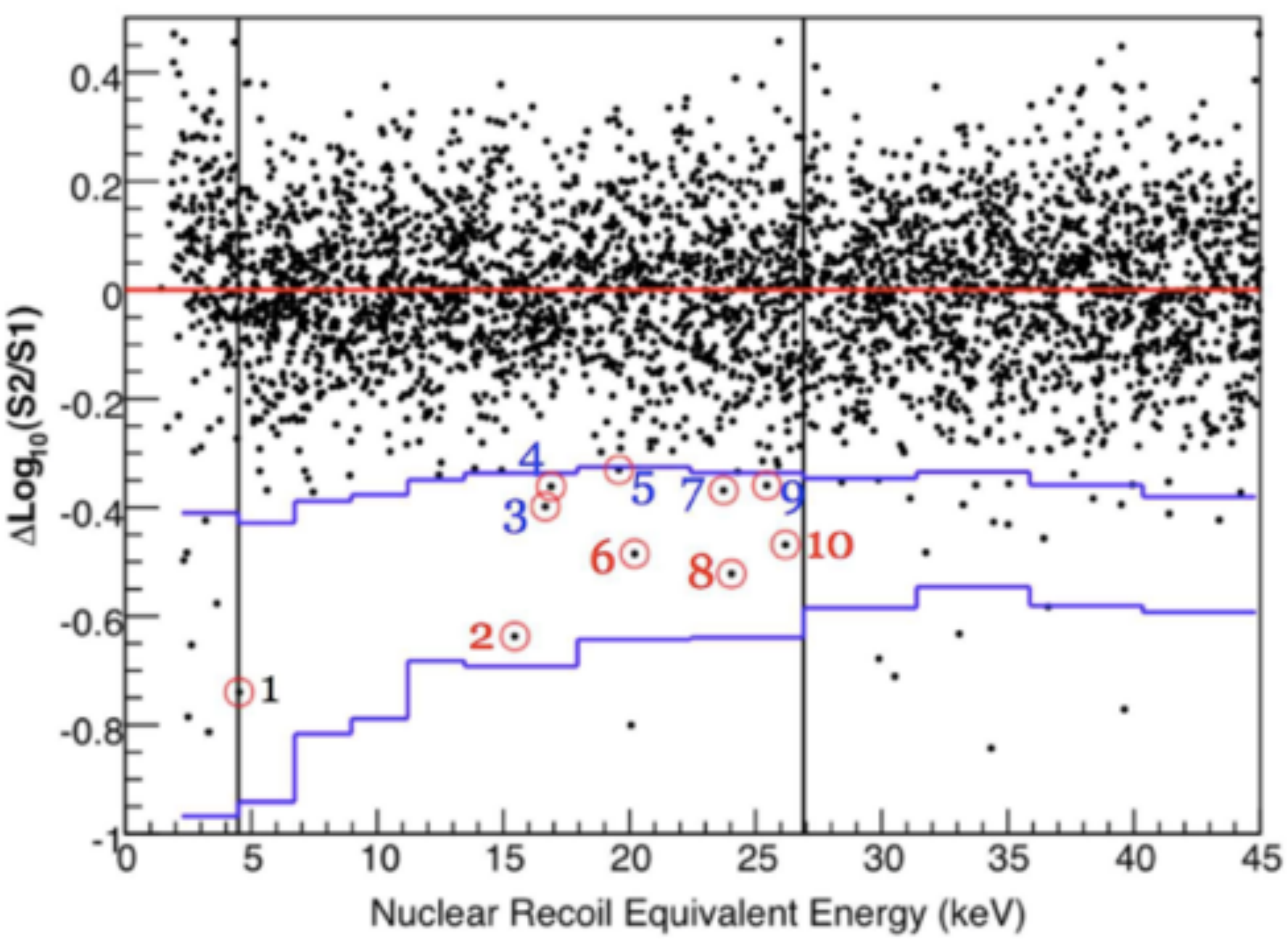
Yellin to the Rescue!



- The events don't match the expectation well
- Use the "maximum gap" and find the cross-section at which 90% of the trials have a gap smaller than this

More Yellin

- This can only set an upper limit, and can never be used for discovery
- Also generates one-sided (upper) limits
- No information about the background goes into this (it may be unknown)



UNIVERSITY OF TORONTO

Binned Maximum Likelihood

- We want to use all the information
- Assume we have one discriminator
- Bin the number of counts in that parameter

$$\mathcal{L} = \prod_{i=1}^{k} P(n_i|\mu(x))$$

$$ln\mathcal{L} = \sum_{i=1}^{k} ln(P(n_i|\mu(x)))$$

Binned Maximum Likelihood

$$ln\mathcal{L} = \sum_{i=1}^{k} ln(P(n_i|\mu(x)))$$

Use Poisson for low statistics (as before)

$$ln\mathcal{L} = \sum_{i=1}^{k} ln(\frac{e^{-\mu_i}\mu_i^{n_i}}{n_i!})$$

But introduce a new term for signal and background counts

$$\mu_x = S \cdot P_s(x) + B \cdot P_b(x)$$

Binned Maximum Likelihood

$$\mu_x = S \cdot P_s(x) + B \cdot P_b(x)$$

S and B are the hypothetical number of signal and background events in bin x

The likelihood can now be minimized to produce the best estimate of signal and background, which is used to set the limit

UNIVERSITY OF TORONTO

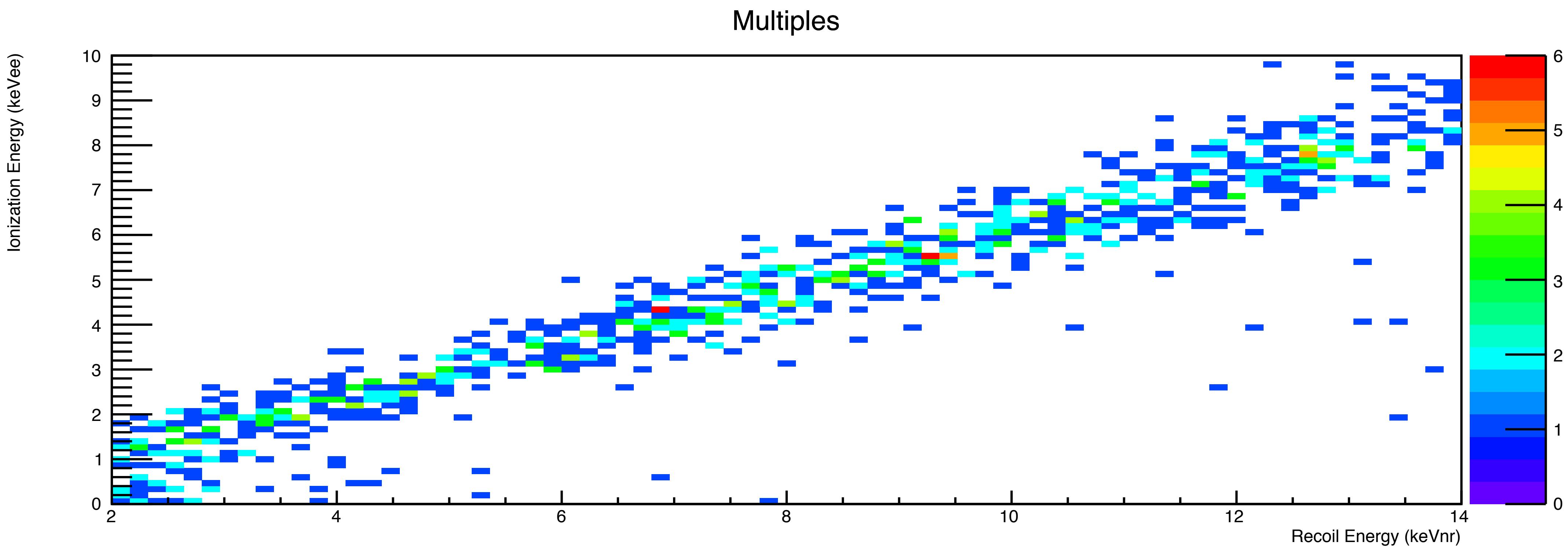
Let's Look at Data

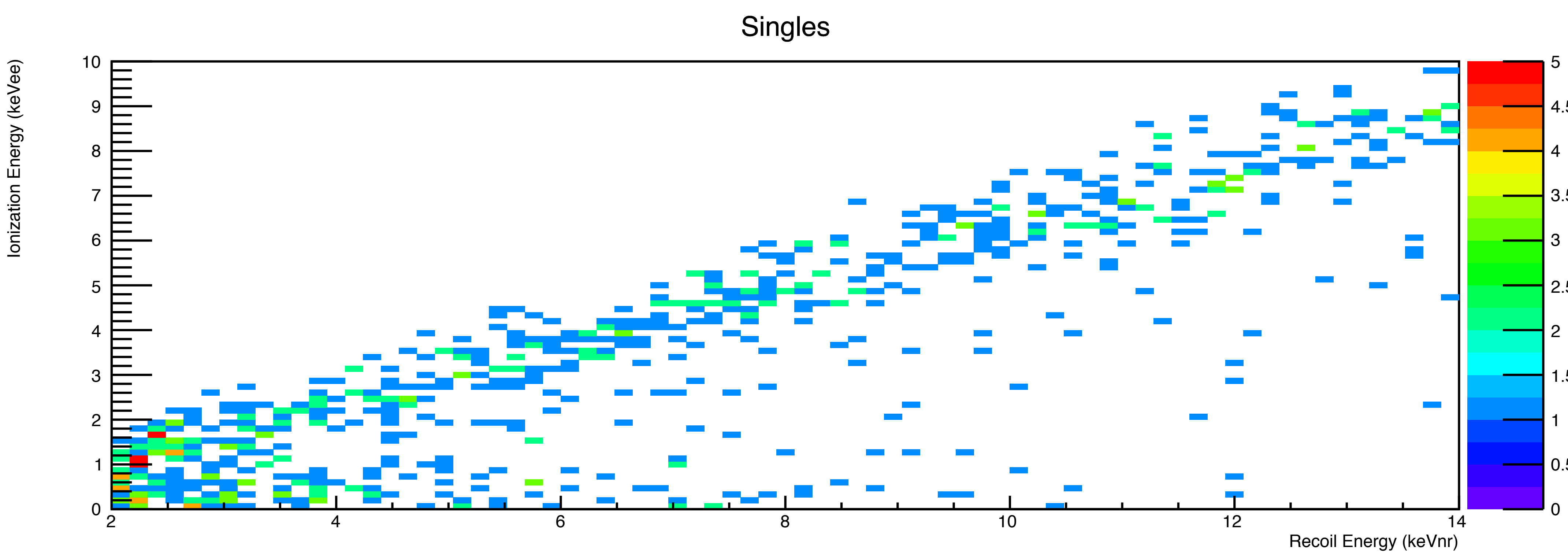
- This is CDMS data taken when I was a postdoc with them
- We have the single and multiple rates for one detector
- Use these to define backgrounds and signal

Don’t have DM Data… as such

- I have calibration data
- Split into “singles” and “multiples”
- Neutron events will happen in the “singles”
- Let’s define “singles” as calibration, “multiples” as data

CDMS Data

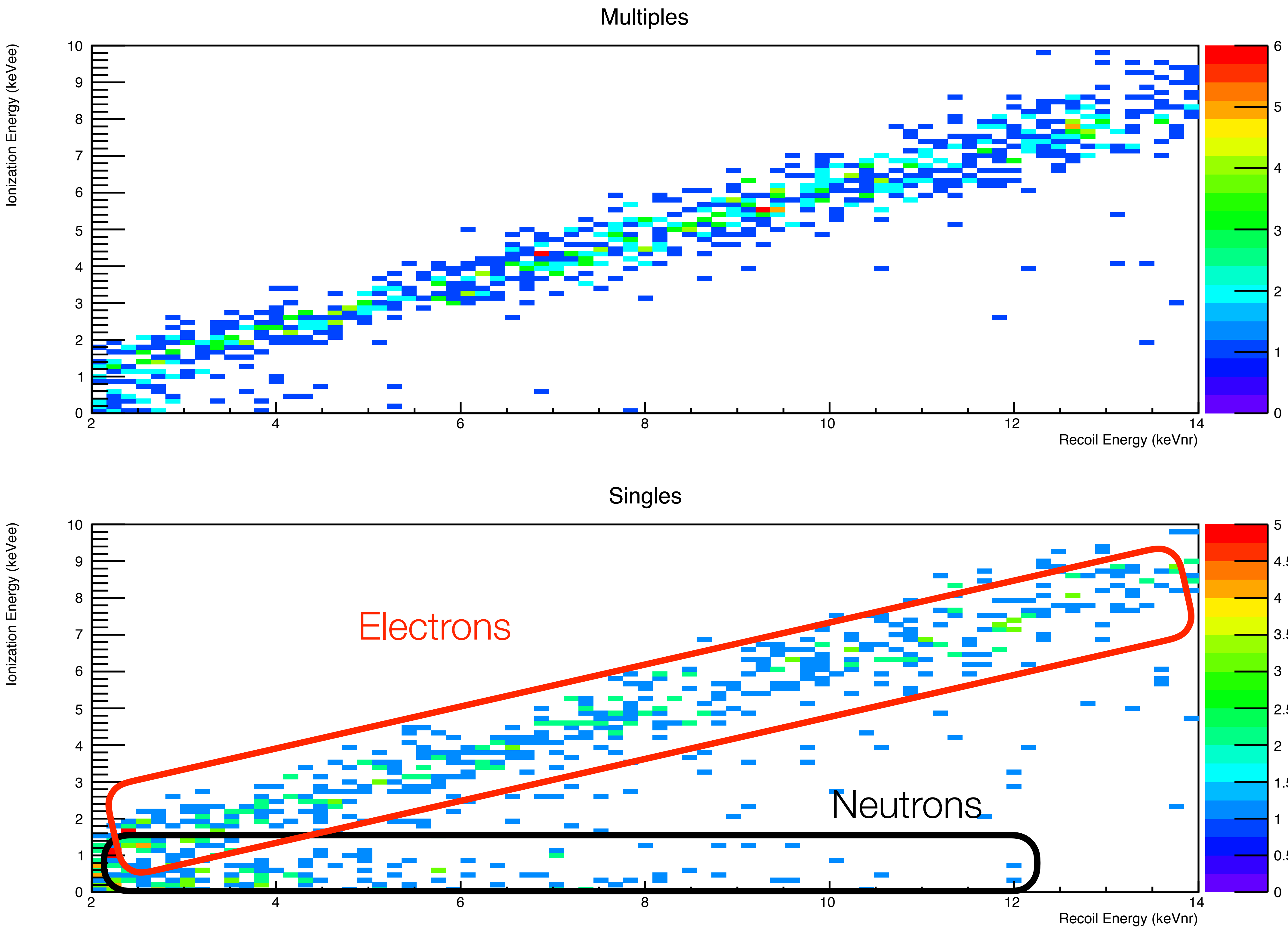




600g Germanium crystal, 1 day exposure

UNIVERSITY OF TORONTO

CDMS Data



600g Germanium crystal, 1 day exposure

Go!

UNIVERSITY OF TORONTO