Julia Kim
Monday, April 11, 2022

# Likelihood-Based Statistical Tests for Use in the Analysis of LIP Signals

## 1. Introduction

In a particle detector, incident particles hit a sensor, depositing amounts of energy which are read out by a data acquisition system. In frequentist testing, a important statistical test to perform upon these measurements is to establish the exclusion or the discovery of a desired *signal*. To do so, it is necessary to determine whether the measurements are compatible with the **null hypothesis**, $H_0$, which assumes the presence of only background, or the **alternative hypothesis**, $H_1$, which assumes the presence of both signal and background. Denoting the parameter of interest, the expected number of signal events, to be $n_s$, this one-tailed hypothesis test would be more briefly written as

$$\begin{cases} H_0 : n_s = 0 \\ H_1 : n_s > 0. \end{cases} \tag{1}$$

(Likewise, the expected number of background events, serving as a nuisance parameter, can be denoted by $n_b$). In order to measure the agreement between the data and $H_0$, a real-valued function of the data, called the **test statistic** $q_0$, is constructed. Under $H_0$, $q_0$ will be distributed according to a given **probability distribution function** (p.d.f), denoted $f(q_0|H_0)$. If this measure of agreement is to be formulated in terms of a practical decision to accept or reject $H_0$, the ***p*-value** for the null hypothesis, $p_0$, can be introduced. This is the probability, under the assumption of $H_0$, of obtaining data as equally or less compatible with $H_0$ than the one observed; in short,

$$p_0 = \int_{q_{0,obs.}}^{\infty} f(q_0|H_0)dq_0, \tag{2}$$

where $q_{0,obs.}$ is the value of $q_0$ observed from the experimental data set. In a hypothesis test with a **confidence level** of 95%, if the $p_0$-value should be less than the desired **significance level** $\alpha = 0.05$ of the test, then the chances to observe the measurements, given $H_0$, are very unlikely and the null hypothesis must be rejected [1]. With smaller and smaller $p_0$-values, there is stronger and stronger evidence compelling the discovery of a signal.

In the case, however, that $H_0$ is accepted and no signal can be discovered, a second important task is to establish an *upper limit* on the number of signal events that could be present in the data set [1]. To do so, various upper limits $\bar{n}_s$ are proposed and their test statistics, $\tilde{q}_{\bar{n}_s}$, denoting the level of agreement between the proposed upper limit and an estimated number of signal events in the data set, are computed. For each proposed $\bar{n}_s$, $\tilde{q}_{\bar{n}_s}$ will be distributed according to a p.d.f, $f(\tilde{q}_{\bar{n}_s}|\bar{n}_s)$. To establish whether the proposed upper limit is sufficiently high, their corresponding $p_{\bar{n}_s}$-values, each given by

$$p_{\bar{n}_s} = \int_{\tilde{q}_{\bar{n}_s,obs.}}^{\infty} f(\tilde{q}_{\bar{n}_s}|\bar{n}_s)d\tilde{q}_{\bar{n}_s}, \tag{3}$$

where $\tilde{q}_{\bar{n}_s,obs.}$ is the value of $\tilde{q}_{\bar{n}_s}$ observed in the same experimental data set, are calculated and plotted. The upper limit, $n_s^{upper}$, at a confidence level of 95% will be equal to the largest value of $\bar{n}_s$, whose $p_{\bar{n}_s}$-value remains greater than or equal to 0.05 [1].

In the present report, these two tests are performed in `RooFit` upon an "experimental" data set of $\sim 10^4$ energies (in keV) deposited by background sources alone. To speak more precisely, the data set is generated from a model p.d.f., given by the weighted sum of a **lightly ionising particle**[1] (LIP) signal distribution (acquired by sampling $10^6$ events from a theoretical LIP model in the `SuperCDMS supersim` package[2]), and a background distribution (acquired by sampling $10^6$ events from a spectrum in `CUTE bgexplorer`[2]), with the

[1]LIPs are hypothetical candidates of dark matter with fractional charge $q = \pm ze$, $z \in (0, 1)$, whose mean energy loss per unit length $\langle dE/dx \rangle$ is proportional to $z^2$ and is therefore much smaller than known minimum ionising particles under similar conditions [2].

[2]The files containing these randomly generated events were provided by Enze Zhang in January 2022.

coefficient multiplying the former distribution suppressed to 0. With the artificial advantage of producing a data set of pure background, it is evident that no discovery of a LIP signal will be made and, therefore, that an upper limit to the number of signal events at a 95% confidence level can be found. The p.d.f.s of the test statistics $q_0$ and $\tilde{q}_{\bar{n}_s}$ required in the analysis are obtained by generating 1,000 Monte Carlo toys from the model p.d.f., in which the number of events $N \sim \text{Poisson}(10^4)$ and in which the coefficients multiplying the LIP signal and the background distributions are appropriately defined. As these numbers $N$ are *"large enough"*, the p.d.f.s are expected to follow well-defined asymptotic behaviour, according to theorems due to Wilks and Wald. In this report, these asymptotic approximations are confirmed to model the distributions of the test statistics well, with reduced chi-squared ($\chi^2_{red.}$) values always close to 1.

## 2. Theory

Suppose the energies $E$ deposited by the signal and background sources to follow p.d.f.s $f_s(E|\boldsymbol{\theta}_s)$ and $f_b(E|\boldsymbol{\theta}_b)$, respectively, where $\boldsymbol{\theta}_s$ and $\boldsymbol{\theta}_b$ represent nuisance parameters that characterise their shapes. In every data set of observed energies $\boldsymbol{E}$, suppose the number of signal events and the number of background events to be Poisson random variables with means $n_s$ and $n_b$, respectively, so that the number of total events, $N$, might be a Poisson variable with mean $n = n_s + n_b$. The p.d.f. for the observable $E$ is thus

$$f(E|n_s, \boldsymbol{\theta}) = \frac{n_s f_s(E|\boldsymbol{\theta}_s) + n_b f_b(E|\boldsymbol{\theta}_b)}{n_s + n_b}, \tag{4}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_s, \boldsymbol{\theta}_b, n_b\}$ denotes the set of all nuisance parameters, whose values are not of primary interest in the analysis [1]. From this, the **extended likelihood function** for any observed data sample $\boldsymbol{E}$ can be constructed. It is given by the product of the probabilities to detect $N$ total events and to obtain the data $\boldsymbol{E}$,

$$L(\boldsymbol{E}|n_s, \boldsymbol{\theta}) = \frac{n^N}{N!} e^{-n} \prod_{i=1}^{N} f(E_i|n_s, \boldsymbol{\theta}). \tag{5}$$

A technique to estimate the optimal values of the parameters, given a data set $\boldsymbol{E}$, is called the method of **maximum likelihood** (ML). In this method, the estimators of the parameters are given by the unique solutions[3] to the likelihood equations,

$$\frac{\partial L}{\partial n_s} = 0 \text{ and } \frac{\partial L}{\partial \theta_i} = 0, i = 1, ..., m, \tag{6}$$

and are denoted with hats, $\hat{n}_s$, $\hat{\boldsymbol{\theta}}$, to distinguish them from the *true* values of the parameters [3]. Rather than using the likelihood function, it is often more convenient to use its logarithm or the negative of its logarithm. Indeed, the logarithm has the pleasant advantage of converting products to sums and reducing exponentials into simpler factors, as shown below,

$$\ln L(\boldsymbol{E}|n_s, \boldsymbol{\theta}) = N \ln(n) - n - \ln(N!) + \sum_{i=1}^{N} \ln f(E_i|n_s, \boldsymbol{\theta}), \tag{7}$$

which often facilitates optimisation. Since ln is a monotonically increasing function, the parameter values which maximise $L$ will also maximise $\ln(L)$ and minimise $-\ln(L)$.

From the likelihood function, and its ML estimators, the **profile likelihood ratio** can be written as

$$\lambda(n_s) = \frac{L(n_s, \hat{\hat{\boldsymbol{\theta}}}(n_s))}{L(\hat{n}_s, \hat{\boldsymbol{\theta}})}, \tag{8}$$

where $\hat{\hat{\boldsymbol{\theta}}}(n_s)$ is the **conditional ML estimator** of $\boldsymbol{\theta}$, which maximises $L$ for a given $n_s$, and $\hat{n}_s$ and $\hat{\boldsymbol{\theta}}$ are the **unconditional ML estimators** of $n_s$ and $\boldsymbol{\theta}$, maximising $L$ over the whole parameter space. As the denominator must exceed or be equal to the numerator, $0 \leq \lambda(n_s) \leq 1$. Note this ratio to be useful, for its

[3]If more than one local maximum exists, the *highest* is always taken.

magnitude serves as a measure of the agreement between a hypothesised $n_s$ and the most likely estimate of the quantity, $\hat{n}_s$, obtained from the data [1]. If, for instance, $n_s$ and $\hat{n}_s$ are close in value, then the likelihood function at $n_s$ and at $\hat{n}_s$ must be close in value and, therefore, $\lambda(n_s)$ must be close to 1.

Though many types of estimators exist, ML estimators are especially useful, being asymptotically unbiased and asymptotically normally distributed [4]. Accordingly, it is expected that, if $\hat{n}_s$ is permitted to assume negative values, and if the data sets $\boldsymbol{E}$ are large, then the sampling distribution of $\hat{n}_s$ must be a Gaussian with a certain mean $n'_s$ and standard deviation $\sigma$. Wishing to confirm and to apply this useful property, $\hat{n}_s$ will be regarded in this report as an *effective* estimator, for which negative values are permitted. With this convention, a **modified profile likelihood ratio** can be constructed to compare a hypothesised $n_s$ to the *closest physical value* of $\hat{n}_s$. This is $\hat{n}_s$ itself when $\hat{n}_s \geq 0$, and 0 when $\hat{n}_s < 0$, as shown below:

$$\tilde{\lambda}(n_s) = \begin{cases} \dfrac{L(n_s, \hat{\hat{\boldsymbol{\theta}}}(n_s))}{L(\hat{n}_s, \hat{\boldsymbol{\theta}})}, & \hat{n}_s \geq 0 \\ \dfrac{L(n_s, \hat{\hat{\boldsymbol{\theta}}}(n_s))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))} & \hat{n}_s < 0, \end{cases} \tag{9}$$

where $\hat{\hat{\boldsymbol{\theta}}}(0)$ is the conditional ML estimator of $\boldsymbol{\theta}$, given $n_s = 0$.

These likelihood ratios $\lambda(n_s)$ and $\tilde{\lambda}(n_s)$ are required in the definitions of the test statistics $q_0$ and $\tilde{q}_{n_s}$, introduced in the two sections below.

### 2.1. Test Statistic for the Discovery of a Positive Signal

To measure the agreement between the data and $H_0$, the following test statistic $q_0$ is constructed:

$$q_0 = \begin{cases} -2\ln(\lambda(0)), & \hat{n}_s \geq 0 \\ 0 & \hat{n}_s < 0. \end{cases} \tag{10}$$

Here, $q_0$ is always non-negative and is *larger* when there is *greater* disagreement between $\hat{n}_s$ and $n_s = 0$ [1]. Indeed, if $\hat{n}_s \gg 0$, then $\lambda(0) \ll 1$ and so $q_0 \gg 0$. If $\hat{n}_s < 0$, then the most likely estimate of the number of signal events is *fewer* than it could physically be, supporting the claim of there being only background events in the data. As there is no disagreement between the conclusions inferred from $\hat{n}_s < 0$ and $n_s = 0$, $q_0$ is accordingly set to its *smallest* value of 0.

### 2.2. Test Statistic for Upper Limits

If the data set is compatible with $H_0$, it is instructive to establish an upper limit on the number of signal events that could be present. For every proposed upper limit, $\bar{n}_s$, the following test statistic $\tilde{q}_{\bar{n}_s}$ is constructed:

$$\begin{aligned} \tilde{q}_{\bar{n}_s} &= \begin{cases} -2\ln(\tilde{\lambda}(\bar{n}_s)), & \hat{n}_s \leq \bar{n}_s \\ 0 & \hat{n}_s > \bar{n}_s. \end{cases} \\ &= \begin{cases} -2\ln \dfrac{L(\bar{n}_s, \hat{\hat{\boldsymbol{\theta}}}(\bar{n}_s))}{L(0, \hat{\hat{\boldsymbol{\theta}}}(0))}, & \hat{n}_s < 0 \\ -2\ln \dfrac{L(\bar{n}_s, \hat{\hat{\boldsymbol{\theta}}}(\bar{n}_s))}{L(\hat{n}_s, \hat{\boldsymbol{\theta}})}, & 0 \leq \hat{n}_s \leq \bar{n}_s \\ 0 & \hat{n}_s > \bar{n}_s. \end{cases} \end{aligned} \tag{11}$$

Similar to the preceding, $\tilde{q}_{\bar{n}_s}$ is always non-negative and is *larger* when there is greater discrepancy between $\bar{n}_s$ and the closest physical value to $\hat{n}_s$ [1]. To illustrate, if $\hat{n}_s \ll 0$ or $0 \leq \hat{n}_s \ll \bar{n}_s$, then $\tilde{\lambda}(\bar{n}_s) \ll 1$ and so $\tilde{q}_{\bar{n}_s} \gg 0$.

To establish an upper limit at a confidence level of 95%, the discrepancy between $\bar{n}_s$ and the closest physical value of $\hat{n}_s$ must be made greater and greater, so that the test statistics $\tilde{q}_{\bar{n}_s}$ might be made larger and larger and so that the $p_{\bar{n}_s}$-values, defined in (3), can *diminish* to values closer and closer to 0.05. If

$\hat{n}_s > \bar{n}_s$, it is clear that the proposed $\bar{n}_s$ cannot be sufficiently high to serve as an adequate upper limit to the number of signal events in the data set. In such a case, $\tilde{q}_{\bar{n}_s}$ is accordingly set to its *smallest* value of 0, so that the $p_{\bar{n}_s}$-value can be equal to its *largest* value of 1; this implies there to be no possibility of it being close to the desired upper limit.

### 2.3. Statement of the Asymptotic Approximations (Without Proof)

If ML estimators $\hat{n}_s$ have a limiting normal distribution and if the sizes of the data samples are large, then these two test statistics of interest will exhibit well-defined asymptotic behaviour [1]. Indeed, according to **Wald's Theorem**, if $\hat{n}_s \sim Normal(n_s', \sigma)$, then the profile likelihood function can be written as

$$-2\ln(\lambda(n_s)) = \frac{(n_s - \hat{n}_s)^2}{\sigma^2} + \mathcal{O}\left(\frac{1}{N}\right), \tag{12}$$

where $N$ is the data sample size. Supposing $N$ to be large, the error term $\mathcal{O}\left(\frac{1}{N}\right)$ can be neglected entirely and expression (10) can be reduced to

$$q_0 = \begin{cases} \hat{n}_s^2/\sigma^2, & \hat{n}_s \geq 0 \\ 0, & \hat{n}_s < 0. \end{cases} \tag{13}$$

As $q_0$ is a squared standard normal variable, $q_0$ follows a *chi-square* distribution, $\chi_1^2$, with a degree of freedom of 1, a result shown also by **Wilks**[4]. For ease in computation, its distribution will be given by its *"finite"* approximation,

$$f(q_0|n_s = 0) = \frac{1}{2}\delta(q_0) + \frac{1}{2}\frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{q_0}}e^{-q_0/2}, \tag{14}$$

which does not diverge at $q_0 = 0$, but is instead is equal to $\frac{1}{2}$.

Likewise, with **Wald's Theorem**, expression (11) can be reduced to

$$\tilde{q}_{\bar{n}_s} = \begin{cases} \dfrac{\bar{n}_s^2}{\sigma^2} - \dfrac{2\bar{n}_s\hat{n}_s}{\sigma^2}, & \hat{n}_s \geq 0 \\ \dfrac{(\bar{n}_s - \hat{n}_s)^2}{\sigma^2}, & 0 \leq \hat{n}_s \leq \bar{n}_s \\ 0 & \hat{n}_s > \bar{n}_s, \end{cases} \tag{15}$$

with the distribution again given by its *"finite"* approximation,

$$f(\tilde{q}_{\bar{n}_s}) = \frac{1}{2}\delta(\tilde{q}_{\bar{n}_s}) + \begin{cases} \dfrac{1}{2}\dfrac{1}{\sqrt{2\pi}}\dfrac{1}{\sqrt{\tilde{q}_{\bar{n}_s}}}e^{-\tilde{q}_{\bar{n}_s}/2}, & 0 < \tilde{q}_{\bar{n}_s} \leq \dfrac{\bar{n}_s^2}{\sigma^2} \\ \dfrac{1}{\sqrt{2\pi}(2\bar{n}_s/\sigma)}e^{-\frac{1}{2}\frac{(\tilde{q}_{\bar{n}_s} + \bar{n}_s^2/\sigma^2)^2}{(2\bar{n}_s/\sigma)^2}}, & \tilde{q}_{\bar{n}_s} > \dfrac{\bar{n}_s^2}{\sigma^2}. \end{cases} \tag{16}$$

Supposing this distribution to hold, the theoretical value of the upper limit on $n_s$ for an observed data sample $\boldsymbol{E}$ at a confidence level of $\alpha$ is given by

$$n_s^{upper} = \hat{n}_{s,obs.} + \sigma\Phi^{-1}(1-\alpha), \tag{17}$$

where $\Phi^{-1}$ is the inverse of the cumulative distribution function of the standard Gaussian. Further details concerning these equations can be found in [1].

---

[4] In particular, **Wilks' theorem** declares that, if the maximum likelihood estimators have a limiting normal distribution, and if the sample size of each data set is large, then the distribution of $-2\ln(\lambda)$ to converge to a $\chi^2$ distribution with a number of degrees of freedom determined by $\dim(\Theta) - \dim(\Theta_0)$, where $\Theta$ and $\Theta_0$ are the full and the null parameter spaces of the model p.d.f., respectively [5]. In the present case, $f(E|n_s, \boldsymbol{\theta})$ is the model p.d.f., $\dim\Theta = |\{n_s\}| + |\boldsymbol{\theta}| = 1 + |\boldsymbol{\theta}|$ and $\dim\Theta_0 = |\boldsymbol{\theta}|$ (as it is supposed, under $H_0$, that $n_s = 0$). Therefore, $-2\ln(\lambda)$ must converge to a $\chi^2$ distribution with a degree of freedom of $\dim(\Theta) - \dim(\Theta_0) = 1 + |\boldsymbol{\theta}| - |\boldsymbol{\theta}| = 1$.

# 3. Results

## 3.1. "Experimental" Data Set

To generate the "experimental" data set upon which these two tests are performed, the model p.d.f. governing the distribution of the deposited energies must first be defined. In `RooFit`, the `supersim` and `CUTE bgexplorer` files, containing $10^6$ LIP signal and background energy deposition events, are imported. The observable `E` is defined as a `RooRealVar`,

```
RooRealVar E("E", "E", 0.1, 10);
```

thereby restricting the energy region of interest (ROI) between 0.1 and 10 keV. With `RooDataHist`, the LIP signal and background events within the ROI are arranged into 3,100 bins of two histograms, which are converted, by means of linear interpolation, into two normalised `RooHistPdf` distributions. These are shown in Figure 1, below.

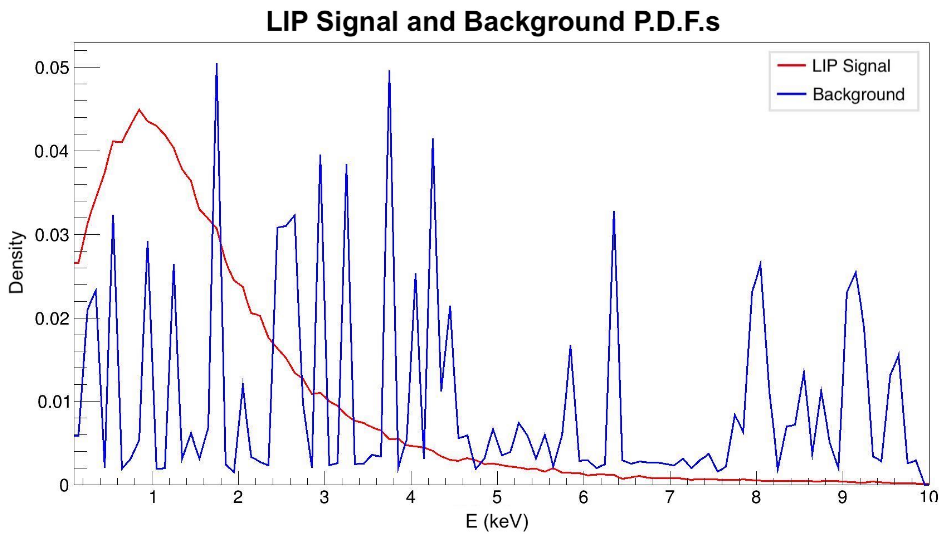


Figure 1: The `RooHistPdf` LIP signal and background distributions.

Having constructed the signal and background distributions, denoted `sig` and `bgd`, respectively, the expected number of signal and background events in every generated data set must be specified. For this purpose, two independent and uncorrelated `RooRealVar` objects, `n_s` and `n_b`, representing the (expected) number of LIP signal and background events,

```
RooRealVar n_s("num_sig","n_s", 0, -2000, 20000);
RooRealVar n_b("num_bgd","n_b", 10000, -2000, 20000);
```

are defined. Wishing to produce a data set of pure background, the initial values of `n_s` and `n_b` are made equal to `0` and `10000`, respectively.[5] Assembling the p.d.f.s and the coefficients into two `RooArgLists`, denoted `components` and `coeffs`, respectively, and passing them into `RooAddPdf`, the model p.d.f.

```
RooAddPdf model("model", "f_{s+b}", components, coeffs);
```

is at last constructed. Observe it to simply correspond to the background distribution, `bgd`, previously defined. To generate a `RooDataSet` consisting of the "experimental" data,

```
RooDataSet *newdata_exp = model.generate(E);
```

[5]Further, the ranges of `n_s` and `n_b` are between `-2000` and `-20000`, though there are many others. Indeed, it is only required that the upper bound be sufficiently larger than 10,000 and the the lower bound be sufficiently smaller than 0. The upper bound must be sufficiently large, as, Poisson fluctuations could cause the generated data sets, depending upon the model p.d.f., to contain more than 10,000 signal or background events. The lower bound must be sufficiently negative, as the ML estimators of these two parameters must be Gaussian distributed.

is required, producing $\sim 10^4$ simulated energy values within the ROI.

From this, the extended negative log likelihood function, `nll`, for `*newdata_exp` can be constructed.

```
RooAbsReal* nll = model_exp.createNLL(*newdata_exp, Extended(), NumCPU(4));
```

As the `nll` is generated in `Extended()` mode, the number of events in any generated data set, and thus in `*newdata_exp`, becomes a Poisson variable with a mean of `10000`. Wishing to acquire the ML estimator of `n_s`, `RooMinuit(*nll).migrad()` is applied upon the extended negative log likelihood function, thereby minimising it. The plot of the function, after `migrad()` is applied, is contained in Figure 2 below.

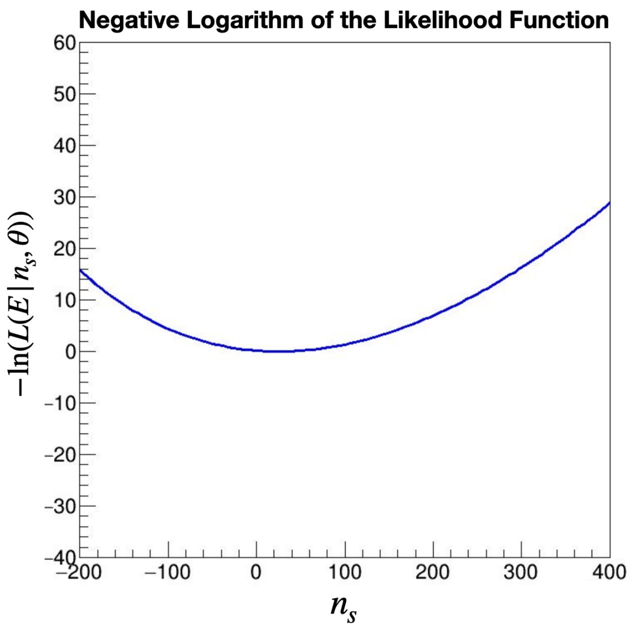


Figure 2: Plot of the negative log of the likelihood function for the data set `*newdata_exp`. The domain of the plot has been restricted from $-200$ to 400, in order that its minimum might be clearly seen.

Though all values within the range of `n_s` from `-2000` to `20000` are considered possible minimising points, the ML estimator, `n_s_Hat`,

```
double n_s_Hat = n_s.getVal();
```

is found, in particular, to be equal to `40.2489`. Accordingly, $\boxed{\hat{n}_{s,obs} = 40.2489}$.

Wishing to determine the null test statistic, `createProfile` is used to generate the negative logarithm of the profile likelihood ratio, `pll_n_s`, in function of `n_s`. To evaluate this `RooAbsReal` function at desired points, `asTF` is used to convert it to a `TF1` object, called `pll`, as follows:

```
RooAbsReal* pll_n_s = nll->createProfile(n_s);
TF1 *pll = pll_n_s -> asTF(n_s);
```

Using Equation (10), the test statistic, stored in the list `Double_t q_0_obs[1] = {}`, is therefore given by

```
if (n_s_Hat.getVal() >= 0){
    Double_t q0 = 2*pll->Eval(0);
    q_0_obs[0] = q0;
} else{
    q_0_obs[0] = 0;
}
```

for which 0.6856 is obtained. Accordingly, $\boxed{q_{0,obs.} = 0.6856}$.

Observing $\hat{n}_{s,obs} \sim 40$, these 15 possible upper limits to the number of signal events in *newdata_exp are stored in the list

```
Double_t n_s_up[15] = {50,60,70,80,90,100,110,115,120,125,130,135,140,145,150};
```

The observed test statistic q_n_s_obs[i] corresponding to each proposed upper limit n_s_up[i], for 0≤i≤14, will be stored in the list

```
Double_t q_n_s_obs[15] = {};
```

Using Equation (11), this test statistic is given by

```
if (n_s_Hat < 0){
  Double_t q_val = 2*((pll->Eval(n_s_up[i])) - (pll->Eval(0)));
  q_n_s_obs[i] = q_val;
} else if(n_s_Hat > n_s_up[i]){
  q_n_s_obs[i] = 0;
} else{
  Double_t q_val = 2*(pll->Eval(n_s_up[i]));
  q_n_s_obs[i] = q_val;
}
```

In printing the list q_n_s_up_obs, the observed test statistics in Table 1 below are obtained.

Table 1: Observed Test Statistics at Proposed Upper Limits

| $\bar{n}_s$ | $\tilde{q}_{\bar{n}_s,obs.}$ |
|---|---|
| 50 | 0.0406 |
| 60 | 0.1626 |
| 70 | 0.3645 |
| 80 | 0.6448 |
| 90 | 1.0020 |
| 100 | 1.4348 |
| 110 | 1.9419 |
| 115 | 2.2228 |
| 120 | 2.5218 |
| 125 | 2.8387 |
| 130 | 3.1734 |
| 135 | 3.5257 |
| 140 | 3.8955 |
| 145 | 4.2826 |
| 150 | 4.6869 |

### 3.2. Test for the Discovery of a Positive Signal

To obtain the $p_0$-value to which $q_{0,obs.} = 0.6856$ corresponds, the null test statistic distribution $f(q_0|H_0)$ must first be constructed. For this purpose, 1,000 Monte Carlo toy data sets are generated from the model p.d.f., model, previously defined. Much of the procedure in section 3.1, excepting the part concerning the upper limits, is performed upon every toy. With every iteration, the ML estimator of n_s is stored in a TH1 histogram of 20 bins, as shown in Figure 3. Using the code implementing Equation (10), the null test statistic of this toy is then generated and stored in a TH1 histogram of 30 bins. The histogram is thereafter converted, by means of linear interpolation, into a normalised RooHistPdf distribution q_0_pdf, as shown in Figure 4. At the end of the iteration, n_s is reset to its initial value of 0.

To obtain the $p_0$-value, it suffices to determine the area under $f(q_0|H_0)$ to the *right* of $q_{0,obs.}$. To do so, createIntegral is applied upon q_0_pdf to integrate it from q_0_obs to "infinity" (or, in practice, to the largest generated null test statistic, 10.5271). In doing so, $\boxed{p_0 = 0.2033}$, which certainly exceeds 0.05.

Accordingly, $H_0$ is accepted at a confidence level of 95%, as desired. In other words, no LIP signal can be discovered in the "experimental" data set at a significance level of $\alpha = 0.05$.

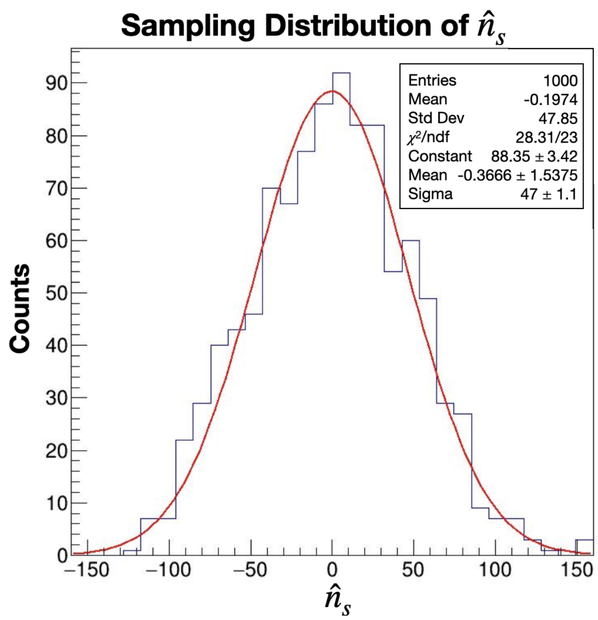


Figure 3: (Non-normalised) sampling distribution of 1,000 ML estimators, fitted with a Gaussian. These are acquired from Monte Carlo toys generated with the model p.d.f. for which $n_s = 0$ and $n_b = 10,000$.

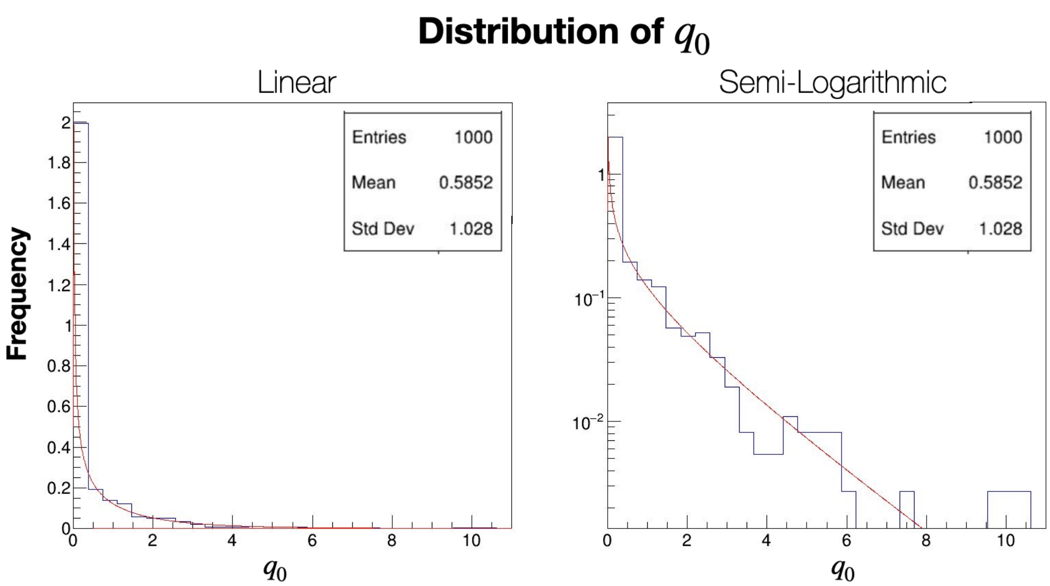


Figure 4: Null test statistic distribution on a linear (*left*) and semi-logarithmic (*right*) scale, fitted with Equation (14). Its $\chi^2_{red.} = 1.38$.

A note upon the fitted functions in Figures 3 and 4 is presently in order. Each bin in every histogram is presumed to follow a Poisson distribution, whose uncertainty is given by the Poisson counting rule: this would be $\sqrt{\text{number of counts}}$, when the histogram is non-normalised, and $C \cdot \sqrt{\text{number of counts}}$, when the histogram is normalised (having been converted from its non-normalised form by a normalisation factor, $C$).

With these errors taken into account, the Gaussian function,

$$\text{counts} = \hat{A}_0 \exp\{-(\hat{n}_s - \hat{\mu}_0)^2/2\hat{\sigma}_0^2\} = (88 \pm 3) \exp\{-(\hat{n}_s - (-0.4 \pm 1.5))^2/2(47 \pm 1)^2\}, \tag{18}$$

where $\hat{A}_0, \hat{\mu}_0, \hat{\sigma}_0 \in \mathbb{R}$ represent the least-squares estimates of its scaling constant, mean and standard deviation, is fitted to the sampling distribution of ML estimators in Figure 3. As its $\chi^2_{red.} = 28.31/23 \simeq 1.23$ is close to the ideal value of 1, the fit must describe the histogram *well*. Having confirmed the ML estimators to be likely normally distributed, it remains, according to Wilks and to Wald, to try to fit the $\chi^2_1$ distribution in equation (14) to $f(q_0|H_0)$. This is done in Figure 4. As its $\chi^2_{red.} = 1.38 \sim 1$, the fit must describe the null test statistic distribution *well*, thereby *validating* the asymptotic approximation.

With these fits, an estimate theoretical value of the upper limit, given in equation (17), can be computed

$$n_s^{upper} = \hat{n}_{s,obs.} + \sigma\Phi^{-1}(1 - 0.05) = 40.2489 + (47)(1.645) = 117.6.$$

Therefore, one must consider values near 120 as possible upper limits, for which are chosen 15 numbers between 50 and 150. For those numbers close to 120, they are separated by increments of 5, rather than 10.

### 3.3. Test for Upper Limits

Consider the proposed upper limit `n_s_up[i]`, for some `0`$\leq$`i`$\leq$`14`. To obtain the $p_{\bar{n}_s}$-value to which `q_n_s_obs[i]` corresponds, the test statistic distribution at this proposed upper limit must be defined. For this purpose, the definitions of `n_s` and `n_b` must be modified to

```
RooRealVar n_s("num_sig","n_s", n_s_up[i], -2000, 20000);
RooRealVar n_b("num_bgd","n_b", 10000 - n_s_up[i], -2000, 20000);
```

1,000 Monte Carlo toy data sets are then generated from the model p.d.f., `model`, in which these new definitions of `n_s` and `n_b` are passed as coefficients in a `RooArgList`. With every iteration, the ML estimator of `n_s` is acquired and stored in a `TH1` histogram of 20 bins. Using the code implementing Equation (11), the associated test statistic for the toy is generated and stored in another `TH1` histogram of 30 bins. The histogram is subsequently converted into a normalised `RooHistPdf` distribution, `q_n_s_pdf`. At the end of the iteration, `n_s` is reset to its initial value of `q_n_s_obs[i]`. To obtain the $p_{\bar{n}_s}$-value, `createIntegral` is applied upon `q_n_s_pdf` to integrate it from `q_n_s_obs[i]` to the "infinity" (or, in practice to largest generated test statistic).

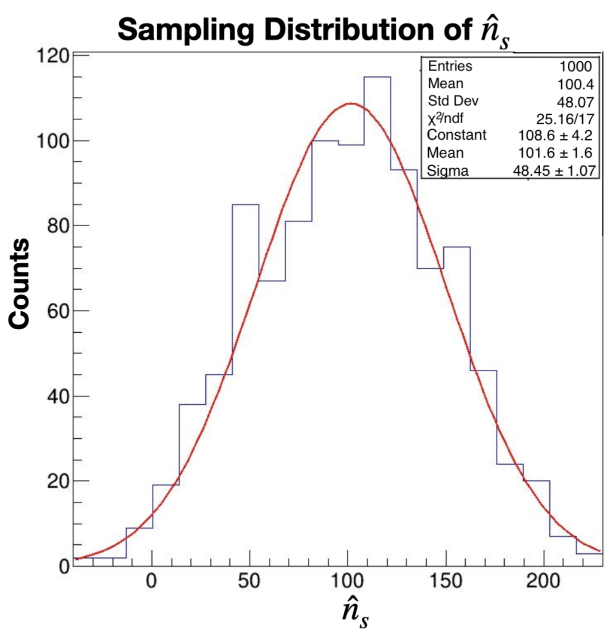


Figure 5: (Non-normalised) sampling distribution of 1,000 ML estimators, fitted with a Gaussian. These are acquired from Monte Carlo toys generated with the model p.d.f., for which $n_s = 100$ and $n_b = 9900$.

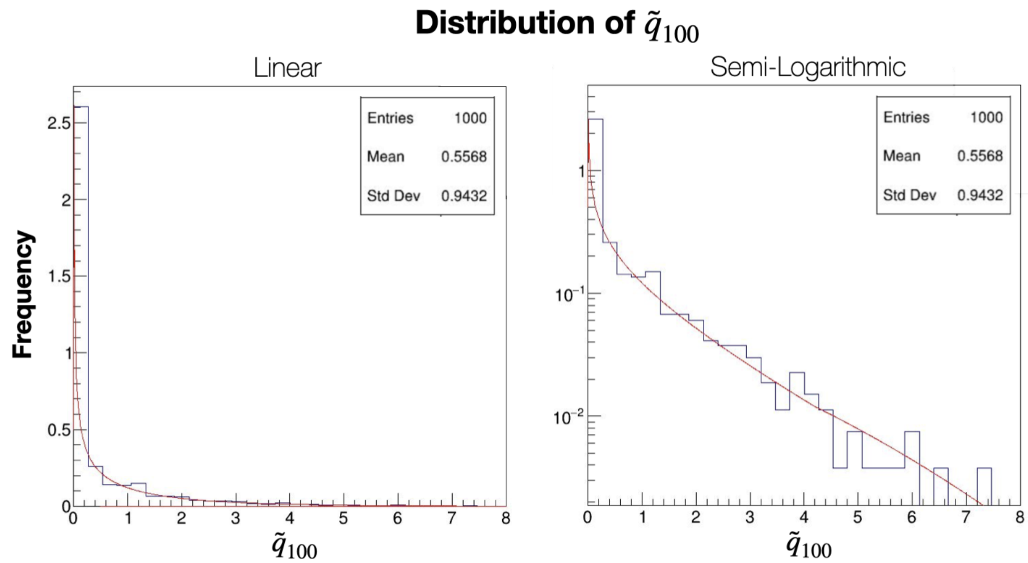


Figure 6: $\tilde{q}_{100}$ test statistic distribution on a linear (*left*) and semi-logarithmic (*right*) scale, fitted with Equation (16). Its $\chi^2_{red.} = 1.31$.

The ML estimator and test statistic distributions, in the case that $\bar{n}_s = 100$, are shown in Figures 5 and 6 above. When $f(\tilde{q}_{100}|100)$ is integrated from $\tilde{q}_{100,obs.} = 1.4348$ to "infinity" (or to the largest generated test statistic, 7.2128), $\boxed{p_{100} = 0.1166}$, which is certainly larger than 0.05. Accordingly, it can be inferred that the upper limit on $n_s$ in the experimental data set must be somewhat larger than 100. With respect to the fitted functions, observe in Figure 5 that the Gaussian function

$$\text{counts} = \hat{A}_{100} \exp\{-(\hat{n}_s - \hat{\mu}_{100})^2/2\hat{\sigma}^2_{100}\} = (109 \pm 4) \exp\{-(\hat{n}_s - (102 \pm 2))^2/2(48 \pm 1)^2\}, \tag{19}$$

where $\hat{A}_{100}, \hat{\mu}_{100}, \hat{\sigma}_{100} \in \mathbb{R}$ represent the least-squares estimates of its scaling constant, mean and standard deviation, is fitted to the sampling distribution of ML estimators. As its $\chi^2_{red.} = 25.16/17 \simeq 1.48$ is close to 1, the fit must describe the histogram *well*. Having again confirmed the ML estimators to be likely normally distributed, it remains, according to Wilks and to Wald, to try to fit the asymptotic distribution in Equation (16), where $\bar{n}_s = 100$ and $\sigma = \hat{\sigma} \simeq 48$, to $f(\tilde{q}_{100}|100)$. This is done in Figure 6. As its $\chi^2_{red.} = 1.31 \sim 1$, the fit must adequately describe the test statistic distribution, thereby *validating* the asymptotic approximation. Note the plots for all other proposed upper limits to be highly similar. The equations of the fitted Gaussians as well as the goodness of fit of the theoretical functions to the ML and test statistic distributions of all other proposed limits are contained in Table 2 in the Appendix. In this table, it can be observed that the asymptotic approximations are always confirmed to hold, with $\chi^2_{red.}$ values always close to 1.

At last, to determine the desired upper limit $n_s^{upper}$ at a 95% confidence level, the $p_{\bar{n}_s}$-values are plotted against the proposed $\bar{n}_s$ values, as shown in Figure 7. In using the definition in [1], the desired upper limit must be 115, for it is the largest proposed value whose $p_{\bar{n}_s}$-value of 0.06163 remains greater than or equal to 0.05. However, with linear interpolation between the points (115, 0.06163) and (120, 0.04974), this estimate can be much *refined*. With this technique, the desired upper limit, at which the $p$-value is equal to 0.05, is $\boxed{n_s^{upper} = 119.9}$ or, to the nearest integer, 120, concluding this statistical test. It is instructive to note this value to be near to the expected value of 117.6, obtained with equation (17).

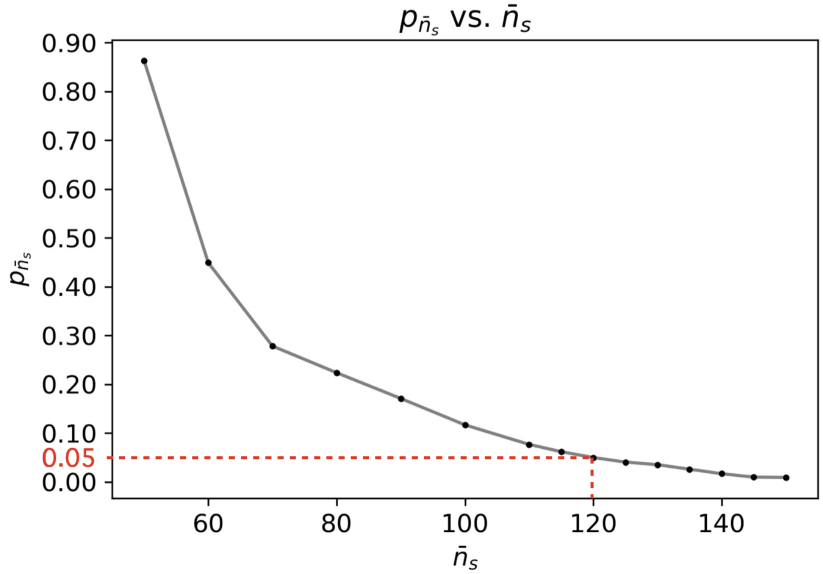


Figure 7: Plot of the $p_{\bar{n}_s}$-values against the proposed $\bar{n}_s$ values. The desired upper limit is found with linear interpolation, as shown in red. The precise $p_{\bar{n}_s}$-values appear in Table 2 of the APPENDIX.

## 4. CONCLUSION

In `RooFit`, two likelihood-based statistical tests are performed upon an "experimental" data set $\boldsymbol{E}$ of $\sim 10^4$ background energy depositions between 0.1 and 10 keV. In the first test, $\boldsymbol{E}$ is found to be *compatible* with the null hypothesis, $H_0$, at a confidence level of 95%. Indeed, its $p_0$-value is discovered to be 0.2033, which certainly exceeds the $\alpha$-threshold of 0.05. The distribution of ML estimators of $n_s$ and the null test statistic distribution, $f(q_0|n_s = 0)$, are generated using 1,000 Monte Carlo toys and are found to excellently match their asymptotic predictions (of being normally and $\chi_1^2$-distributed, respectively) under the conditions required by Wilks and Wald. In the second test, the upper limit to the number of signal events present in $\boldsymbol{E}$, at a confidence level of 95%, is found to be 119.9. This number is certainly close to the expected value of 117.6, derived from asymptotics. At every proposed upper limit, $\bar{n}_s$, the distribution of ML estimators and the test statistic distribution, $f(\tilde{q}_{\bar{n}_s}|\bar{n}_s)$, are generated using 1,000 Monte Carlo toys and are determined to again excellent match their asymptotic predictions (of being normally distributed and being distributed according to Equation (16), respectively). Accordingly, this signifies the choice of the size $N \sim Poisson(10^4)$ of the Monte Carlo toys is *large* enough for the purposes of testing **Wilks'** and **Wald's Theorems**.

In practice, these statistical tests are performed not upon measurements, known to be composed of background alone, but upon *real* data sets, whose proportion of (dark matter) signal to background is yet to be discovered. However, much of the same concepts and methods discussed in this paper can be applied to the analysis of physical measurements, for the purposes of signal discovery and limit setting.

## 5. REFERENCES


[1] McLean , K. (2014, August). *Implementing Likelihood-based Statistical Tests using a Toy Model in RooFit and RooStats and Exploring the Basics of TMA*. Retrieved March 23, 2022, from here.

[2] Alkhatib I et al.; SuperCDMS Collaboration. *Constraints on Lightly Ionizing Particles from CDMSlite.* Phys Rev Lett. 2021 Aug 20;127(8):081802. doi: 10.1103/PhysRevLett.127.081802. PMID: 34477436.

[3] Cowan, G. (1998). The method of maximum likelihood. In *Statistical Data Analysis* (1st ed., pp. 70–93). Clarendon Press.

[4] Watkins , J. (2011, November). *Topic 15: Maximum Likelihood Estimation.* Retrieved March 23, 2022, from here.

[5] Bonakdarpour, M. (2016, January 14). *Likelihood Ratio: Wilks's Theorem.* fiveMinuteStats. Retrieved March 22, 2022, from here.

## 6. Appendix

At every proposed upper limit, $\bar{n}_s$, the Gaussian function

$$\text{counts} = \hat{A}_{\bar{n}_s} \exp\{-(\hat{n}_s - \hat{\mu}_{\bar{n}_s})^2 / 2\hat{\sigma}^2_{\bar{n}_s}\}, \tag{20}$$

where $\hat{A}, \hat{\mu}, \hat{\sigma} \in \mathbb{R}$ represent the least-squares estimates of its scaling constant, mean and standard deviation, is fitted to the sampling distribution of its 1,000 ML estimators, $\hat{n}_s$. Likewise, equation (16), with the given $\bar{n}_s$ and with $\sigma = \hat{\sigma}$, is fitted to the generated test statistic distribution, $f(\tilde{q}_{\bar{n}_s}|\bar{n}_s)$. The least-squares estimates of the parameters of the Gaussian function and the goodness of fit of these two fitted functions appear in Table 2 below. The corresponding $p_{\bar{n}_s}$ values at every proposed limit are also contained in the table.

Table 2: $p$-values and Asymptotic Fits to Distributions at Proposed Upper Limits

| $\bar{n}_s$ | $p_{\bar{n}_s}$ | **ML Sampling Distribution** | | | | **Test Statistic Distribution** |
|---|---|---|---|---|---|---|
| | | **Fitted Parameters** | | | $\chi^2_{red.}$ | $\chi^2_{red.}$ |
| | | $\hat{A}_{\bar{n}_s}$ | $\hat{\mu}_{\bar{n}_s}$ | $\sigma_{\bar{n}_s}$ | | |
| 50 | 0.8626 | 112 ± 4 | 50 ± 2 | 48 ± 1 | 1.51 | 1.31 |
| 60 | 0.4488 | 116 ± 4 | 59 ± 1 | 50 ± 1 | 1.42 | 1.35 |
| 70 | 0.2781 | 116 ± 4 | 70 ± 2 | 50 ± 1 | 1.46 | 1.36 |
| 80 | 0.2235 | 116 ± 5 | 80 ± 2 | 50 ± 1 | 1.48 | 1.34 |
| 90 | 0.1707 | 118 ± 3 | 90 ± 1 | 50 ± 1 | 1.49 | 1.33 |
| 100 | 0.1166 | 109 ± 4 | 102 ± 2 | 48 ± 1 | 1.48 | 1.31 |
| 110 | 0.0766 | 112 ± 4 | 110 ± 2 | 50 ± 1 | 1.52 | 1.27 |
| 115 | 0.0616 | 114 ± 4 | 115 ± 1 | 50 ± 1 | 1.41 | 1.27 |
| 120 | 0.0497 | 116 ± 4 | 121 ± 1 | 49 ± 1 | 1.48 | 1.33 |
| 125 | 0.0404 | 110 ± 4 | 125 ± 1 | 50 ± 1 | 1.47 | 1.29 |
| 130 | 0.0351 | 117 ± 4 | 132 ± 2 | 49 ± 1 | 1.46 | 1.27 |
| 135 | 0.0258 | 114 ± 4 | 135 ± 1 | 51 ± 1 | 1.50 | 1.28 |
| 140 | 0.0166 | 114 ± 4 | 141 ± 1 | 50 ± 1 | 1.43 | 1.29 |
| 145 | 0.0096 | 118 ± 4 | 146 ± 1 | 50 ± 1 | 1.46 | 1.28 |
| 150 | 0.0090 | 120 ± 4 | 151 ± 1 | 50 ± 1 | 1.45 | 1.32 |