Investigating the XENON1T Low-Energy Electronic Recoil Excess Using NEST

M. Szydagis,* C. Levy,† G. M. Blockinger, A. Kamaha, N. Parveen, and G.R.C. Rischbieter
Department of Physics, University at Albany, State University of New York, Albany, NY USA
(Dated: July 6, 2020)

The search for dark matter, the missing mass of the universe, is one of the most active fields of study within particle physics. The XENON1T experiment recently observed a 3.5σ excess consistent with solar axions, a possible dark matter candidate. Here we utilize the Noble Element Simulation Technique (NEST) software to simulate the XENON1T detector, reproducing the excess. We report that, despite a detector efficiency modeling systematic, and non-ideal energy reconstruction sub-keV, the XENON1T excess cannot be explained by any simple detector parameter mismatch. However, using NEST, we can reproduce their excess in multiple unique ways, most easily via the addition of 31 ± 11 37Ar decays. Furthermore, this results in new, modified background models, reducing the significance of the excess to only 1.5-2σ. This is an independent confirmation that the XENON1T excess appears to be a real effect, but that it can be explained with more mundane, known physics.

Keywords: NEST, XENON1T, solar axion, low energy, dark matter direct detection, electronic recoils, xenon

I. INTRODUCTION

There is overwhelming evidence, via astrophysical and cosmological observations [1, 2], that the universe is made of non-luminous matter interacting rarely with baryons. The search for the aptly-named “dark matter” has been an active field for decades. Experiments have been looking for different types of it, particularly WIMPs (Weakly Interacting Massive Particles), and axions, via direct interactions through nuclear recoils (NR) and/or electronic recoils (ER), respectively. The case for the axion is more-overserved with the strong CP problem in QCD [3]. While no experiment has made an unambiguous conclusive detection of dark matter that has not already been contested and/or explained, the newest result from the XENON1T experiment [4] exhibits an excess over their background at low energies, for ER. While XENON1T was built to look predominantly for WIMPs, it is also sensitive to axions via ER, particularly solar axions, which is one possible explanation for the reported excess. In this work, we will not study potential solar axion detection. Instead, using the NEST software [5], we focus on independently confirming a real excess, but providing likely explanations for it from known physics.

Liquid xenon (LXe) detectors such as XENON1T need to be simulated with high precision, as in all rare event searches, before potentially new physics can be properly identified. The publicly available NEST simulation software is a toolkit that is widely used in the LXe community, and whose development team includes members of the LUX / LZ, XENON1T/νT, n(EXO), and DUNE experiments. NEST has served numerous noble-element-based experiments during the nine years since its inception [6], proving that it can accurately simulate and reproduce the results of various LXe (and liquid argon) detectors [7–10], by incorporating the vast amount of data available, from calibrations/backgrounds (BGs).

II. NOBLE ELEMENT SIMULATION TECHNIQUE

In a detector-agnostic way, NEST is capable of modeling average yield, i.e., numbers of quanta (photons or electrons) produced per unit energy, by various types of interactions: NR, ER, α, 83mKr, and heavy non-Xe ion recoils like 206Pb [11, 12]. It is also capable of simulating detector specifics like energy resolution, both standard deviation of monoenergetic peaks and the widths of the log(S2) and log(S2/S1) “bands” (where S1 and S2 refer to the primary and secondary scintillation signals in noble elements). NEST can thereby simulate the leakage of ER events into the NR region, and quantify the background discrimination in WIMP searches. In its simulating both the mean yields and resolution, NEST is able to model efficiencies, and so thresholds. We heavily take advantage of this capability in this work. Lastly, NEST can reproduce S1 and S2 pulse shape characteristics, but they are unnecessary here except for the S1 coincidence window.

Unfortunately, the XENON1T solar axion result does not make use of NEST. We therefore re-analyze all their data here, using NEST in an attempt to explain excess events as an artefact, from either uncertainty in detector parameters or an unexpected background. NEST average yield and/or width parameters did not need to be varied to best-fit XENON1T data, just detector-specific values. This is made clear in Fig. 1. At the very lowest energies, the light yield is going to zero, as in opposite fashion the charge yield asymptotes to its maximum possible value, with NEST uncertainty spanning the possibilities ranging from taking the inverse of the “traditional” W value of 13.7±0.2 eV [11] (73 quanta/keV) to the reciprocal of the recent measurement from EXO, 11.5±0.5 eV (that is, 87 quanta/keV) [17]. However, within the region of greatest interest for our analysis, indicated by the dashed vertical lines in Fig. 1, the default NEST yields simulation for electrons is in outstanding agreement with all the existing relevant data sets and models. Disagreement at energies orders of magnitude away from this ROI are less relevant, but also still small.

* mszydagis@albany.edu
† clevy@albany.edu
It is therefore no surprise we find NEST able to “post-dict” the XENON1T results at 81 V/cm without any free parameters. This occurred despite the fact that there is little calibration data at this low drift field (compared to past experiments operated at $O(100-1000)$ V/cm) upon which to base NEST’s low-field yields model for ER: $L_y$ (photons/keV) and $Q_y$ (e$^-$/keV). So, we were able to use Fig. 1’s central red and blue lines, without floating yields.

It is also worth noting that, despite there being a recent stable release of NEST, the beta yield model has not been officially updated in over 2 years. Recent LUX work with a $^{14}$C beta source [15] is not the default but instead a NEST option, to avoid potential over-fitting to LUX at the expense of earlier global data. NEST was never used for a $^{220}$Ru calibration before now, being driven primarily by tritium, yet it works successfully, as will be seen next.

| geometric and analysis parameters |  |
|-----------------------------------|---|
| Minimum Drift Time [µs]           | 70 [26] |
| Maximum Drift Time [µs]           | 740 [26] |
| Fiducial Radius [mm]              | 370 [4, 26] |
| Detector Radius [mm]              | 960 [27] |
| LXe-GXe Border [mm]               | 1031.5 [27] |
| Anode Level [mm]                  | 1034 [18] |
| Gate Level [mm]                   | 1029 [18] |
| Cathode Level [mm]                | 60 [18] |

TABLE I. Summary of XENON1T detector parameter values implemented for NEST in this work. Please note that the $g_1$ does not match a published value, as standard phe units include the 2-phe effect (whereby 1 VUV photon can make 2 phe in a PMT [28]). We therefore apply a lower $g_1$, in our style of detector modeling, using the unit of “phd” (detected quanta vs. standard phe). We simulated data taken under the conditions of their experiment in NEST, then compare that output to official XENON1T results.

For NEST to accurately and precisely simulate a detector, the first key input involves a proper detector parameter file. For complete transparency, Table I defines all parameters used as input to NEST that can be found publicly, except for the precise dimensions of the fiducial volume, which were set in NEST to best reproduce the fiducial mass of 1042±12 kg. The most important values NEST must have are $g_1$, $g_2$, and the drift electric field.

We further assumed a 3-fold coincidence requirement, across 212 active PMTs, applying a 50.0 ns coincidence window [18]. Based on all of these inputs, NEST will output a $g_2$ (an emergent property based on gas light collection, extraction, and other separate effects modeled from first principles [19]) of 9.85 phd/e (or, 11.57 phe/e). This can be separated into an electron extraction efficiency of 95%, derived from PIXeY / LLNL [20, 21], and an underlying SE=10.37 phd/e=12.18 phe/e. In using Poissonian statistics, we modeled a SE (1σ) width of 3.2 phe/e. The pressure and temperature reported lead to a simulated density of 2.860 g/mL and e$^-$ drift speed of 1.26 mm/µs, a velocity which does appear to make the physical coordinates of their reported detector geometry match with the max and max drift times of the fiducial volume. The density also leads to an expected $W=13.5$ eV according to NEST (which models the work function for creation of quanta as being dependent on density, including across phases) which conveniently splits the difference between the Dahl and neriX values of 13.3-13.7 eV [22, 23]. This is a very small effect, however, and an overall scaling of $O(1\%)$. It is therefore a negligible systematic.

### III. METHODS

The primary method employed here is simple: we first reproduce XENON1T’s calibration data, striving to understand their energy resolution, detector efficiency, and background model. We simulated data taken under the conditions of their experiment in NEST, then compare that output to official XENON1T results.

For NEST to accurately and precisely simulate a detector, the first key input involves a proper detector parameter file. For complete transparency, Table I defines all parameters used as input to NEST that can be found publicly, except for the precise dimensions of the fiducial volume, which were set in NEST to best reproduce the fiducial mass of 1042±12 kg. The most important values NEST must have are $g_1$, $g_2$, and the drift electric field.

![FIG. 1. NESTv2.1 $L_y$ (blue), $Q_y$ (red) for betas at 81 V/cm. Bands represent ±10%, a typical estimate of the systematic uncertainty in NEST, driven primarily by uncertainties in S1 and S2 gains in the data ($g_1$ and $g_2$) [13]. XENON100’s $^3$H-based-model is in grey, while XENON1T’s is in black using $^{220}$Ru at the closest E-fields with which we can compare, 90 and 82 V/cm respectively [14]. The circles and diamonds are the XENON1T results at 81 V/cm without any free parameters. This occurred despite the fact that there is little calibration data at this low drift field (compared to past experiments operated at $O(100-1000)$ V/cm) upon which to base NEST’s low-field yields model for ER: $L_y$ (photons/keV) and $Q_y$ (e$^-$/keV). So, we were able to use Fig. 1’s central red and blue lines, without floating yields.

XENON100's $^3$H-based-model is in grey, while XENON1T's is in black using $^{220}$Ru at the closest E-fields with which we can compare, 90 and 82 V/cm respectively [14]. The circles and diamonds are the XENON1T results at 81 V/cm without any free parameters. This occurred despite the fact that there is little calibration data at this low drift field (compared to past experiments operated at $O(100-1000)$ V/cm) upon which to base NEST’s low-field yields model for ER: $L_y$ (photons/keV) and $Q_y$ (e$^-$/keV). So, we were able to use Fig. 1’s central red and blue lines, without floating yields.

It is also worth noting that, despite there being a recent stable release of NEST, the beta yield model has not been officially updated in over 2 years. Recent LUX work with a $^{14}$C beta source [15] is not the default but instead a NEST option, to avoid potential over-fitting to LUX at the expense of earlier global data. NEST was never used for a $^{220}$Ru calibration before now, being driven primarily by tritium, yet it works successfully, as will be seen next.

![TABLE I. Summary of XENON1T detector parameter values implemented for NEST in this work. Please note that the $g_1$ does not match a published value, as standard phe units include the 2-phe effect (whereby 1 VUV photon can make 2 phe in a PMT [28]). We therefore apply a lower $g_1$, in our style of detector modeling, using the unit of “phd” (detected quanta vs. standard phe). We simulated data taken under the conditions of their experiment in NEST, then compare that output to official XENON1T results.

| primary scintillation (S1) parameters |  |
|--------------------------------------|---|
| $g_1$ [phd/photon]                   | 0.13 [24] |
| Single Photocount Resolution         | 0.4 [25] |
| Single Photoelectron Threshold [pke] | 0 (eff used instead) |
| Single Photoelectron Efficiency*     | 0.93 [26] |
| Baseline Noise                       | 0 (assumed small) |
| Double phe Emission Probability      | 0.17 [18] |

| ionization or secondary scintillation (S2) |  |
|--------------------------------------------|---|
| $g_{1gas}$ [phd/photon]                    | 0.1 [24, 26] |
| Single e$^-$ (SE) Size Fano-like Factor     | 1.0 |
| S2 Threshold [pme]                         | 500 (total, uncorr) [4] |
| Gas Extraction Region Field (kV/cm) | 10.8 (estimated) [26] |
| Electron Lifetime [µs]                     | 650 [26] |

| thermodynamics properties |  |
|---------------------------|---|
| Temperature [K]           | 177.13 [26] |
| Gas Pressure [bar]        | 1.94 (absolute) [26] |

TABLE I. Summary of XENON1T detector parameter values implemented for NEST in this work. Please note that the $g_1$ does not match a published value, as standard phe units include the 2-phe effect (whereby 1 VUV photon can make 2 phe in a PMT [28]). We therefore apply a lower $g_1$, in our style of detector modeling, using the unit of “phd” (detected quanta vs. standard phe). We simulated data taken under the conditions of their experiment in NEST, then compare that output to official XENON1T results.

For NEST to accurately and precisely simulate a detector, the first key input involves a proper detector parameter file. For complete transparency, Table I defines all parameters used as input to NEST that can be found publicly, except for the precise dimensions of the fiducial volume, which were set in NEST to best reproduce the fiducial mass of 1042±12 kg. The most important values NEST must have are $g_1$, $g_2$, and the drift electric field.

We further assumed a 3-fold coincidence requirement, across 212 active PMTs, applying a 50.0 ns coincidence window [18]. Based on all of these inputs, NEST will output a $g_2$ (an emergent property based on gas light collection, extraction, and other separate effects modeled from first principles [19]) of 9.85 phd/e (or, 11.57 phe/e). This can be separated into an electron extraction efficiency of 95%, derived from PIXeY / LLNL [20, 21], and an underlying SE=10.37 phd/e=12.18 phe/e. In using Poissonian statistics, we modeled a SE (1σ) width of 3.2 phe/e. The pressure and temperature reported lead to a simulated density of 2.860 g/mL and e$^-$ drift speed of 1.26 mm/µs, a velocity which does appear to make the physical coordinates of their reported detector geometry match with the min and max drift times of the fiducial volume. The density also leads to an expected $W=13.5$ eV according to NEST (which models the work function for creation of quanta as being dependent on density, including across phases) which conveniently splits the difference between the Dahl and neriX values of 13.3-13.7 eV [22, 23]. This is a very small effect, however, and an overall scaling of $O(1\%)$. It is therefore a negligible systematic.
A. Energy Resolution

We confirm the veracity of detector parameters and the fluctuation model, covering both correlated and anti-correlated noise, by verifying NEST’s predicted resolution for XENON1T vs. energy [31] in Fig. 2. This reveals that the “linear” noise, set by default (unrealistically) to 0.0 in NEST is closer to 0.6%, but even without this addition (which is detector-based and not NEST-yield-based) the match is still excellent on the first try with truly no free parameters, as we had claimed earlier. The difference is <1% (relative) comparing to results with and without added noise. It is modeled as an additional, uncorrelated Gaussian smearing, applied separately to the S1 and S2 channels; it is directly proportional to the pulse areas.

This accounts for imperfect position-dependent light collection, field uniformity, liquid leveling, plus similar known and unknown effects. Typical linear noise values, even given high-statistics 83mKr and/or 131mXe calibrations for efficiency and field mapping, are ~1-4%, with near-identical values for S1 and S2 (given the same DAQ being used for all pulse types) whenever NEST is used to match the past world data from different experiments [22, 32]. We do set the noise to 0.6% here, as it appears to create a better match to XENON1T, particularly at lower energies, as seen in Fig. 2. Nevertheless, we have effectively performed an unbiased side-band calibration of the noise level here, as the lowest data point within Fig. 2 is at 41.5 keV, but the solar axion result reports no energies higher than 30 keV for its analyses [4].

While beyond the scope of this paper, Fig. 2 also suggests XENON1T has, in a two-phase TPC, achieved close to the best possible resolution at the 2.6 MeV Q-value of neutrinoless double-beta decay [35], given their closeness to the NEST 0.0%-noise result. This claim is founded on NEST’s reliance on the results of [33] (shown as an inset of Fig. 2) for determining the noise floor as a function of field. It is unlikely a single-phase detector had a similar noise level to one constructed nearly 30 years later (XENON1T). Despite only Qν not combined resolution being reported, the symmetry of recombination fluctuations made it possible to use these ionization-only data to also model S1 widths. Later work enabled extrapolation from 570 keV down to lower energies: using for example XENON10 we found more calibration peaks both higher and lower in recombination probability [36].

B. NEST Reproduction of the 220Rn Calibration

To further confirm NEST simulates XENON1T well, we validate it against 220Rn data. We simulate 107 212Pb beta-decay events along with associated gamma-rays [37] as well as a flat (i.e. uniform in energy) spectrum as the calibration is itself approximatable as flat. Fig. 3, top is comparison to both: this not only demonstrates that we reproduce their 220Rn calibration, but coincidentally also explains the outliers of [4] as due to gamma/x-rays, which have different yields compared with betas at this energy scale [38]. Our hypothesis can also explain why this type of event is seen only in 220Rn or background data, but not 3H/14C calibrations in XENON100/LUX. However, these may be X/X/MSSI (Multiple-Scatter Single-Ionization) backgrounds, possibly more insidious in this higher S1 range up to 70phe, as opposed to 20-50 phe in earlier experiments [39]. This does, however, depend on exact detector geometry.

The flat ER background spectrum shows that even in this crude way, we can still reproduce the calibration. We cannot be quantitative at this stage of our analysis, appealing only to visual agreement, as the ER band is only provided as clean curves easy to interpret at 120 V/cm [27]. We attempted to use [27] to match [4], but the band was much higher due to higher S2 (lower S1) at higher field in [27]. At low field, the yields do change rapidly as a function of electric field, in an anti-correlated way [40].

C. XENON1T ER Background NEST Generator

Of equal importance to being capable of reproducing the 220Rn calibration is the ability to generate the BGs using NEST, to obtain simulated points. To that end, a custom generator was created to follow the XENON1T ER BG model, corrected for detection efficiency, below 30 keV. By not including the detector efficiency, we ensure the ER BG generator inputs the “true energy” into NEST, as an unaltered and uncorrected energy spectrum, independent of detector effects. Efficiency roll-off is thus not doubly-applied (before plus after).
FIG. 3. (Top) NEST reproducing $^{220}$Rn calibration superimposed over all observed events (black, from [4]): $^{220}$Rn with Gaussian fits (dash red), flat BG raw means (solid blue), flat BG fit (dash blue), gamma (solid green), gamma fit (dash green). Upper/lower lines indicate $1\sigma$. (Middle) NEST scatter plot overlaid on the XENON1T BG: flat model (cyan squares) and custom generator (orange diamonds). Data from [4] again in black. (Bottom) Repeating orange from the middle, but fewer events, and adding yellow X’s ($^{37}$Ar ) and pink pluses (exponential BG), the potential excess explanations.

To show our generator functions, we simulate BG with it, and compare the outputs to data along with our first simplified flat model again (Fig. 3 middle, and bottom). 1D unbinned KS tests in S1 and S2, running the generator repeatedly with different seeds on different systems and with different stats (≥409 real points) produced p-values of 0.1-0.3 with both models, without a consistent improvement when applying noise, as small pulse areas are less effected. These indicate statistical consistency.

D. XENON1T’s Energy Reconstruction

As the excess was measured in the energy space histogram not S2 vs. S1 scatter, we next explored energy reconstruction. While the combined-energy scale outperforms the older S1-only [41] or ionization-only employed e.g. by $\nu$ projects [42], it is prone to break down at low energy. We suspected, then confirmed, XENON1T used reconstructed not true energy (estimated via MC). Fig. 4 shows the output from the NEST reconstructed energy, which differs drastically from the true energy especially in the sub-keV regime. While this is not new, as shown in [23], the recent XENON1T paper does not appear to account for this. While important for other analyses, the discrepancy is insufficient to explain the excess. It is only particularly evident below 1 keV, outside of the region of interest for the excess, but can exceed a factor of 2.

FIG. 4. NEST output comparing true and reconstructed energy, using XENON1T parameters. The thickness of the line indicates statistical uncertainty. The disagreement is an emergent property stemming from many causes, including inherent skew in recombination probability, and triggering on upward fluctuations instead of true mean S1 and S2 pulse sizes, near thresholds. Inset, upper right: Data from [23] are included for qualitative comparison only, as direct agreement would only be established by modeling neriX separately within NEST.

E. Detector Efficiency

ER detection efficiency is next verified in three ways: using the true energy, NEST reconstructed energy (which matches the default XENON1T method), but additionally using the $^{220}$Rn beta spectrum. Fig. 5 demonstrates the level of disagreement among these scenarios. Below 1 keV, XENON1T is systematically overestimating efficiency (mustard line) compared to reconstructed energy (green) or true (blue). While this is not negligible, it is once more not near the ~2-4 keV ROI of the excess, and thus cannot impact the XENON1T result, though any signal model at sub-keV would be affected.
FIG. 5. The dependence of the relative efficiency on the energy. The mustard is XENON1T’s efficiency model and black is data, both from [4], the latter using the $^{220}\text{Rn}$ calibration, which we reproduce using NEST: first with a flat beta model (cyan) and then with the correct $^{220}\text{Rn}$ energy spectrum (red). Red and cyan each follow black well: this provides further evidence we can replicate XENON1T’s analyses. Green is NEST efficiency versus the reconstructed energy from an analytical fit (Gompertz) to a series of monoenergetic sims. Blue is versus true energy. Inset: Zoomed-in near the excess, with linear y. An overall (∼flat) reduction in efficiency across all energy is not portrayed, to focus upon shape. It is however important to note XENON1T’s high-energy asymptote is slightly below 1.0. We are aware this is not complete absolute efficiency.

Note the mustard line is systematically above the black points across the first four bins in a row plotted in Fig. 5 (though only at 1-2σ). What we claim to be the true efficiency in dark blue is at times lower, other times higher, than the $^{220}\text{Rn}$ points, but diverges from mustard as energy goes to 0.0. A continuous spectrum such as $^{220}\text{Rn}$ is not ideal for determining efficiency, even though this was a LUX method [43] (though not for a potential signal). A dense series of monoenergetic MC peaks is more appropriate, as is naturally done with NEST, which can be tuned and verified to match a particular detector’s data, as done for NRs in [29]. This may explain the difference between the green and mustard, the latter likely not using this technique. Contamination between energy bins occurs due to finite resolution in real data [40] that is of course changing rapidly with energy, especially as it decreases (Fig. 2). If one tries to study the efficiency as a function of reconstructed energy with MC peaks instead of estimating the true, both the mustard and black may be too high, above the green. Nevertheless, all points or lines ultimately agree on ∼1.0 at 2-4 keV.

IV. RESULTS

The primary results of our simulations are depicted in Fig. 6. The top only shows the region of interest below 10 keV but we explored up to 30 keV as shown at bottom. Black circles are always real data points as reported by [4]. We first modeled XENON1T’s ER background using NEST, assuming a flat background (cyan squares), then using our custom generator (orange diamonds again).

The difference between ‘B0’ (red) and the other curves near 1 keV in Fig. 6 is due to NEST’s lower estimate of the true efficiency, but the disagreement is far from the ROI. However, $^{37}\text{Ar}$ does fall well within 2-4 keV; it is thus our main hypothesis for explaining the excess. We took XENON1T’s ∼flat BG, and initially added 40 $^{37}\text{Ar}$ events over the full 0.65 tonne-year exposure as an educated guess based upon the visual size of the excess. We later refined this to 31±11 as a best fit to the data. $^{37}\text{Ar}$ exhibits two low-$E$ peaks: 0.27 and 2.82 keV. While the latter is the one of interest here, as it corresponds to the location of the excess in XENON1T’s main analysis, the lower-$E$ peak may permit us to distinguish between $^{37}\text{Ar}$ and other potential BGs. Our simulation corresponds with $48^{+17}_{-18}$ $^{37}\text{Ar}$ decays per tonne-year of exposure in the end. See Fig. 3 bottom to note where $^{37}\text{Ar}$ lies within S2 vs. S1 parameter space.

We also model an exponential background added to a flat ER background (pink in Fig. 6). However, it is not
motivated by a specific new BG physically. It is purely mathematical, but shows that adding either a spectrum, or a monoenergetic peak, can reproduce the excess. As the flat + exponential model fits so well, we try to motivate the excess by an underestimation in the BG model, via an overestimation of efficiency. However, we find the efficiency would have to be over 2σ off for several data points in a row, in the ROI, to justify such a drastically different BG, as shown in Fig. 7. This is less compelling.

Table II shows the χ², and σ discrepancies, between our models and the data points (black circles of Fig. 6). For completeness, and to reproduce the XENON1T numbers, we considered the 1-7 keV range, which confirmed that our BG model is similarly discrepant with the data (3.4σ) when looking at counts, showing once more how well NEST reproduces XENON1T. However, due to the size of the error bars, we find that the fits, and therefore χ² are over-constrained over this range. We thereby choose to fit to the full energy range out to 30 keV instead. This shows that our best fit to the data is using an exponential background, followed by 37Ar then tritium.

37Ar does not add events to the 1.5 and 3.5 keV bins equally, when at 2.8 keV it should be symmetric around the 2.5 keV bin. This is due to its positive skew, Fig. 8. At low energies close to threshold, event triggering occurs on high-S1 tails. In addition, skew in NEST can enter at the level of the recombination probability for the ionization electrons, derived from LUX calibrations [44]. This skew appears not only in ER bands but in fits to monoenergetic peaks. In Fig. 8 the 2.82 keV peak from 37Ar is simulated. The histogram shows a raw skew of 0.21, while a skew-Gaussian fit has a skew parameter of 1.26. This effect, while already observed for 37Ar [16, 45] is again not specific to 37Ar [44]. The effect will be more prominent, however, for monoenergetic peaks like it, than for a broad spectrum of different energies like tritium.

This effect is in fact already known to XENON1T: after finding low outliers below their ER band (shown earlier...
to be likely due to gamma and/or gamma-X events) not just high outliers above the band (as expected based on the skew observed in their calibration bands) they added a BG “mis-modeling” parameter into their WIMP search to compensate for the lower (sub-band) outliers [27].

Another important check upon the validity of the $^{37}$Ar hypothesis comes from looking at the S2-only analysis. Note that this will be units of the total S2 signal, as opposed to bottom-PMT-array, and it is uncorrected, as the lack of S1 makes 3D position correction impossible. If the excess is due to $^{37}$Ar then we expect additional excess at low S2s due to the 0.27 keV peak from the $^{37}$Ar, along with more events at high S2s due to the 2.82 keV peak. Our NEST simulation is compared to the XENON1T S2-only cross-check [24] and it is shown in Fig. 9. Within the statistics of the existing data provided by XENON1T, the S2-only analysis can neither rule out, nor rule in, the $^{37}$Ar hypothesis. It is not, however, inconsistent with it, and can thus be the means to explain the excess event counts with respect to the S2-only BG model in most bins, even if they are not individually statistically significant.

\[ \text{Excess over BG} \text{ (0.27/2.8)}\times1900=180 \text{ would reproduce the first dots). Errors on y are Poisson; on x, bin width. Lastly, while a naive scaling (0.27/2.8)$\times1900=180 \text{ would reproduce the first bin excess, Q}_{0} \text{ energy dependence does not justify it.} \]

The comparison at the lowest energy bins is less compelling, with the excess over BG occurring at lower S2 than simulated with NEST at 0.27 keV with the proper branching ratio. However, Fig. 1 hints that this could be explained within NEST’s large uncertainties on $Q_{y}$ for this extreme low-energy regime. Moreover, as this is uncorrected S2 we would need a full X-Y map and e-lifetime (vs. time) to simulate XENON1T more precisely. Lastly, few-e BGs from multiple sources, e.g. grid wire emission, may be coming into play at the first bin. Because of these enormous systematics, we do not pursue this avenue further, not considering tritium nor other earlier options.

\[ \text{FIG. 9. S2-only data from NEST (gold) simulated by adding the same amount of }^{37}\text{Ar as in the primary analysis to the XENON1T S2-only BG model (black steps). Excess over BG at } \sim 2000 \text{ phe (or PE) is clearly well explained with 2.82 keV in NEST consistent with the XENON1T data points (black dots). Errors on y are Poisson; on x, bin width. Lastly, while a naive scaling (0.27/2.8)$\times1900=180 \text{ would reproduce the first bin excess, Q}_{0} \text{ energy dependence does not justify it.} \]

\[ \text{The comparison at the lowest energy bins is less compelling, with the excess over BG occurring at lower S2 than simulated with NEST at 0.27 keV with the proper branching ratio. However, Fig. 1 hints that this could be explained within NEST’s large uncertainties on }Q_{y}\text{ for this extreme low-energy regime. Moreover, as this is uncorrected S2 we would need a full X-Y map and e-lifetime (vs. time) to simulate XENON1T more precisely. Lastly, few-e BGs from multiple sources, e.g. grid wire emission, may be coming into play at the first bin. Because of these enormous systematics, we do not pursue this avenue further, not considering tritium nor other earlier options.} \]

\[ \text{FIG. 9. S2-only data from NEST (gold) simulated by adding the same amount of }^{37}\text{Ar as in the primary analysis to the XENON1T S2-only BG model (black steps). Excess over BG at } \sim 2000 \text{ phe (or PE) is clearly well explained with 2.82 keV in NEST consistent with the XENON1T data points (black dots). Errors on y are Poisson; on x, bin width. Lastly, while a naive scaling (0.27/2.8)$\times1900=180 \text{ would reproduce the first bin excess, Q}_{0} \text{ energy dependence does not justify it.} \]

\[ \text{The comparison at the lowest energy bins is less compelling, with the excess over BG occurring at lower S2 than simulated with NEST at 0.27 keV with the proper branching ratio. However, Fig. 1 hints that this could be explained within NEST’s large uncertainties on }Q_{y}\text{ for this extreme low-energy regime. Moreover, as this is uncorrected S2 we would need a full X-Y map and e-lifetime (vs. time) to simulate XENON1T more precisely. Lastly, few-e BGs from multiple sources, e.g. grid wire emission, may be coming into play at the first bin. Because of these enormous systematics, we do not pursue this avenue further, not considering tritium nor other earlier options.} \]

\[ \text{The comparison at the lowest energy bins is less compelling, with the excess over BG occurring at lower S2 than simulated with NEST at 0.27 keV with the proper branching ratio. However, Fig. 1 hints that this could be explained within NEST’s large uncertainties on }Q_{y}\text{ for this extreme low-energy regime. Moreover, as this is uncorrected S2 we would need a full X-Y map and e-lifetime (vs. time) to simulate XENON1T more precisely. Lastly, few-e BGs from multiple sources, e.g. grid wire emission, may be coming into play at the first bin. Because of these enormous systematics, we do not pursue this avenue further, not considering tritium nor other earlier options.} \]

V. DISCUSSION

We show here that the excess seen in XENON1T data can be effectively reproduced by NEST as due to known physics. Upon incorporation of $^{37}$Ar into the XENON1T BG model, the disagreement between the model and real XENON1T data drops from 3.5 to $< 2\sigma$. This likely can be lowered even further, if we were to fully consider all uncertainties in NEST, but we conservatively do not, using only the default beta yields model, unadulterated in any way, throughout this paper. However, if one reconsiders Fig. 1, different E-field is not sufficient to completely explain an at least 1σ-level difference between $^{37}$Ar PIXeY data [16] and NEST. Furthermore, recent work by XELDA [46] indicates that there may be 5-10% differences in yields at different energies and fields, not only between gammas and betas but among many more different sub-types of ER. The PIXeY data most especially works in our favor here: if we increased the charge yield at 2.82 keV it could better explain the excess observed in S2, at low S1s, in the data scatter plot of Fig. 3 (around S1 ~ 7, S2 just below 2000 phe).

Beyond yield uncertainties, there is also uncertainty on the newly-modeled skew, using [44]. Advantage is never taken of this, using again the central NEST default values only, based on LUX and ZEPLIN, while a higher skewness, within the error bar for it within [44] could easily not only add more points at higher S2 in the first few S1 bins of Fig. 3, but also add more counts to the 3.5 and 4.5 keV bins, making the $^{37}$Ar as good if not a better fit to the XENON1T ER data, when compared again to the less well-motivated (physically) exponential hypothesis. That latter notion can itself still be motivated, based on past claims of dark matter evidence [47] that may be explicable with exponential (or similar: power-law) rising backgrounds at low energies, across different technologies. We do not speculate on any specific physics to explain it in LXe here, however.

\[ \text{In addition, sub-optimal efficiency and energy reconstruction contribute to systematics in the XENON1T analysis; however, those did not impact the excess and the overall XENON1T result. We acknowledge that we did not have access to actual XENON1T data and therefore had to digitize all their plots for comparison. This can lead to a small error, although NEST is incredibly robust in its predictions as shown in the past and we have put in a system of checks to try and minimize our errors. Therefore the authors do not believe that this would impact the stated results significantly. That said, and as mentioned before, NEST is an open-source software. We thus urge the XENON1T collaboration to reproduce our results using their data, or make their data available publicly. While the work presented here stops short of using Profile Likelihood Ratio (PLR) analysis, a PLR analysis on the NEST results will yield more precise conclusions. But, once more, this will not change the fact that, to first order, we have independently reproduced the XENON1T excess, finding it consistent with an $^{37}$Ar background.} \]
ACKNOWLEDGMENTS

This work was supported by the University at Albany SUNY under new faculty startup funding for Prof. Levy and by the DOE under Award DE-SC0015535. The authors wish to thank the LZ and LUX collaborations for useful recent discussion as well as continued support for NEST work, and recognition of its high precision plus its extreme predictive power. We also wish to thank all NEST collaboration members, especially those on XENON1T advocating for its increased use.

We do not speculate on how the $^{37}\text{Ar}$ was introduced into the detector, and we note that such an unexplained excess of $^{37}\text{Ar}$ was previously found in LUX [48].

Other possible future work could include redoing the entire analysis using the EXO-200 reported value $W = 11.5$ eV, though this would be highly non-trivial: simple re-scaling of $g_1$ and $g_2$ to account for this $W$ would disrupt NEST agreement with data on the carefully-crafted fluctuations model (Fano factor for total quanta, excitation and ionization, and non-binomial recombination fluctuations). Evidence in favor of our present assumptions ultimately lies in reproduction of XENON1T’s data.

[1] V. C. Rubin, One hundred years of rotating galaxies, Publications of the Astronomical Society of the Pacific 112, 747 (2000).
[2] Y. Akrami et al. (Planck), Planck 2018 results. I. Overview and the cosmological legacy of Planck, (2018), arXiv:1807.06205 [astro-ph.CO].
[3] R. D. Peccei and H. R. Quinn, CP conservation in the presence of pseudoparticles, Phys. Rev. Lett. 38, 1440 (1977).
[4] E. Aprile et al. (XENON), Observation of Excess Electronic Recoil Events in XENON1T, (2020), arXiv:2006.09721 [hep-ex].
[5] M. Szydagis et al., Open access noble element simulation technique, https://zenodo.org/record/3905382#XvLEAZWkEy1.
[6] M. Szydagis, N. Barry, K. Kazkaz, J. Mock, D. Stolp, M. Sweany, M. Tripathi, S. Uvarov, N. Walsh, and M. Woods, NEST: A Comprehensive Model for Scintillation Yield in Liquid Xenon, arXiv e-prints, arXiv:1106.1613 (2011), arXiv:1106.1613 [physics.ins-det].
[7] D. S. Akerib et al. (LUX Collaboration), First results from the lux dark matter experiment at the sanford underground research facility, Phys. Rev. Lett. 112, 091303 (2014).
[8] D. S. Akerib et al. (LUX-ZEPLIN Collaboration), Projected wimp sensitivity of the lux-zeplin dark matter experiment, Phys. Rev. D 101, 052002 (2020).
[9] X. Ren et al. (PandaX-II Collaboration), Constraining dark matter models with a light mediator at the pandaxii experiment, Phys. Rev. Lett. 121, 021304 (2018).
[10] E. Aprile et al. (XENON Collaboration), Low-mass dark matter search using ionization signals in xenon100, Phys. Rev. D 94, 092001 (2016).
[11] B. Lenardo, K. Kazkaz, A. Manalaysay, J. Mock, M. Szydagis, and M. Tripathi, A Global Analysis of Light and Charge Yields in Liquid Xenon, IEEE Trans. Nucl. Sci. 62, 3387 (2015), arXiv:1412.4417 [astro-ph.IM].
[12] J. Cutter, The noble element simulation technique v2, (2017).
[13] D. S. Akerib et al. (LUX Collaboration), Calibration, event reconstruction, data analysis, and limit calculation for the lux dark matter experiment, Phys. Rev. D 97, 102008 (2018).
[14] E. Aprile et al. (XENON Collaboration 5), Signal yields of kev electronic recoils and their discrimination from nuclear recoils in liquid xenon, Phys. Rev. D 97, 092007 (2018).
[15] D. Akerib et al. (LUX), Improved Measurements of the $\beta$-Decay Response of Liquid Xenon with the LUX Detector, Phys. Rev. D 100, 022002 (2019), arXiv:1903.12372 [physics.ins-det].
[16] E. Boulton et al., Calibration of a two-phase xenon time projection chamber with a $^{37}$Ar source, JINST 12 (08), P08004, arXiv:1705.08958 [physics.ins-det].
[17] G. Anton et al. (EXO-200), Measurement of the scintillation and ionization response of liquid xenon at MeV energies in the EXO-200 experiment, Phys. Rev. C 101, 065501 (2020), arXiv:1908.04128 [physics.ins-det].
[18] E. Aprile et al. (XENON), The XENON1T Dark Matter Experiment, Eur. Phys. J. C 77, 881 (2017), arXiv:1708.07051 [astro-ph.IM].
[19] D. Akerib et al. (LUX), Improved modeling of $\beta$ electronic recoils in liquid xenon using LUX calibration data, JINST 15 (02), T02007, arXiv:1910.04211 [physics.ins-det].
[20] B. Edwards et al., Extraction efficiency of drifting electrons in a two-phase xenon time projection chamber, JINST 13 (01), P01005, arXiv:1710.11032 [physics.ins-det].
[21] J. Xu, S. Pereverzev, B. Lenardo, J. Kingston, D. Naim, A. Bernstein, K. Kazkaz, and M. Tripathi, Electron extraction efficiency study for dual-phase xenon dark matter experiments, Phys. Rev. D 99, 103024 (2019), arXiv:1904.02885 [physics.ins-det].
[22] C. E. Dahl, The physics of background discrimination in liquid xenon, and first results from Xenon10 in the hunt for WIMP dark matter, Ph.D. thesis, Princeton U. (2009).
[23] L. Goetzke, E. Aprile, M. Anthony, G. Plante, and M. Weber, Measurement of light and charge yield of low-energy electronic recoils in liquid xenon, Phys. Rev. D 96, 103007 (2017), arXiv:1611.10322 [astro-ph.IM].
[24] E. Shockley, Search for new physics with electronic recoil events in xenon11, Talk (2020).
[25] A. Behrens, Light Detectors for the XENON100 and XENON1T Dark Matter Search Experiments, Ph.D. thesis, Universitaet Zurich (2014).
[26] E. Aprile et al. (XENON), Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, Phys. Rev. Lett. 121, 111302 (2018), arXiv:1805.12562 [astro-ph.CO].
[27] E. Aprile et al. (XENON), First Dark Matter Search Results from the XENON1T Experiment, Phys. Rev. Lett. 119, 181301 (2017), arXiv:1705.06655 [astro-ph.CO].
[28] C. Faham, V. Gehman, A. Currie, A. Dobi, P. Sorensen, and R. Gaitskell, Measurements of wavelength-dependent double photoelectron emission from single photons in VUV-sensitive photomultiplier tubes, Journal of Instrumentation 10 (09), P09010.
[29] D. S. Akerib et al. (LUX Collaboration), Improved limits on scattering of weakly interacting massive particles from reanalysis of 2013 lux data, Phys. Rev. Lett. 116, 161301 (2016).
[30] D. A. et al., Wimp-nucleon cross-section results from the second science run of zeplin-iii, Physics Letters B 709, 14 (2012).
[31] E. Aprile et al. (XENON), Energy resolution and linearity in the keV to MeV range measured in XENON1T, (2020), arXiv:2003.03825 [physics.ins-det].
[32] D. S. Akerib et al. (LUX Collaboration), Signal yields, energy resolution, and recombination fluctuations in liquid xenon, Phys. Rev. D 95, 012008 (2017).
[33] E. Aprile, R. Mukherjee, and M. Suzuki, Performance of a liquid xenon ionization chamber irradiated with electrons and gamma-rays, Tech. Rep. (Columbia Univ., New York, NY, 1991).
[34] E. Conti, R. DeVoe, G. Gratta, T. Koffas, D. Akimov, G. Bower, M. Breidenbach, R. Conley, M. Danilov, Z. Djurcic, A. Dolgolenko, C. Hall, A. Odian, A. Piepke, C. Y. Prescott, P. C. Rowson, K. Skarpaas, J.-L. Vuilleumier, K. Wamba, and O. Zeldovich, Correlated fluctuations between luminescence and ionization in liquid xenon, Phys. Rev. B 68, 054201 (2003).
[35] M. Redshaw, E. Wingfield, J. McDaniel, and E. Myers, Mass and double-beta-decay Q value of Xe-136, Phys. Rev. Lett. 98, 053003 (2007).
[36] E. Aprile, J. Angle, F. Armeodo, L. Baudis, A. Bernstein, A. Bolozdynya, P. Brusov, L. Coelho, C. Dahl, L. DeViveiros, and et al., Design and performance of the xenon10 dark matter experiment, Astroparticle Physics 34, 679698 (2011).
[37] R. F. Lang, A. Brown, E. Brown, M. Cervantes, S. Macmullin, D. Masson, J. Schreiner, and H. Simgen, A 226Rn source for the calibration of low-background experiments, JINST 11 (04), P04004, arXiv:1602.01138 [physics.ins-det].
[38] M. Szydagis, A. Fyhrie, D. Thorngren, and M. Tripathi, Enhancement of NEST Capabilities for Simulating Low-Energy Recoils in Liquid Xenon, JINST 8, C10003, arXiv:1307.6601 [physics.ins-det].
[39] G. Rischbieter, Background modeling in the lux detector for an effective field theory dark matter search, (2020).
[40] J. Balajthy, Purity Monitoring Techniques and Electronic Energy Deposition Properties in Liquid Xenon Time Projection Chambers, Ph.D. thesis.
[41] E. Aprile et al. (XENON100 Collaboration), First axion results from the xenon100 experiment, Phys. Rev. D 90, 062009 (2014).
[42] C. Adams et al. (MicroBooNE), Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons, JINST 11 (03), P03022, arXiv:1907.11736 [physics.ins-det].
[43] D. Akerib et al. (LUX), Tritium calibration of the LUX dark matter experiment, Phys. Rev. D 93, 072009 (2016), arXiv:1512.03133 [physics.ins-det].
[44] D. Akerib et al. (LUX), Discrimination of electronic recoils from nuclear recoils in two-phase xenon time projection chambers, (2020), arXiv:2004.06304 [physics.ins-det].
[45] E. Boulton, Applications of Two-Phase Xenon Time Projection Chambers: Searching for Dark Matter and Special Nuclear Materials, Ph.D. thesis.
[46] D. Temples, Understanding neutrino background implications in lxe-tpc dark matter searches using 127xe electron captures (2019).
[47] C. Aalseth et al. (CoGeNT), Results from a Search for Light-Mass Dark Matter with a P-type Point Contact Germanium Detector, Phys. Rev. Lett. 106, 131301 (2011), arXiv:1002.4703 [astro-ph.CO].
[48] D. Akerib et al. (LUX), Search for annual and diurnal rate modulations in the LUX experiment, Phys. Rev. D 98, 062005 (2018), arXiv:1807.07113 [astro-ph.CO].