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 12222-0100, New York, USA

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 potentially consistent with dark matter, or with solar axions. Here, we will use the Noble Element Simulation Technique (NEST) software to simulate the XENON1T detector, reproducing the excess. We utilize different detector efficiency and energy reconstruction models, but they primarily impact sub-keV energies and cannot explain the XENON1T excess. However, using NEST, we can reproduce their excess in multiple, unique ways, most easily via the addition of 31 ± 11 ^{37}Ar decays. Furthermore, this results in new, modified background models, reducing the significance of the excess to $\leq 2.2\sigma$ at least using non-Profile Likelihood Ratio (PLR) methods. This is independent confirmation that the excess is a real effect, but potentially explicable by known physics. Many cross-checks of our ^{37}Ar hypothesis are presented.

I. INTRODUCTION

There is overwhelming evidence, via astrophysical and cosmological observations [1,2], that the Universe is made of nonluminous 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, particularly weakly interacting massive particles (WIMPs) via direct nuclear recoils (NRs) and/ or electronic recoils (ERs). While no experiment has made an unambiguous conclusive detection of dark matter or of axions [3] that has not already been contested and/or explained, the newest results from the XENON1T experiment [4] do exhibit an excess over their background for low-energy ER. While XENON1T was built to look predominantly for WIMPs, it is sensitive to the axion via ER, particularly solar axions, one potential explanation for the reported excess. For this work, we will not study potential solar axion detection, nor a neutrino magnetic moment or bosonic WIMPs. Instead, using the Noble Element Simulation Technique (NEST) software [5], we focus on independently confirming a real excess, then seek alternate explanations.

---
*mszydagis@albany.edu
†clevy@albany.edu

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. While XENON1T has its own Monte Carlo (MC) framework [6], whose advantage is in simulating features unique to the detector, 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/nT, n(EXO), and DUNE experiments. NEST has served numerous noble-element-based experiments during the nine years since its inception [7], proving that it can accurately simulate and reproduce the results of various LXe (and liquid argon) detectors [8–11], by incorporating the immense amount of data available from calibrations and 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, ^{83m}Kr, and heavy non-Xe ion recoils like ^{206}Pb [12,13]. 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
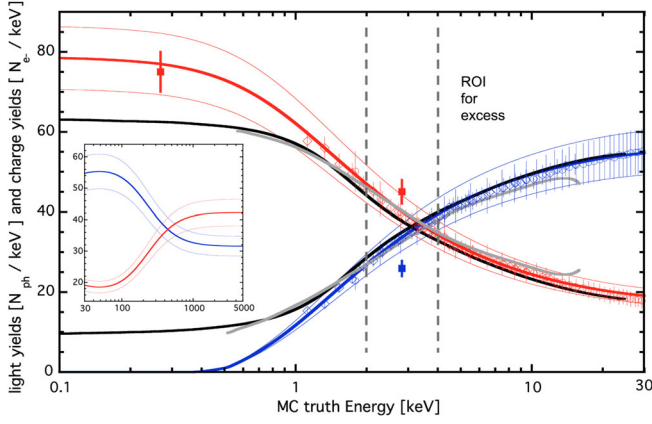
FIG. 1. NESTv2.1 L_y (blue), Q_y (red) for betas at 81 V/cm. Bands represent ±10%, a typical estimate of the systematic uncertainty in NEST, driven primarily by uncertainties in S1 and S2 gains in the data (g_1 and g_2) [15]. XENON100's ³H-based-model is in gray, with XENON1T's [6] in black using ²²⁰Rn at the closest E-fields with which we can compare, 90 and 82 V/cm, respectively [16]. The circles and diamonds are 80 V/cm ¹⁴C and ³H LUX data sets, respectively [17], while squares are ³⁷Ar data from PIXeY [18] at ~100 V/cm. One reason for the slight discrepancy is the L_y increasing (Q_y anticorrelated) with lower drift field. Inset: yields out to 5 MeV.

of this capability in this work. Lastly, NEST can reproduce S1 and S2 pulse shape characteristics, but they are unneeded here except for the S1 coincidence window.

We reanalyze [4] here, utilizing NEST to try and explain excess events as being, e.g., from an unexpected BG. NEST average yield and width parameters did not need to be varied to fit to XENON1T data, as they are detector-independent. Only the detector-specific values were changed to match XENON1T. This is made clear in Fig. 1. At sub-keV energies, light yield goes to 0, as, in opposite fashion, charge 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 [12] (73 quanta/keV) to the reciprocal of the recent measurement from EXO, 11.5 ± 0.5 eV (i.e., 87 quanta/keV) [14]. However, in the region of greatest interest for our analysis, indicated by vertical dashes in Fig. 1, the default NEST yields MC 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 region of interest (ROI) is less relevant, but also still small (Fig. 1 inset).

It is therefore no surprise we find NEST able to "postdict" the XENON1T results at 81 V/cm without any free parameters. This occurred despite the fact that there is less 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 the central red and blue lines of Fig. 1, without floating yields.

It is also worth noting that, despite there being a recent new stable release of NEST, the beta yield model has not been officially updated in over two years. Recent LUX work with a ¹⁴C beta source [17] is not the default but instead a NEST option, to avoid potential overfitting to LUX at the expense of earlier global data. The default NEST yields applied in this work were fit to LUX tritium data but not to the LUX ¹⁴C data. NEST was never used for a ²²⁰Rn calibration before now, being driven primarily by tritium, yet it works successfully, as will be seen next.

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, and then compared 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 threefold coincidence requirement, across 212 active PMTs (Photomultiplier Tubes), applying a 50.0 ns coincidence window [24]. 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 [27]) 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 [28,29], 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 13.4–13.7 eV [30,31]. This is a very small effect, however, and an overall scaling of $O(1\%)$. It is therefore a negligible systematic.

A. Energy resolution

We confirmed the veracity of detector parameters and the fluctuation model, covering both correlated and anticorrelated noise, by verifying NEST's predicted resolution for XENON1T as a function of energy [32] in

TABLE I. Summary of XENON1T detector parameter values implemented for NEST in this work. Please note the g_1 does not match a published value, as standard phe units include the 2-phe effect (whereby one VUV photon can make 2 phe within a PMT [19]). We therefore quote a different g_1, in our style of detector modeling, using the unit of "phd" (detected photons) developed by LUX [20], with the 2-phe effect separately simulated, probabilistically (not a constant offset) as done also by XENON1T. Lastly, in NEST $z = 0$ (the vertical axis) is at bottom, requiring a translation from XENON1T's definition, of $z = 0$ at the top (gate grid wires).

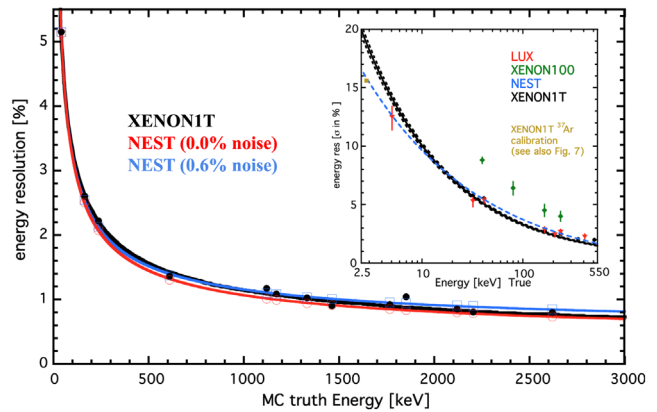|	
Primary scintillation (S1) parameters	
g_1 (phd/photon)	0.13 [21]
Single photoelectron resolution	0.4 [22]
Single photoelectron threshold (phe)	0 (*eff used)
Single photoelectron efficiency*	0.93 [23]
Baseline noise	0 (assumed small)
Double phe emission probability	0.2 ± 0.05 [24,25]
Ionization or secondary scintillation (S2)	
g_1^{gas} (phd/photon)	0.1 [21,23]
Single e^- (SE) size Fano-like factor	1.0
S2 threshold (phe) top + bottom	500 (uncorr) [4]
Gas extraction region field (kV/cm)	10.8 (est.) [23]
Electron lifetime (μs)	650 [23]
Thermodynamics properties	
Temperature (K)	177.15 [23]
Gas pressure (bar)	1.94 (abs) [23]
Geometric and analysis parameters	
Minimum drift time (μs)	70 [23]
Maximum drift time (μs)	740 [23]
Fiducial radius (mm)	370 [4,23]
Detector radius (mm)	960 [26]
LXe-GXe border (mm)	1031.5 [26]
Anode level (mm)	1034 [24]
Gate level (mm)	1029 [24]
Cathode level (mm)	60 [24]

FIG. 2. Energy resolution versus energy, comparing black dots, real data from XENON1T [32], to NEST with 0% noise (hollow red circles) and with 0.6% noise (cyan squares). Lines are analytic fits (power laws plus constants, with powers consistent with the theoretical 0.5). Black line is XENON1T model. Inset: the resolution at lower energy (down to 2.5 keV) with XENON1T's empirical function extrapolated from higher energies in black [33]. NEST with 0.006 noise is the cyan dash, extending the same simulations from the primary figure. Once g_1, g_2, and E-field are established, they drive the resolution in this energy range, from first principles. Data sets from other experiments are displayed as points, in other colors, but are not expected to match as resolution is unique per experiment. The yellow square in the inset will be addressed later.

Fig. 2. This reveals that the "linear" noise, set by default (unrealistically) to 0.0 in NEST is closer to 0.6%. Even without the addition of noise, the energy resolution predicted by NEST (without free parameters) is in good agreement with the XENON1T data; this implies that XENON1T achieved extremely low levels of noise and other effects, not captured within NEST by default. The difference is $<1\%$ (relative) comparing to results with/ without noise. It is modeled as additional, uncorrelated Gaussian smearing and applied separately to the S1 and S2 pulses; it is directly proportional to each of the pulse areas in phe.

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 ^{83m}Kr and/or ^{131m}Xe calibrations for efficiency and field mapping, are $\sim 1\% - 4\%$, 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 [30,34]. We do set the noise to 0.6% here, as it appears to create a better match to XENON1T, particularly at lower energies, as shown 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 signal model does not extend beyond 30 keV [4].

B. NEST reproduction of the ^{220}Rn calibration

To further confirm NEST simulates XENON1T well, we validate it against ^{220}Rn data. We simulate 10^7 ^{212}Pb beta decays that dominate [35] as well as a flat (i.e., uniform in energy) spectrum, as the ^{212}Pb is close to flat. Figure 3 top compares with both. This demonstrates we reproduce ^{220}Rn while Fig. 3 second from top potentially explains outliers in [4] as due to gamma/x-rays, as they have different yields compared with betas at this energy scale [36]. Our hypothesis can also explain why this type of event is seen in BG data, but not ^{3}H/^{14}C calibrations in XENON100/ LUX. However, these may be gamma-X/MSSI (multiple-scatter single-ionization) BGs, possibly more insidious in this higher S1 range up to 70 phe, as opposed to 20–50 phe in earlier experiments [37]. Detector geometry plays a strong role in gamma-X.

The flat ER BG spectrum shows that even in this crude way we still reproduce XENON1T well. To be quantitative, we compare not the flat MC but Rn MC with data. $(NEST - data)/data$ (NEST is red in Fig. 3 top) has a median offset from the data (black line [38]) of -1.0% for band mean, with (nonsystematic) max/min deviation of
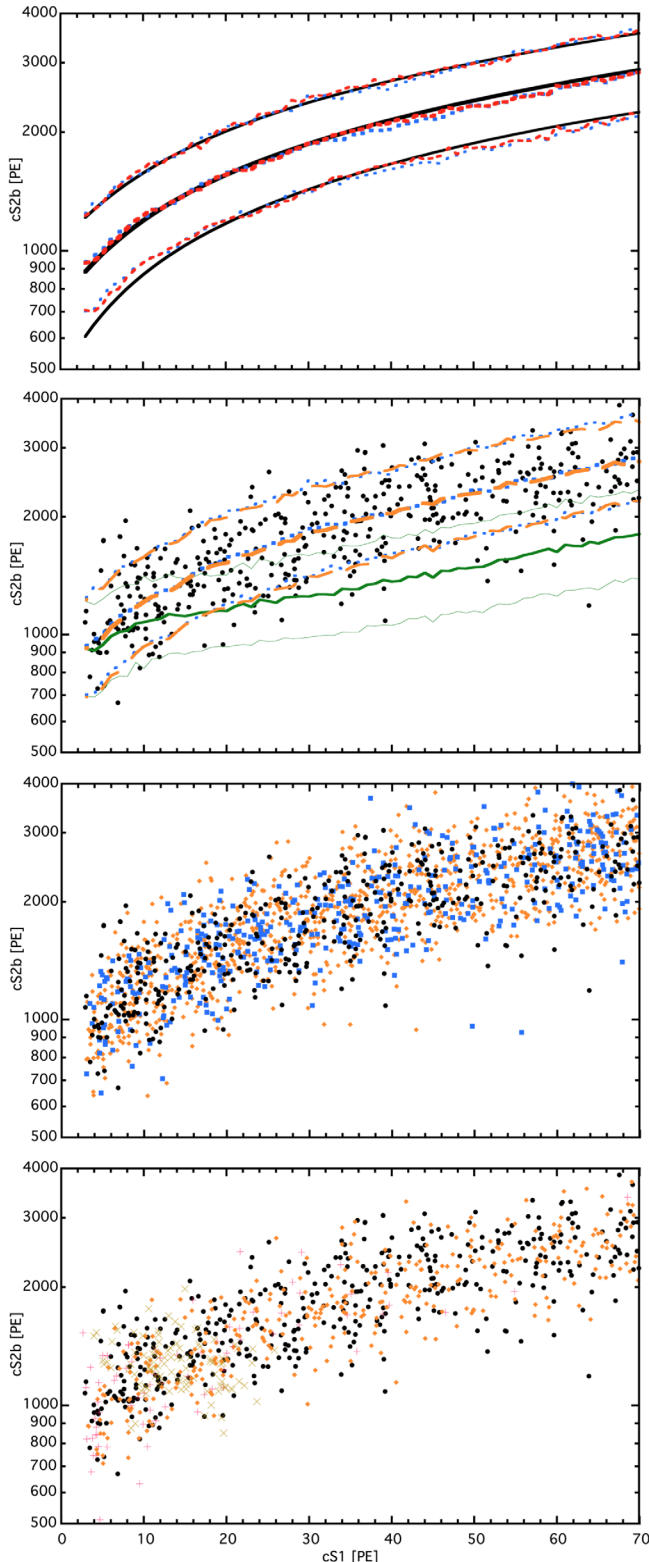
$\pm 5\%$. For band width, the median offset is $+1.3\%$ with max/min $\pm 12\%$. This is quite comparable to what can be achieved with NEST with direct access to data [27].

C. XENON1T ER background NEST generator

Of equal importance to reproduction of the ^{220}Rn calibration is BG generation, for obtaining simulated points: orange, in lower half of Fig. 3, contrasted with flat in cyan and data in black. A custom generator was created to follow the XENON1T ER BG model, corrected for detection efficiency, below 30 keV, allowing for a significant buffer beyond the excess ROI. By not including detector efficiency initially, we ensure the generator inputs the "true energies" into NEST, as an unadulterated, uncorrected energy spectrum, independent of detector effects.

NEST's PDF is the sum of all XENON1T BGs in Table 1 of [4], which includes the mean rate for each isotope fit by XENON1T, and the energy spectrum shapes assumed. Our check of the excess remains independent, as the use of NEST instead of XENON1T MC leads to some variations in energy resolutions, as seen in Fig. 2 and next in Fig. 4 where the 163.9 keV peak (^{131m}Xe) differs slightly. The sum of all BGs is indistinguishable statistically (Kolmogorov-Smirnov (KS) test, quoted below) from flat due to fluctuations, prior to addition of ^{37}Ar. See Fig. 3 for scatter plot or Fig. 5 for binning in energy.

In Fig. 4, we explicitly show what the XENON1T BG looks like before efficiency. It is quite flat for $\sim$4–30 keV, but has a slight positive slope, after all radioisotopes are combined that contribute, which we do not neglect. Computing this was a necessary initial step. This is not in [4] but was derived by combining all isotopic contributions and dividing by the efficiency. The resultant shape better motivates qualitatively our investigation later into a BG that rises as energy goes to zero.

To show our generator functions, we simulate BG with it, and compare the outputs to data along with our first simplified flat model once again. One-dimensional unbinned KS tests in both S1s and S2s, running the generator repeatedly with different seeds on different

FIG. 3. Top: NEST in dashed red reproducing high-statistic ^{220}Rn calibration superimposed in solid black [38], with skew-Gaussian fits, compared to flat BG (dashed cyan). Points for black not provided but can be seen in [38]. Contours indicate 10–50–90%. Second: as both the BG and ^{220}Rn are ~flat that sim is repeated again in cyan, but now our custom generator appears (orange) and a generic flat gamma band (solid green). Science data from [4] as black dots. Third: NEST scatter plot overlaid on the XENON1T BG: a flat model (cyan squares) and custom generator (orange diamonds). XENON1T search data as black points again. Bottom: repeating orange from last plot, but fewer events, and adding yellow X's (^{37}Ar) and pink pluses (exponential BG), potential excess explanations.
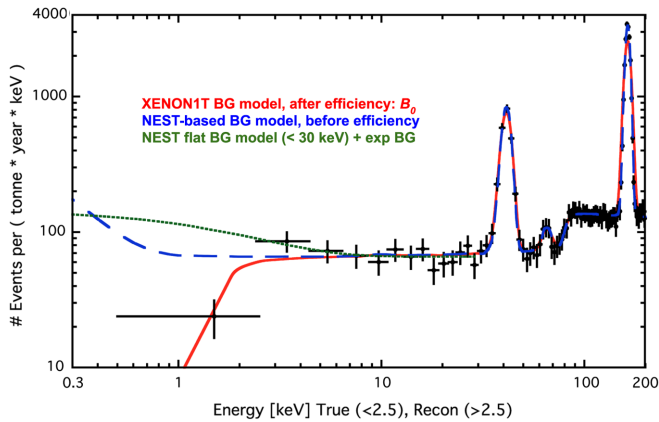
FIG. 4. Smooth depiction of BGs assumed by NEST in blue (long dash), compared to B0 in red (solid), and observed data in black (points with errors) [4]. The exponential is in green (dotted) which best fits the excess (no ^{37}Ar necessary). Given the uncertainties, including systematics in B0 and in efficiency, not depicted, it might be possible to reconcile the blue with the green. Figure 6 implies that is not necessarily the case, but a combination may be possible where the number of necessary ^{37}Ar events to explain the data is dropped, reducing the steepness required in the green to explain the excess. (Note green becomes pink and blue becomes orange in other plots.)
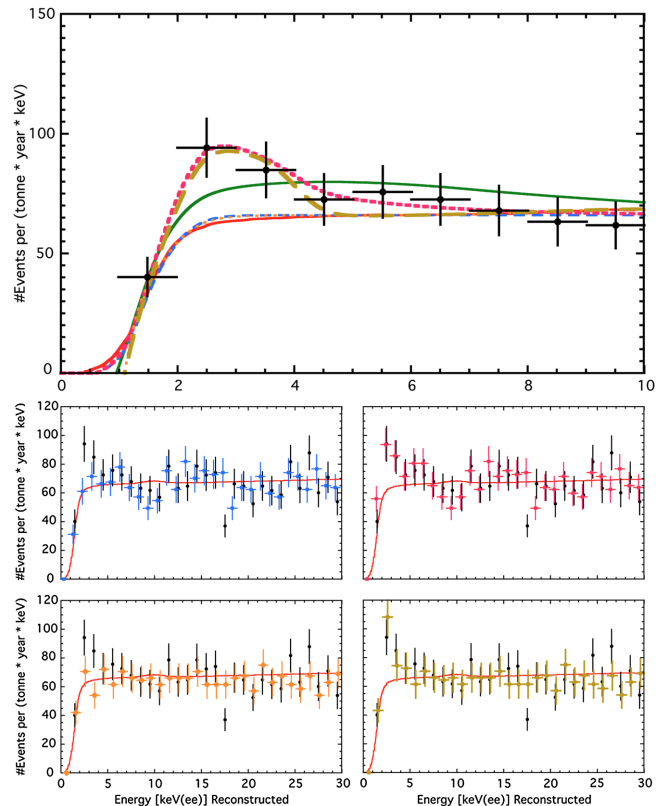

FIG. 5. A summary of every model studied with NEST: data and background B0 model from [4] are black dots and solid red line, respectively. Top: our flat ER BG (cyan), the same flat ER BG with a low-energy exponential added (pink), the NEST custom generator for mimicking B0 (orange), the same custom generator with ^{37}Ar (yellow), then with tritium added (thick solid green line). Bottom: the discrete NEST outputs in the same colors as at top, but after realistic full, detector MC. For clarity, every point has been offset from its actual value by $O(0.1)$ keV and ^{3}H is omitted. The flat BG is 66 per tonne-year-keV. Random seed used is identical per each row.

systems and with different events counts (both greater than and equal to the 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 affected by it.

These p-values are not indicative of any significant degree of statistical inconsistency. The reason they are not uniformly distributed up to 1.0 is likely the divergence at the lower half of the band for the lowest S1s, most easily observed in Fig. 3 (top). This issue is not challenging to understand, but difficult to model without access to all information on XENON1T. It is likely due to a combination of wall and accidental-coincidence BGs that are not included in NEST. This feature can be observed in both [4] and Fig. 16 of [38]. Looking at the low-S1 upper half of the band, it is clear this is a problem with the symmetry of the band, but not with band width overall. Alphas on the wall or from ^{220}Rn itself degraded in energy, as well as heavy recoiling atoms/ions such as Pb from the Rn chain and naked betas, may experience charge loss, lowering the S2. As this is a problem only at lower S2s, and appears below the average S1 for ^{37}Ar, our results stand in spite of this, but this is a problem even in the fiducial volume far from the walls: degraded position resolution at walls can cause a low S2 to be reconstructed inside of the fiducial volume. For higher S2s, apparent leakage of events above the upper Gaussian contour can be ascribed to multiple-scatter ER identified as single-scatter and/or Xe-intrinsic upward skew, as detailed later for peaks.

D. Energy reconstruction and efficiency

The excess was measured in binned energy space not S2 versus S1 scatter, so that defined the next investigation. XENON1T reports reconstructed energies, but the nonlinear deconvolution into true energy was estimated via MC [6,38] for their PLR. There may be differences from NEST, but primarily at sub-keV, thus irrelevant.

The efficiency was verified many ways, but again NEST agrees with what was reported except sub-keV, not relevant here, and also within large errors (in the Appendix).

IV. RESULTS

The NEST-simulated energy histograms are depicted in Fig. 5. 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," the XENON1T BG model after efficiency application in red, and the other curves near 1 keV in Fig. 5 is due to NEST's lower intrinsic efficiency, as predicted based on g_1, g_2, and field, but this (insignificant) disagreement is far from the ROI. However, ^{37}Ar does fall well within the ROI and, based also on LUX experiences [39,40], is our primary attempt to explain the excess. We at first added 50 ^{37}Ar events over the full 0.65 tonne-year exposure, estimated from the raw size of the excess, later refined to 31 ± 11 counts as best fit. ^{37}Ar exhibits two low-E peaks: 0.27 and 2.82 keV. While the latter is the one of interest here, as it may lie near the location of the excess in XENON1T's main analysis, the lower-E peak may permit us to distinguish between ^{37}Ar and other potential BGs. Our MC simulation corresponds to 48^{+17}_{-18} ^{37}Ar decays per tonne-year of exposure. Figure 3 bottom shows them in S2 versus S1.

We also model an exponential background added to a flat ER background (pink in Fig. 5). 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. 6. This is less compelling.

Lastly, we model tritium (^{3}H), but also find it to be less compelling. It is not only a worse fit than ^{37}Ar and the exponential (if you account for shape using χ^2, and do not just look at Poisson statistics), it is lower than the other hypotheses in the 2.5 keV bin, farther from the data. It also raises the counts in the lowest energy bin due to this being a
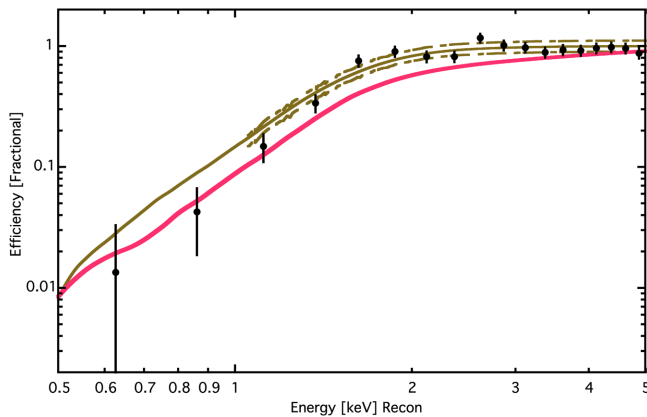
continuous source, unlike ^{37}Ar which is monoenergetic. The exponential hypothesis suffers less from this raising of counts for the 1.5 keV bin considerably above the data, as, counterintuitively, exponentially more counts at low energies implies more counts at true energies which are unable to fluctuate up effectively, in *reconstructed* energy space. We fully recognize these statements could be strengthened with a PLR, but without access to all data in all dimensions including position this is unrealistic for non-XENON1T members.

Table II has the χ^2's and the σ discrepancies between our models and the data points (black dots from Fig. 5). For completeness, and to reproduce the XENON1T numbers, we considered the 1–7 keV range. However, due to the size of the error bars, we find that the fits, and thus χ^2's, are overconstrained over this range. Therefore, we choose to fit to a larger energy range (1–30 keV, as per Fig. 4 of [4]). This shows that our best fit to the data is using an exponential BG, followed by ^{37}Ar then tritium.

^{37}Ar does not span 1.5–3.5 keV bins equally, when at 2.8 keV it should be ~symmetric about 2.5 keV. This is due to positive skew (Fig. 7). At near-threshold energies, event triggering occurs on high-S1 tails. Moreover, skew in NEST enters at the level of recombination probability for S2 electrons, derived from LUX calibrations [41]. It appears not only in ER bands but monoenergetic peaks. Figure 7 shows the ^{37}Ar 2.82 keV peak. A fit of the NEST histogram to a skew-normal distribution has skewness α (described later) of 1.3 ± 0.2, compared with 1.5 ± 0.2 for preliminary XENON1T calibration data. The skewness effect, already observed for ^{37}Ar [18,42], is again not specific to it [41]; the effect will be more prominent for

FIG. 6.   XENON1T efficiencies from Figs. 2 and 6 of [4] copied here in mustard (with error) and black (solid circles), where ^{220}Rn was used to calculate the latter. Pink shows what efficiency would be to justify an exponentially falling BG.

TABLE II.   The goodness of fit quantifying the level of agreement with data, for two broad energy ranges (encompassing the 2–4 keV range where the excess seems largest). The number of free parameters assumed for B0 and the PDF, our custom B0-like generator, is four, representing the four largest low-energy BGs (^{214}Pb, ^{85}Kr, solar ν's, materials) and three for the exponential (amplitude, shape, offset). With ^{3}H and ^{37}Ar, one additional parameter was varied for the PDF, the number of decays. Flat had only one free parameter. Using naïve counting, in a tighter energy window, all excess hypotheses do well, as denoted by σ_p (p = Poisson). σ_χ is derived from the χ^2/d.o.f (degrees of freedom).

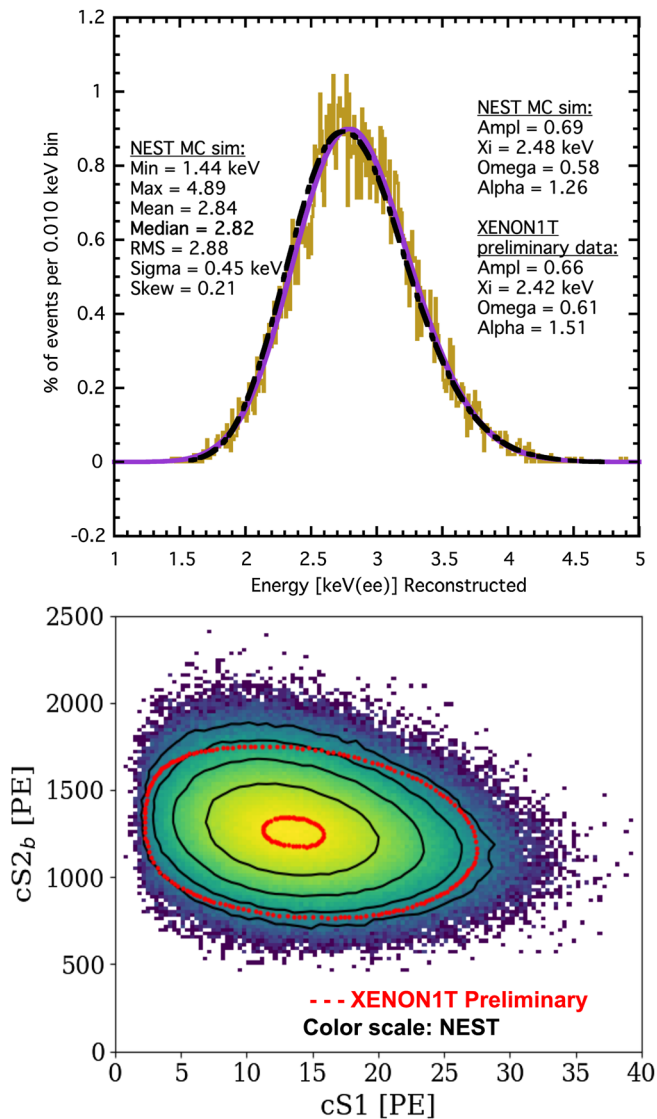| 1–30 keV (σ_χ) and 1–7 keV (σ_p)					
Hypothesis (color)	χ^2	d.o.f	χ^2/d.o.f	σ_χ	σ_p
Flat BG (cyan)	41	29-1	1.46	1.92	2.65
B_0 (red)	48	29-4	1.92	2.91	3.35
PDF (orange)	47	29-4	1.88	2.80	2.70
PDF + ^{37}Ar (yellow)	38	29-5	1.60	2.16	0.41
Flat + exponential (pink)	33	29-3	1.26	1.38	−0.54
PDF + ^{3}H	45	29-5	1.88	2.80	−0.28

FIG. 7. Top: ^{37}Ar peak. NEST utilizing XENON1T detector parameters in gold, best fit (skew-normal) in magenta. Black dash is a preliminary XENON1T calibration [21] showing again remarkable agreement with (default) NEST. Numbers at left are raw histogram statistics; at right best-fit parameters, defined on next page, for both NEST and data, with errors ~0.1 in each due to high statistics. Bottom: S2 versus S1 for 10^6 ^{37}Ar events. The color scale and black contours are both NEST's; in red are inner/outermost (arbitrary) contours of slide 68 of [21]. As raw data were marked as preliminary, not provided by XENON1T, only a qualitative comparison can be performed.

monoenergetic peaks than for a broad spectrum of different energies like tritium, due to smearing.

In Fig. 7 bottom, we use NEST to further study actual ^{37}Ar, which was a XENON1T calibration, not just potential BG or excess hypothesis, affording us an opportunity of a deeper independent study. For this plot, we separate combined energy into the S1 and S2 areas. The non-Gaussian, triangular shape qualitatively agrees with data. This should make the probability of a NEST mismodeling

of ^{37}Ar in XENON1T impacting our result *de minimis*. To allow additional, quantitative comparison, in combined energy space, we quantify our work in Fig. 7 top.

A similar asymmetry was in fact already reported by XENON1T: after discovering low outliers, lying below their ER band (Sec. III B: possibly γ's and/or γ-X), not just high outliers above the band (as expected based upon the skew observed in their calibration bands), they added a BG "mismodeling" parameter into their WIMP search to compensate for any lower (i.e., subband) outliers [26,43]. They did ultimately determine though that fewer WIMP-signal-like (NR-like) ER tail events in science data compared to calibration were a better fit [6,23] and also provided an explanation for remaining outliers as being driven primarily by surface BGs, which experience charge loss, lowering their S2, similar to what was found on LUX [44], and mentioned earlier. We presented here a novel explanation that can perhaps account for a fraction of the outliers in Fig. 3 (XENON1T's Fig. 5). Further evidence in favor of gammas is in the Appendix. They are not likely to explain all outliers as the NR band would be too contaminated for a WIMP search then.

Another important check upon the validity of the ^{37}Ar hypothesis comes from looking at the S2-only analysis. Note that this will be in 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 [21] and it is shown in Fig. 8. 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.

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 this could be explained within NEST's large uncertainties on Q_y for this extreme low-energy regime. Furthermore, as this is uncorrected S2, we would need a full XY map and e^--lifetime (versus time) to simulate XENON1T more precisely. Lastly, few-electron BGs from multiple sources, e.g., grid wire emissions [11], may be coming into play for the first bin. Because of these enormous systematics, we do not pursue the S2-only avenue further, not considering, e.g., tritium.

A drawback to the ^{37}Ar hypothesis is the best fit to a peak for a bosonic dark matter search being 2.3 keV: in XENON1T's Fig. 11 ([4] v2) 2.8 keV is strongly disfavored. We now reconcile our hypothesis with this analysis.
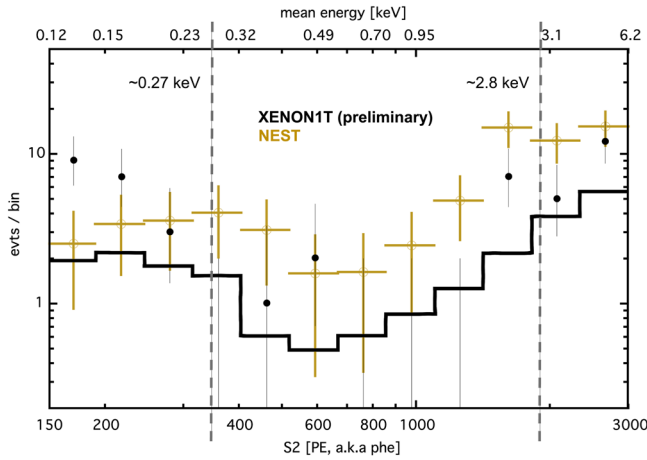
FIG. 8.   S2-only data from NEST (gold) simulated by adding the same amount of ^{37}Ar as in the primary analysis to the preliminary XENON1T S2-only BG model [21] (black steps). Excess over BG at ~2000 phe (or photoelectrons (PE)) is consistent with 2.82 keV in NEST, consistent with the preliminary XENON1T data points [21] (black dots). Errors on y are Poisson; on x, bin width. Lastly, while a naïve scaling $(0.27/2.8) * 1900 = 180$ would reproduce the first bin excess, the energy dependence of Q_y does not justify that.

XENON1T states more than once that the functional form used in [4] was Gaussian, so their peak search does not account for the inherent asymmetry due to skew at keV-scale energies (Fig. 7 again) demonstrated by their own calibration, which they do not include in their analysis. The formulas for a skew-normal fit are as follows:

(1) $y = A e^{\frac{-(x-\xi)^2}{2\omega^2}} \left[1 + \text{erf}\left(\alpha \frac{x-\xi}{\omega\sqrt{2}} \right) \right]$

(2) $\mu = \xi + \omega\delta\sqrt{\frac{2}{\pi}}$

(3) $\delta = \alpha/\sqrt{1 + \alpha^2}$

Here A is amplitude, μ mean, ω related to σ (i.e., a measure of the width), and α is related to the amount of skew. ξ can be either lower or higher than the mean, peak, or median, for positive and negative skew, respectively. We scan over both normal Gaussian fits (in gold) and skew versions (in purple) in the data in our Fig. 5 (XENON1T's Fig. 4), conducting a monoenergetic peak search. The results are depicted in Fig. 9. For the Gaussian case, we reproduce XENON1T's 2.3 keV value, with a similar error bar, further evidence they fit to a Gaussian, but in the skew case we find a higher best-fit mean, 2.5 keV, within a greater error range, spanning 2.3 and 2.8 keV. Contrasting the two methods, one can see that proper accounting of skew can easily shift 2.3 to 2.8 keV. Our ^{37}Ar hypothesis should therefore still be seriously considered. A PLR would likely have a more constraining uncertainty. Lastly, many phenomenological papers reinterpreting the excess [45–49] infer a 2.8 keV peak in independent analyses completely unrelated to NEST, or skew. This is additional evidence 2.8 is not unreasonable.
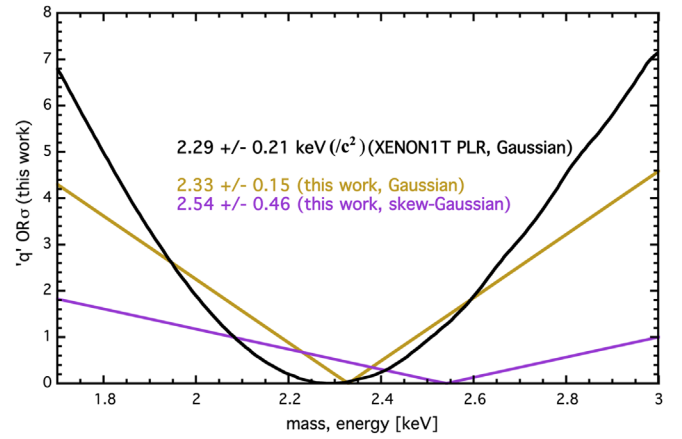

FIG. 9.   XENON1T's quoted log-likelihood ratio for different bosonic WIMP rest-mass energies from [4] in black. The best fit was 2.3 keV. In gold, for comparison is the number of σ of disagreement from χ^2-based, not PLR, fits to the data, with NEST. The nature of the different statistical test causes the nonsmooth V shape, and lower significances of discrepancy in the "wings," as expected for this type of method. Despite these differences, we find a near-identical best-fit energy as XENON1T, with 2.82 keV discrepant by a similar amount: $>3\sigma$ at least. In purple, the fit function is changed to a skew Gaussian, lowering the disagreement to 0.6σ and bringing the best fit closer (higher E). This too is natural, as the simulation showed positive skew, quantitatively confirmed with data, and a greater number of free parameters introduces new correlations.

Recognizing [4] states ^{37}Ar is unlikely, we sought additional validation beyond mean energy and S2 only. First, we refer back to Fig. 2 inset, which shows NEST's width (15.79% at 2.8 keV) in cyan, better matching the digitized XENON1T ^{37}Ar calibration data width (15.88%) in yellow, compared with the XENON1T model (18.88% at 2.8 keV) in black, implying a possible discrepancy in energy resolution. [46], an independent reanalysis of the XENON1T data, agrees with the lower resolution predicted by NEST. The XENON1T analysis [21] states that the ^{37}Ar calibration data show a resolution of 18.12%; however, both NEST and our digitization of the real data agree upon 16% rounded across three methods (Gauss, skew, raw σ). Our only explanation is a fit to only the right half of the peak yields 18.1%, but this just underscores again our point that skew or asymmetry cannot be ignored.

Next, we considered time dependence in actual data in Fig. 10. While errors are large and XENON1T's PLR has already established the points are consistent with none, we find in Fig. 10 statistical consistency with the ^{37}Ar lifetime. While our hypothesis tests are "goodness of fit" not likelihood ratios, multiple tests all concur, despite distillation and gettering removing Ar in principle [4]. The unlikely possibility exists that, e.g., a small leak, outgassing, or activation introduces minute quantities of it, or it is introduced by other means as-yet not understood. This could address why the excess was
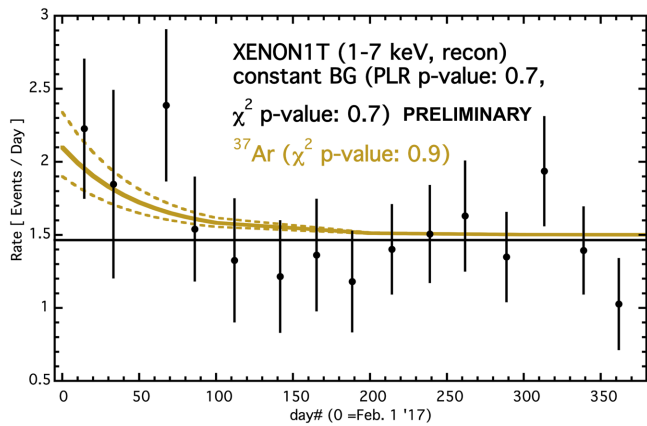
FIG. 10. The time dependence of preliminary XENON1T science data [21] in black. Without the time variation (flat black line), we find $p = 0.7$, matching the p of the XENON1T PLR, in spite of using χ^2-testing instead, showing the similarity of our analysis. We then introduce 31 ± 11 ^{37}Ar decays (yellow lines), as determined from fitting the excess in energy space, with no free parameters, then float the lifetime, counts, and both. The χ^2/d.o.f for all scenarios is <1.0; all hypotheses have corresponding p-values of ~ 0.9. While this does not confirm ^{37}Ar, it certainly does not rule it out either. Moreover, when the decay count is kept fixed at 31 ^{37}Ar events, the best-fit mean $(1/e)$ lifetime is 50^{+40}_{-30} days. When both the lifetime and counts are allowed to float free, the best fit is 57 ± 31 events, and 36 ± 21 days. ^{37}Ar's actual lifetime is 50.6 days (half-life 35.04 days [50]). While errors on lifetime are large, multiple fit versions agreeing at 1σ is a positive hint and shows once more ^{37}Ar is worth investigating thoroughly.

present in both of the two XENON1T science runs [4]. This lifetime consistency implies an introduction mechanism occurring only at the beginning of runs. While we cannot explain conclusively why XENON1T would have 31 ^{37}Ar events, we note LUX observed excess events at an energy consistent with ^{37}Ar. If the LUX peak was new physics, XENON1T would have observed 200–500 events, based on the exposure increase between LUX and XENON1T, not 30 [40]. This discrepancy cannot be accounted for by different efficiencies, since they were similar for both experiments ($\sim 100\%$ at 2.8 keV for ER).

V. DISCUSSION

The excess seen by XENON1T can be effectively reproduced by NEST, and second it may be caused by known physics, other than tritium or other sources already considered [51]. On incorporation of ^{37}Ar into the BG model, disagreement between model and data is 2.2σ (0.4σ Poissonian). This is not completely comparable to PLR, but uses χ^2, like [51], but it might be possible to show even better agreement if we were to fully consider every uncertainty in NEST; we conservatively do not, relying on the default beta yields model.

There is uncertainty for the newly modeled skew [41]. Advantage is never taken of this, using again the central NEST values only based on LUX/ZEPLIN [41,52]. Higher skew, within error, could easily not only add more points at higher S2 in the first few S1 bins of Fig. 3 but also add more counts into the 3.5 and 4.5 keV bins and make ^{37}Ar as good if not a better fit to the XENON1T ER data, when compared again to the less well-motivated (from physics) exponential. That latter notion can itself still be motivated, based on past claims of new physics evidence [53] which may be explicable with exponential (or similar: power-law) rising backgrounds at low energy, across different technologies. We do not speculate on any specific physics to explain it in LXe.

Efficiency and energy reconstruction may contribute to systematics, but primarily at sub-keV; thus, these cannot impact the excess and overall XENON1T result. We acknowledge we had no access to actual XENON1T data and thus had to digitize their plots for comparisons. This can lead to a small error; although, NEST is incredibly robust in its predictions as depicted in the past, and we have put in a system of checks to try to minimize our errors. Therefore, the authors do not believe these would impact our reported results significantly. That said, and as mentioned before, NEST is an open-source software. We urge the XENON1T Collaboration to reproduce our work using their data and/or make their data available publicly. While the results presented here stop short of using PLR, such an analysis for the NEST results will yield more robust conclusions. Although, once more, it is unlikely to change the fact that to first order we have independently reproduced the XENON1T excess and find it consistent with ^{37}Ar. We do not claim to know how it could be introduced, but note such an unexplained excess was previously found in LUX [40].

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 nontrivial: simple rescaling 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 nonbinomial recombination fluctuations). Evidence in favor of our present assumptions ultimately lies in reproduction of XENON1T's data.

APPENDIX: ADDITIONAL VALIDATIONS

This Appendix is secondary evidence to corroborate several of our conclusions. First, we show that the photo-absorption process is capable of reducing the charge yield by ~half (NEST actually assumes a smaller difference) compared to Compton scattering. The 2.8 keV ^{37}Ar peak is the result of e^- capture, so it was not immediately clear which of the two ER models was most appropriate, historically named gamma (photoabsorption would be better) and beta models, even though later data showed that betas agree with Compton scatters within uncertainties, in terms of yield measurements [17,31].

We base our claim of a difference primarily on [54]. In the main text body, this is referred to as the difference between the nominal gamma/x-ray NEST model as opposed to the beta model which covers Compton as well. L_y data were converted into Q_y (even at 0 V/cm) by assuming anticorrelation holds (total of 73 quanta/keV). See Fig. 11. Relative yields were converted into absolute numbers of photons per keV to high precision by converting between 32.1 keV (^{83m}Kr) and 122 keV (^{57}Co) yields, which are nearly identical [55], and then assuming 63 ± 2 photons/keV at 0 V/cm for ^{57}Co γ-rays, a well-established value, given the historic role of this source in calibrating LXe detectors [12]. While many intermediate steps appear in this analysis, each is robustly justifiable.

If one reconsiders Fig. 1, different E-fields may be insufficient to completely explain at least 1σ of difference between the ^{37}Ar PIXeY data [18] and NEST. Recent work by XELDA [57] indicates that there may be 5%–10% differences in yield at different energies and fields, not only between gammas and betas but among many different ER subtypes. The PIXeY data set 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 of 7, S2 just below 2000 phe). This might further help explain reconstruction of 2.8 keV as 2.5 or as low as 2.3.

1. XENON1T's energy reconstruction

As the excess was measured for the energy space histogram not in S2 versus S1 scatter, we also explored energy reconstruction. While the combined-energy scale outperforms the older S1-only [58] or ionization-only employed, e.g., by ν projects [59], it is prone to breakdown at low energy. XENON1T reports reconstructed energy, not true energy that they estimated via MC [6,38]. Figure 12 shows the output from the NEST reconstructed energy, which differs drastically from the true energy especially in the sub-keV regime, in agreement with neriX [31].

While important for other analyses, and although it can create differences of a factor of 2, the discrepancy is not relevant here. It is only particularly evident <1 keV, outside the region of interest for XENON1T's excess.
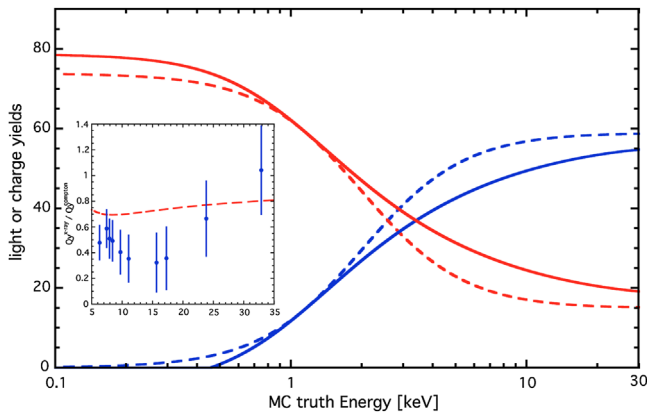

FIG. 11. Comparison of NEST gamma (dash) and beta (solid) models below 30 keV, for 81 V/cm field. The light yield is in blue and charge yield in red. The inset depicts NEST's ratio of charge yields in red dash, along with a comparison to data, dividing the x-ray results of Obodovskii and Ospanov [56] by the Compton scatters from Baudis *et al.*, which also cites the former. Both data sets are from zero field, which is why NEST does not agree well with the data points despite being partly based on them. Direct evidence of the ratio at 81 V/cm does not exist. That being said, NEST is constrained by lower (0) and higher fields; its γ model is extrapolated at nonzero field from high energy. This plot supplements Figs. 1 and 3 (green).
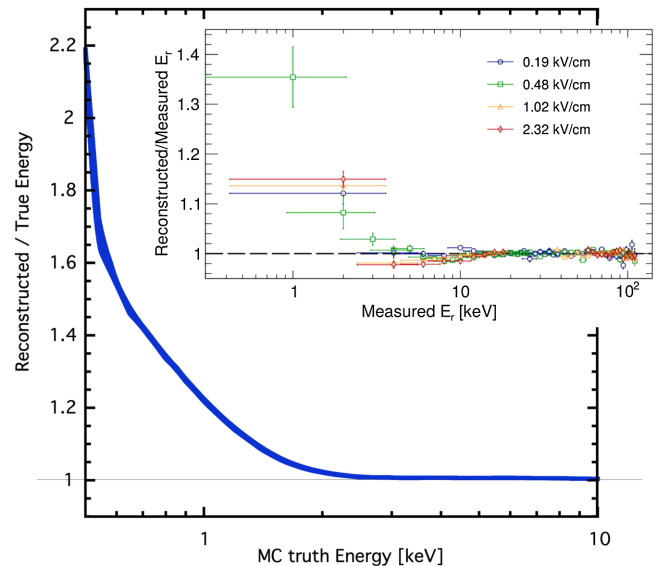

FIG. 12. NEST output comparing true to reconstructed energy, using XENON1T parameters. The thickness of the line indicates statistical uncertainty. The disagreement is an emergent property stemming from many causes, including inherent skew in recombination probability, and triggering on upward fluctuations instead of true mean S1 and S2 pulse sizes, near thresholds (the Eddington bias [60]). Inset: neriX data [31], included for qualitative comparison only, as direct agreement would only be seen by modeling the neriX detector in NEST.

2. Detection efficiency

All of the techniques for estimating detector efficiency ultimately agree on high efficiencies at 2–4 keV, of relevance to the excess. Despite not accounting for detector specifics such as unique S1 pulse shapes [38], comparing NEST with data (Rn to Rn, red to black in Fig. 13), the reduced $\chi^2 = 1.4$ below 5 keV and 1.6 for 1–5 keV. These were calculated with systematics in both the data (Fig. 2 in [4]) and in NEST (difference among red, cyan, green in Fig. 13). This points to NEST's robustness in modeling efficiency, even at energies of only a few keV.

Specifically, ER detection efficiency was verified in four ways: true energy for the x-axis (dark blue line), NEST reconstructed energy (green line) which should match the default XENON1T method (mustard line), simulating a flat energy spectrum (light blue, i.e., "cyan" points), and utilizing the ^{220}Rn beta spectrum (red points), with the latter two but especially red meant to match black. The last three methods all use reconstructed energy, but differ in energy spectrum. Both mustard and black come from XENON1T: the former is their MC estimate and latter their Rn cross-check. NEST cases are compared to them.

Figure 13 demonstrates a good level of agreement among NEST's four scenarios, with the most significant comparisons being red and cyan against black, and green against mustard. Below 1 keV, mistaking the reconstructed energy for true (blue) may cause an overestimation of efficiency but this is challenging to conclude with great certainty given the large errors bars including systematics. One of these systematics is the possibility that the ER light yield
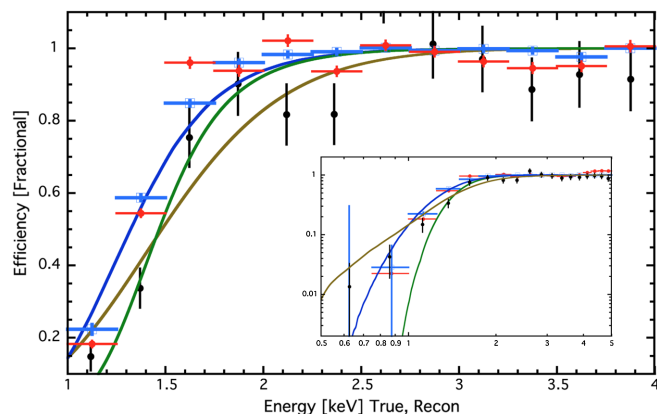

FIG. 13.   The dependence of the relative efficiency on the energy. The mustard is XENON1T's efficiency model and black is data, both from [4], the latter using the ^{220}Rn calibration, which we reproduce using NEST: first with a flat beta model (cyan) and then with the correct ^{220}Rn energy spectrum (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: zoom-out for larger range, log axes. An overall ($\sim$flat) reduction in efficiency across all energy is not portrayed, to focus on shape (actual asymptote <1).

is higher than in NEST near 1 keV and lower energies, closer to what was assumed by XENON1T (Fig. 1) or in earlier NEST versions before sub-keV Q_y's were published (driving L_y estimates downward via anticorrelation). This could easily raise all NEST points and curves up to the mustard in the inset at the very lowest energies. While not directly relevant to the main point of this paper to explain the XENON1T excess since not in the $\sim$2–4 keV ROI or higher, we nonetheless continue to briefly discuss the region below 1 keV in this Appendix, as it may be of interest to the broader community.

The mustard line is above the black points for the first four bins in a row for the inset. However, the differences are always at $\sim 1 - 2\sigma$. What we claim to be the efficiency versus true energy in (dark) blue is sometimes lower, sometimes higher, than the ^{220}Rn points, but diverges from mustard as energy goes to zero. A continuous spectrum such as ^{220}Rn is not best for determining efficiency, even though this was one LUX method [61] (though not for a potential signal). ^{220}Rn was only the cross-check though for XENON1T. Alternatively, a dense series of monoenergetic MC peaks, as naturally done with NEST, can be tuned and verified to match a particular detector's data set, as performed for NRs for LUX [20]. This should explain the difference between the green and mustard if the latter does not originate from a series of monoenergetic peaks. Much contamination between energy bins occurs due to finite resolution in continuous real data [62] that is of course changing rapidly versus energy, with resolution becoming poorer as energy decreases (Fig. 2). If one prefers to study efficiency as a function of reconstructed energy with MC peaks instead of true, both mustard and black may be too high, above green. It is acceptable for black to disagree with green as ^{220}Rn in black (in data) comes from a particular energy spectrum, but mustard and green should agree. Some difference in method and in L_y have already been listed as the possible explanation. Individual PMT effects are another possibility: differing quantum efficiencies and 2-PE probabilities by PMT across the detector, and XYZ-dependent photon collection. Additionally, the exact S1 pulse shape and area-dependent efficiency for detection of single-PE and/or few-PE pulses for the extreme energies, especially given XENON's uniquely long-tailed S1 pulses, with ringing, may be relevant.

Lastly, the green curve, the NEST efficiency versus reconstructed energy for XENON1T, does not agree better with the light blue (cyan) nor red points, also from NEST and also versus reconstructed energy, because it comes from the series of monoenergetic peaks, the results from which are effectively splined together, while red and cyan are influenced by cross-contamination between energy bins as mentioned earlier, from energies both higher and lower than the central value of a particular bin. The reason that red and cyan are not even self-consistent is the difference in energy spectrum (Rn versus flat). Rn is no longer approximately flat in energy near 1 keV.

[1] V. C. Rubin, One hundred years of rotating galaxies, Publ. Astron. Soc. Pac. **112**, 747 (2000).

[2] Y. Akrami *et al.* (PLANCK Collaboration), Planck 2018 results. I. Overview and the cosmological legacy of Planck, Astron. Astrophys. **641**, A1 (2020).

[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 Collaboration), Excess electronic recoil events in XENON1T, Phys. Rev. D **102**, 072004 (2020).

[5] M. Szydagis *et al.*, Open-access noble element simulation technique, https://zenodo.org/record/3905382#.XvlEAZNKjv1.

[6] E. Aprile *et al.* (XENON Collaboration), XENON1T dark matter data analysis: Signal and background models and statistical inference, Phys. Rev. D **99**, 112009 (2019).

[7] 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, J. Instrum. **6**, P10002 2011.

[8] 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).

[9] D. S. Akerib *et al.* (LUX-ZEPLIN Collaboration), Projected WIMP sensitivity of the LUX-ZEPLIN dark matter experiment, Phys. Rev. D **101**, 052002 (2020).

[10] X. Ren *et al.* (PandaX-II Collaboration), Constraining Dark Matter Models with a Light Mediator at the PandaX-II Experiment, Phys. Rev. Lett. **121**, 021304 (2018).

[11] E. Aprile *et al.* (XENON Collaboration), Low-mass dark matter search using ionization signals in XENON100, Phys. Rev. D **94**, 092001 (2016).

[12] 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).

[13] J. Cutter, *The Noble Element Simulation Technique v2* (NorCal HEP-EXchange, 2017), https://indico.physics.lbl.gov/event/560/contributions/1332/attachments/1209/1341/cutter_nest_norcal_2017.pdf.

[14] G. Anton *et al.* (EXO-200 Collaboration), Measurement of the scintillation and ionization response of liquid xenon at MeV energies in the EXO-200 experiment, Phys. Rev. C **101**, 065501 (2020).

[15] 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).

[16] E. Aprile *et al.* (XENON Collaboration), Signal yields of keV electronic recoils and their discrimination from nuclear recoils in liquid xenon, Phys. Rev. D **97**, 092007 (2018).

[17] D. Akerib *et al.* (LUX Collaboration), Improved measurements of the β-decay response of liquid xenon with the LUX detector, Phys. Rev. D **100**, 022002 (2019).

[18] E. Boulton *et al.*, Calibration of a two-phase xenon time projection chamber with a ^{37}Ar source, J. Instrum. **12**, P08004 (2017).

[19] 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, J. Instrum. **10**, P09010 (2015).

[20] 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).

[21] E. Shockley, Search for new physics with electronic recoil events in XENON1T, LNGS Seminar, 2020, https://agenda.infn.it/event/23228/.

[22] A. Behrens, Light detectors for the XENON100 and XENON1T dark matter search experiments, Ph.D. Thesis, Universitaet Zurich, 2014.

[23] E. Aprile *et al.* (XENON Collaboration), Dark Matter Search Results from a One Ton-Year Exposure of XENON1T, Phys. Rev. Lett. **121**, 111302 (2018).

[24] E. Aprile *et al.* (XENON Collaboration), The XENON1T dark matter experiment, Eur. Phys. J. C **77**, 881 (2017).

[25] B. L. Paredes, H. Araújo, F. Froborg, N. Marangou, I. Olcina, T. Sumner, R. Taylor, A. Tomás, and A. Vacheret, Response of photomultiplier tubes to xenon scintillation light, Astropart. Phys. **102**, 56 (2018).

[26] E. Aprile *et al.* (XENON Collaboration), First Dark Matter Search Results from the XENON1T Experiment, Phys. Rev. Lett. **119**, 181301 (2017).

[27] D. Akerib *et al.* (LUX Collaboration), Improved modeling of β electronic recoils in liquid xenon using LUX calibration data, J. Instrum. **15**, T02007 2020.

[28] B. Edwards *et al.*, Extraction efficiency of drifting electrons in a two-phase xenon time projection chamber, J. Instrum. **13**, P01005 2018.

[29] 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).

[30] 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 University, 2009.

[31] 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).

[32] E. Aprile *et al.* (XENON Collaboration), Energy resolution and linearity of XENON1T in the MeV energy range, Eur. Phys. J. C **80**, 785 (2020).

[33] E. Aprile *et al.* (XENON1T Collaboration), Observation of two-neutrino double electron capture in ^{124}Xe with XENON1T, Nature (London) **568**, 532 (2019).

[34] D. S. Akerib *et al.* (LUX Collaboration), Signal yields, energy resolution, and recombination fluctuations in liquid xenon, Phys. Rev. D **95**, 012008 (2017).

[35] R. F. Lang, A. Brown, E. Brown, M. Cervantes, S. Macmullin, D. Masson, J. Schreiner, and H. Simgen, A ^{220}Rn source for the calibration of low-background experiments, J. Instrum. **11**, P04004 (2016).

[36] M. Szydagis, A. Fyhrie, D. Thorngren, and M. Tripathi, Enhancement of NEST capabilities for simulating low-energy recoils in liquid xenon, J. Instrum. **8,** C10003 2013.

[37] G. Rischbieter, Background modeling in the LUX detector for an effective field theory dark matter search, April 2020 APS Meeting.

[38] E. Aprile *et al.* (XENON1T Collaboration), XENON1T dark matter data analysis: Signal reconstruction, calibration and event selection, Phys. Rev. D **100,** 052014 (2019).

[39] A. Dobi, Measurement of the electron recoil band of the LUX dark matter detector with a tritium calibration source, Ph.D. Thesis, University of Maryland College Park, 2014.

[40] D. Akerib *et al.* (LUX Collaboration), Search for annual and diurnal rate modulations in the LUX experiment, Phys. Rev. D **98,** 062005 (2018).

[41] D. Akerib *et al.* (LUX Collaboration), Discrimination of electronic recoils from nuclear recoils in two-phase xenon time projection chambers, Phys. Rev. D **102,** 112002 (2020).

[42] E. Boulton, Applications of two-phase xenon time projection chambers: Searching for dark matter and special nuclear materials, Ph.D. Thesis, Yale University, 2019.

[43] N. Priel, L. Rauch, H. Landsman, A. Manfredini, and R. Budnik, A model independent safeguard against background mismodeling for statistical inference, J. Cosmol. Astropart. Phys. 05 (2017) 013.

[44] D. Akerib *et al.* (LUX Collaboration), Results from a Search for Dark Matter in the Complete LUX Exposure, Phys. Rev. Lett. **118,** 021303 (2017).

[45] H. An, M. Pospelov, J. Pradler, and A. Ritz, New limits on dark photons from solar emission and keV scale dark matter, arXiv:2006.13929.

[46] I. M. Bloch, A. Caputo, R. Essig, D. Redigolo, M. Sholapurkar, and T. Volansky, Exploring new physics with O(keV) electron recoils in direct detection experiments, arXiv:2006.14521.

[47] H.-J. He, Y.-C. Wang, and J. Zheng, EFT analysis of inelastic dark matter for xenon electron recoil detection, arXiv:2007.04963.

[48] G. Alonso-Álvarez, F. Ertas, J. Jaeckel, F. Kahlhoefer, and L. J. Thormaehlen, Hidden photon dark matter in the light of XENON1T and stellar cooling, arXiv:2006.11243.

[49] L. A. Anchordoqui, I. Antoniadis, K. Benakli, and D. Lust, Anomalous $U(1)$ gauge bosons as light dark matter in string theory, Phys. Lett. B **810,** 135838 (2020).

[50] D. Akimov *et al.*, Experimental study of ionization yield of liquid xenon for electron recoils in the energy range 2.8–80 keV, J. Instrum. **9,** P11014 (2014).

[51] B. Bhattacherjee and R. Sengupta, XENON1T excess: Some possible backgrounds, arXiv:2006.16172.

[52] V. N. Lebedenko *et al.* (ZEPLIN-III Collaboration), Results from the first science run of the ZEPLIN-III dark matter search experiment, Phys. Rev. D **80,** 052010 (2009).

[53] C. Aalseth *et al.* (CoGeNT Collaboration), Results from a Search for Light-Mass Dark Matter with a P-type Point Contact Germanium Detector, Phys. Rev. Lett. **106,** 131301 (2011).

[54] L. Baudis, H. Dujmovic, C. Geis, A. James, A. Kish, A. Manalaysay, T. M. Undagoitia, and M. Schumann, Response of liquid xenon to Compton electrons down to 1.5 keV, Phys. Rev. D **87,** 115015 (2013).

[55] A. Manalaysay, T. M. Undagoitia, A. Askin, L. Baudis, A. Behrens, A. Ferella, A. Kish, O. Lebeda, R. Santorelli, D. Venos, and A. Vollhardt, Spatially uniform calibration of a liquid xenon detector at low energies using ^{83m}Kr, Rev. Sci. Instrum. **81,** 073303 (2010).

[56] I. Obodovskii and K. Ospanov, Scintillation output of liquid xenon for low-energy γ-quanta, Pribory I Tekhnika Eksperimenta (USSR) **26,** 42 (1994), https://inis.iaea.org/search/search.aspx?orig_q=RN:26045323.

[57] D. Temples, Understanding neutrino background implications in LXe-TPC dark matter searches using ^{127}Xe electron captures (TAUP 2019), http://www-kam2.icrr.u-tokyo.ac.jp/indico/event/3/session/10/contribution/414/material/slides/0.pdf.

[58] E. Aprile *et al.* (XENON100 Collaboration), First axion results from the XENON100 experiment, Phys. Rev. D **90,** 062009 (2014).

[59] C. Adams *et al.* (MicroBooNE Collaboration), Calibration of the charge and energy loss per unit length of the MicroBooNE liquid argon time projection chamber using muons and protons, J. Instrum. **15,** P03022 (2020).

[60] E. Aprile *et al.* (XENON Collaboration), Light Dark Matter Search with Ionization Signals in XENON1T, Phys. Rev. Lett. **123,** 251801 (2019).

[61] D. Akerib *et al.* (LUX Collaboration), Tritium calibration of the LUX dark matter experiment, Phys. Rev. D **93,** 072009 (2016).

[62] J. Balajthy, Purity monitoring techniques and electronic energy deposition properties in liquid xenon time projection chambers, Ph.D. Thesis, University of Maryland College Park, 2018.