Statistical re-examination of reported emission lines in the X-ray afterglow of GRB 011211

Robert E. Rutledge* and Masao Sako*†
Theoretical Astrophysics, California Institute of Technology, MS 130-33, Pasadena, CA 91125, USA

Accepted 2002 September 6. Received 2002 September 4; in original form 2002 June 5

ABSTRACT

A 0.2–12 keV spectrum obtained with the XMM–Newton EPIC/pn instrument of GRB 011211, taken in the first 5 ks of a 27-ks observation, was found by Reeves et al. in 2002 to contain emission lines that were interpreted to be from Mg XI, Si XIV, S XVI, Ar XVIII and Ca XX, at a lower redshift \( z_{\text{obs}} = 1.88 \) than that of the host galaxy \( z_{\text{host}} = 2.14 \). We examine the spectrum independently, and find that the claimed lines would not be discovered in a blind search. Specifically, Monte Carlo simulations show that the significance of reported features, individually, is such that they would be observed in 10 per cent of featureless spectra with the same signal-to-noise ratio. Imposing a model in which the two brightest lines would be Si XIV and S XVI \( K\alpha \) emission velocity shifted to between \( z = 1.88 \) and 2.40, such features would be found in \( \sim 1.3–1.7 \) per cent of observed featureless spectra (that is, with 98.3–98.7 per cent confidence). When we account for the number of trials implicit in a search of five energy spectra (as were examined by Reeves et al.), and permit a wider \( z \) phase space search \( (z = 2.14 \pm 1.0) \), the detection confidence of the two-line complex decreases to 77–82 per cent. We find the detection significance to be insufficient to justify the claim of detection and the model put forth to explain them. \( K\alpha \) line complexes are also found at \( z = 1.2 \) and 2.75 of significance equal to or greater than that at \( z = 1.88 \). Thus, if one adopts the \( z = 1.88 \) complex as significant, one must also adopt the other two complexes to be significant. The interpretation of these data in the context of the model proposed by Reeves et al. is therefore degenerate, and cannot be resolved by these data alone. Our conclusions are in conflict with those of Reeves et al., because our statistical significance accounts for the multiple trials required – but not accounted for by Reeves et al. – in a blind search for emission features across a range of energies. In addition, we describe a practical challenge to the reliability of Monte Carlo \( \Delta \chi^2 \) tests, as employed by Reeves et al.

Key words: gamma rays: bursts – gamma rays: observations.

1 INTRODUCTION

It was recently reported that the X-ray afterglow of the \( \gamma \)-ray burst GRB 011211, as observed with XMM–Newton EPIC/pn, contained spectral emission lines (Reeves et al. 2002a, hereafter R02) – the first report of multiple X-ray emission lines from a GRB. These lines, at 0.45, 0.70, 0.89, 1.21 and 1.44 keV, were interpreted to be from He-like Mg xi (rest energy 1.35 keV) and H-like Si xiv (2.0 keV), S xvi (2.62 keV), Ar xviii (3.32 keV) and Ca xx (4.10 keV), redshifted to \( z = 1.88 \). The difference between this and the known redshift of the host galaxy \( z_{\text{host}} = 2.14 \) was modelled as due to supernova ejecta travelling at \( v = 25800 \pm 1200 \) km s\(^{-1}\), which had originated during a supernova 4 d prior to when the GRB jet illuminated it, producing the afterglow [a more detailed analysis by the same authors (Reeves et al. 2002b, hereafter R02b) was completed after this paper was in its initial form].

The statistical significance of the individual lines was not reported in R02; it was stated that joint analysis of the lines taken together produced an improvement in the \( \chi^2 \) value which, by an \( F \)-test, yielded a significance level of 99.7 per cent. In addition, it was found that Monte Carlo (MC) simulations were unable to produce the same improvement in \( \chi^2 \) found between the best-fitting power-law model and the best-fitting five emission-line model more than 0.02 per cent of the time. Specifically, it was found that the best-fitting \( \chi^2 \) value for a power-law model was improved by fitting to a model of a MEKAL plasma with emission lines at rest energies corresponding to unresolved Mg xi, Si xiv, S xvi, Ar xviii and Ca xx redshifted to...
The implications of the model discussed by R02—a delay between a supernova and a GRB on a time-scale of days, the formation of a thin shell of supernova ejecta, an apparent underabundance of Fe relative to the detected nuclei—provide severe constraints on γ-ray burst emission models. In addition, as demonstrated by R02, the future detection of multiple emission lines can provide extremely strong constraints on the production mechanisms, owing to the inherent required outflow velocity and emission time-scales that can be derived from them, not to mention the implied association with supernovae. Similar spectra observed with greater signal-to-noise ratio (S/N) in the future would greatly aid in unravelling the emission mechanisms and geometry of γ-ray bursts. Therefore, it is of wide theoretical (e.g. Lazzati, Ramirez-Ruiz & Rees 2002; Kumar & Narayan 2002) and observational interest to interpret the observed X-ray spectrum of this GRB 011211 further, in the hope of determining what more could be learned from future, more precise, observations.

In Section 2, we describe the observation, and perform a basic spectral analysis using continuum models. In Section 3, we compare Monte Carlo (MC) realizations of acceptable continuum models with the GRB spectrum, and find that the reported features would be produced in ∼10 per cent of the continuum model spectra, due only to Poisson noise. In Section 4, we adopt the model that the two apparently most significant lines are Kα lines of SiXIV and SXIV; we perform a blind search for features of the same significance in MC realizations of continuum spectra, and find that they would be reported in ∼1.2–2.6 per cent of such spectra, again, due only to Poisson noise. We describe in Section 5 a practical challenge to the reliability of the MC Δχ² analysis produced by R02. We conclude in Section 6 that the lines are not individually significant in the absence of an imposed model, and are only marginally significant when the adopted model is imposed.

These conclusions conflict with those of R02. R02 derived the model (that is, the observed line energies, or redshift) from the data, and then applied statistics for detection as if the energies were known prior to examining the data (that is, single-trial statistics). This is not appropriate when the model line energies are derived directly from the X-ray data, and not from an a priori model—one derived without examination of the X-ray data (for example: line energies of multiple features with redshifts of the host galaxy). We adopt statistics appropriate to a blind search for these features, across a range of energies or redshifts (multi-trial statistics). This accounts for the diminished significance we find for the features. We further discuss the reasons for this conflict and conclude in Section 6.

2 OBSERVATION AND OBSERVED SPECTRUM

We analysed the identical source and background XMM–Newton EPIC/pn (Strüder et al. 2001) spectrum as used by R02 (their fig. 2), which was kindly made available to us by the authors in electronic form (J. N. Reeves, private communication). We used the same response matrix (epn_ff20_sdi9_thin.rsp). The spectrum used 5000 s of real-time observation beginning at 07:14:33 UT on 2001 December 12, with a total live time of 4440 s. The pn spectrum used counts comprising patterns 0–4 (singles and doubles), from a circular region 46 arcsec in radius centred on the source, excluding flagged events (for which the keyword FLAG! = 0) and excluding a region near the edge of the charge-coupled device (CCD) chip.

We performed a basic spectral analysis using xspec v11.1.0 (Arnaud 1996). We used data in the 0.2–12 keV energy range. We performed a non-standard spectral binning, implemented to maximize the signal-to-noise ratio associated with the reported emission features at the reported energies. We first binned data with energy bins centred at the five best-fitting line energies found by R02, with bin sizes approximately equal to the full width at half-maximum (FWHM) EPIC/pn energy response at each energy (respectively: 62, 66, 68, 72 and 75 eV; see equation 1). The remaining data were binned with 60 eV or greater (<1 keV), and 70 eV or greater (>1 keV). Between 0.2 and 5 keV, each bin has >15 counts (although they were not binned on this basis), for which χ² fitting is valid.

We fit an absorbed photon power-law spectrum to the data, shown in Fig. 1. The model spectrum was statistically acceptable (model parameters are listed in Table 1, along with the obtained χ² values). We also fit the model with a thermal bremsstrahlung spectrum (wabs*brms), and derived an acceptable best fit. Finally, we found best-fitting model parameters for the values of the power-law photon slope (models 2 and 3) and kTbrms (models 5 and 6) at the 90 per cent confidence limits of the best fit, which will be used in MC simulations in Section 3.

3 INDIVIDUAL EMISSION-LINE SIGNIFICANCES

We first determine which of the reported lines are individually statistically significant, when one is searching for emission lines at a priori known energies. We compared the observed spectrum with MC simulations of six featureless spectra—the three absorbed power-law and three absorbed thermal bremsstrahlung models, which are models 1–6 in Table 1.

We used a ‘matched filter’, convolving the observed pulse-invariant (PI) counts spectrum with a Gaussian energy response, with the energy resolution response of the detector. The matched filter approach maximizes the signal-to-noise ratio as a function of energy of unresolved lines in the X-ray PI spectrum.

Based on fig. 18 in the XMM–Newton v1.1 Users’ Handbook (Dahlem 1999), we modelled the photon energy redistribution as a Gaussian response, with FWHM...
FWHM($E$) = $57 + 13(E/1 \text{ keV}) - 0.29(E/1 \text{ keV})^2$ eV.  \hspace{1cm} (1)

This approximation was derived from the line in this figure. The EPIC/pn energy resolution has been demonstrated to be stable over nine months of in-flight calibration (Strüder et al. 2001). We expect that this analysis (and that of R02, since that work is based on the same energy response matrices) is valid as long as the energy resolution is within 20 per cent of this approximation (corresponding to three of $\sim 15$ PI channels at 0.75 keV).

We performed a convolution between the raw PI spectrum (that is, number of counts versus PI bin) and the Gaussian energy response function, as a function of energy:

\[ C(E_i) = \sum_{j} I(j) \frac{1}{\sqrt{2\pi} \sigma(E_j)} \exp \left( \frac{-1}{2} \left( \frac{E_i - E_j}{\sigma(E_j)} \right)^2 \right) \delta E_j, \]

where $N$ is the number of PI bins, and we sum across PI bins that are within $\pm 3\sigma(E_i)$ of $E_i$. $I(j)$ is the raw PI spectrum, which contains both source and background counts, and $j = 1, 2, \ldots, N$ is the PI bin number. The centroid (average) energies and energy widths ($\Delta E_i$) of the PI bins were taken from the EBOUNDS extension of the response matrix, where $i$ is the PI bin number and $\sigma(E) = \text{FWHM}(E)/2.35$. We do not correct the PI spectrum for the detector area; however, the detector area does not change dramatically across the FWHM of the lines. If the area did change dramatically across the FWHM of a line, and a statistical excess were observed in the area-corrected PI spectrum but not in the raw PI spectrum, then such an excess could well be due to calibration uncertainties.

The resulting $C(E_i)$ is shown in Fig. 2(a). By visual inspection, there are indeed features in the spectrum near energies where the reported lines occur. To determine if these features are significant, we produced MC spectra of models 1–6 (see Section 2). The MC realizations of the raw PI spectra were performed as follows. We simulated the spectral models 1–6 in XSPEC, using the same response matrix as above, so that the resulting PI spectra (without Poisson noise added) were convolved as the observed spectrum through the telescope and detector response. The simulated PI spectra $N(E)$ each had a total of $> 9 \times 10^8$ counts in PI bins between 0.2 and 3 keV.

We then produced integrated spectra

\[ I(E) = \int_{0.2 \text{ keV}}^{E} N(E) \, dE / \int_{0.2 \text{ keV}}^{3 \text{ keV}} N(E) \, dE, \]

so that $I(0.2 \text{ keV}) = 0$ and $I(3 \text{ keV}) = 1$ (the integrated normalized model is used for the MC simulation as described below). These

![Figure 2](https://example.com/f2.png)

Figure 2. (a) The solid line is $C(E)$ (equation 2) from the observed raw PI spectrum – the convolution between the raw spectrum and the EPIC/pn energy response. The broken lines are the maximum $C(E)$ for spectral models 1–6, showing the 99 and 99.9 per cent confidence single-trial upper limits. (b)–(f) The solid lines are the same observed convolved spectrum as in (a). The dotted lines are five (in the five separate panels) randomly selected Monte Carlo spectra using model 1. Features of similar magnitude to those found in the observed spectra are apparent in each; these are due to the Poisson noise distribution (in energy) in a spectrum with a finite number of detected counts.

© 2003 RAS, MNRAS 339, 600–606
constitute our six acceptable featureless spectral models; we will compare the data with results from all six, as a firm conclusion that emission lines are present should be independent of the underlying broad-band model assumed.

We implemented a background spectral model, to simulate the ~10 per cent of the counts due to background. Taking background from a different part of the detector, we find that it can be parametrized by a broken photon power law (bknpower), with $\alpha_1 = 2.4$ at low energies, break energy 1.35 keV, and $\alpha_2 = 0.44$ at high energies, between 0.20 and 7.3 keV (there is a strong background line at 8 keV). In fact, there are statistically significant deviations from this pure continuum model between 0.55 and 0.6 keV; we ignore these deviations in our background model. In our MC simulation the effect of ignoring what would appear to be a line in the observed spectrum is conservative, in the sense that, by ignoring its presence in the background model, we could detect as ‘significant’ a line in the 0.55–0.6 keV range that is in fact produced by instrument background.

We simulated spectra between 0.2 and 3 keV, in which there were 560 counts in the observed spectrum, of which we estimate that $\sim 66 \pm 1.2$ counts are due to background. We drew, for each MC realization, a number of background counts which is a random Poisson deviate (using poidev, Press et al. 1995) with an average of 66 counts, with the remaining (of 560) counts from the source. To produce a simulated spectrum, we generate a random uniform deviate $r$ between 0 and 1, and we place a count in the PI bin in which $I(E) = r$.

To produce our confidence limits to $C(E)$, we produced 1667 MC realizations each of models 1–6 for a total of 10 002 realizations. We set the 99 and 99.9 per cent confidence limits at the 100th and 10th greatest values, respectively, of $C(E)$ of all such realizations. This ensures that the conclusions are not dependent upon the assumed featureless spectral model.

The results of this MC simulation are shown in Fig. 2(a). Two of the reported features (near 0.7 and 0.85 keV, Si xiv and S xvi) have single-energy-trial probabilities of >99 per cent confidence in comparison with the featureless spectral models (the claimed S xvi line peaks just below the 99.9 per cent confidence limit; we will treat it as having met 99.9 per cent confidence, while the reader may regard this as an upper limit). The remaining three lines are not significant in comparison with single-energy-trial probability of 99 per cent confidence.

In Fig. 2, we also show $C(E)$ for single MC realizations of the best-fitting power-law spectrum (model 1), which also contain apparent features. The bumps in the single simulated spectra appear because, in any spectrum that contains Poisson noise, the counts will not be distributed uniformly in energy, but will be clustered in energy simply due to counting statistics.

### 3.1 Multi-energy-trial (blind search) probabilities

Since it was necessary to perform a blind search for emission-line features in the GRB spectrum – as the redshifted line energies were not known a priori, but were measured from the data – it is necessary to estimate the chance probability that the reported features are produced from a featureless spectrum during a blind search for such features.

We produced 10 000 MC realizations for each of models 1–6 as described in the previous section. We compared the $C(E)$ of these between 0.4 and 1.5 keV against the single-energy-trial 99 and 99.9 per cent confidence limits we found in the previous section, for the models 1–6 individually.

| Model | >99 per cent (one $z$ bin) | >99.9 per cent (one $z$ bin) |
|-------|---------------------------|-----------------------------|
| 1     | 0.78                      | 0.16                        |
| 2     | 0.78                      | 0.15                        |
| 3     | 0.78                      | 0.14                        |
| 4     | 0.78                      | 0.15                        |
| 5     | 0.79                      | 0.16                        |
| 6     | 0.78                      | 0.17                        |

In Table 2 we list the fraction of the 10 000 MC spectra in which $C(E)$ in at least one PI bin reaches a single-energy-trial probability of 99 or 99.9 per cent confidence. These fractions are 78–79 and 14–17 per cent, respectively; if finding a single-energy-trial 99 per cent (Si xiv) and 99.9 per cent (S xvi) feature were statistically independent, then the probability of observing both a 99 and 99.9 per cent single-energy-trial ‘line’ in a single spectrum, such as we find in the present spectrum of GRB 011211, is $\sim 10$ per cent.

Therefore, in a blind search of the EPIC/pn spectrum for emission features, we would expect to find features that have single-energy-trial significance equal to or greater than those observed in one of approximately 10 observed featureless spectra.

### 4 LINE COMPLEX SIGNIFICANCE AS A FUNCTION OF REDSHIFT

In this section, we examine if the reported lines, taken together, implicate $Ke$ emission features from the particular redshift of $z = 1.88$ as reported by R02. We do so by summing the $C(E/(1+z))$, using the values of the rest energies of the reported lines:

$$
\chi(z) = \sum_{i=1}^{N_{lines}} \sum_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma(E_j)} \exp\left( -\frac{1}{2} \left( \frac{E_j - (E_{line}/(1+z))}{\sigma(E_j)} \right)^2 \right),
$$

(3)

where $j$ denotes the PI bin number, $E_j$ is the centroid energy of the $j$th PI bin, $i$ denotes the index $[1, 5]$ of the five lines detected, and $E_i$ denotes the rest energies of the five lines identified by R02, which were 1.35 (Mg ii), 2.00 (Si xiv), 2.62 (S xvi), 3.32 (Ar xvii) and 4.10 keV (Ca xx). We examined the range of $0 < z < 3$, with a step size of $\Delta z = 0.015$. We use PI bins with energies 0.1–7 keV, to cover the spectrum past the rest-frame energy of Ca xx. We find 638 counts in this energy range, of which we estimate 80 are from background. We use only bins that are within $3\sigma(E_j/(1+z))$ of each $E_i/(1+z)$. The result of this convolution, if the reported lines are real, should be a maximum in $\chi(z)$ near the optimal redshift value, in excess of that found from MC realizations of data with featureless spectra.

The average value of $\chi$ will systematically increase with increasing $z$ as the lines are shifted to lower energies, where the intensity is higher in the power-law spectrum and the detector effective area is larger and so there are a greater number of counts. To examine if any particular maximum in $\chi(z)$ is significant, we performed this convolution for 10 000 MC realizations using the simulated spectral models 1–6, taking the 100th and 10th highest values, as described in the previous section, to produce the 99 and 99.9 per cent confidence limits respectively.
The results of the calculation using all five reported lines, as well as the 99 and 99.9 per cent MC confidence limits, are shown in Fig. 3. The value of $\chi(z)$ is in excess of the 99 per cent MC confidence limit at $z = [1.86-1.98]$ and $z = [2.62-2.865]$, and in excess of the 99.9 per cent MC confidence limit at $z = 2.76$.

We also performed this convolution and Monte Carlo simulation using what appear to be the most significant two lines from fig. 2 of R02 (Si xiv and S xvi), the results of which are also shown in Fig. 3. The value of $\chi(z)$ is in excess of the 99 per cent MC confidence limit at $z = [1.155-1.275]$ and $z = [1.80-2.01]$, and in excess of the 99.9 per cent confidence limit at $z = [1.86-1.95]$.

4.1 Multi-redshift-trial (blind search) probabilities

What fraction of featureless spectra, with the same number of source and background counts as the observed spectrum, would produce values of $\chi(z)$ of comparable significance to the excess in $\chi(z = 1.88)$ from the observed spectrum? If one examines $\chi(z)$ only at $z = 1.88$, the answer is $< 1$ per cent, which is the single-$z$-trial probability. However, the reported $z = 1.88$ is different from the known redshift of the host galaxy $z = 2.14$; it is therefore unlikely that $z = 1.88$ was the only redshift that would be considered consistent with an a priori model by R02. The pertinent statistical question to ask, then, is what is the fraction of featureless X-ray spectra, examined for a redshifted pair of S xvi and Si xiv lines, that would produce a value of $\chi(z)$ comparable to the single-$z$-trial significance observed, allowing for a blind search at values of $z$ between 1.88 and 2.40 (a range of equal magnitude redshift and blueshift from the host galaxy)?

To address this, we simulated 10 000 MC spectra of each of spectral models 1–6, and found $\chi(z)$ in the same way as for the observed spectrum in the previous section. We used only the two-line model, as this gave the apparently most significant result near $z = 1.88$. We compared $\chi(z)$ with the 99 and 99.9 per cent MC limits, found in the previous section, and noted that these were exceeded in at least one $z$ bin for the 99 per cent confidence limit, and in at least seven consecutive $z$ bins for the 99.9 per cent confidence limit between $z = 1.88$ and $z = 2.40$. We require seven consecutive $z$ bins as this is the number of $z$ bins in $\chi(z)$ we find in excess of the single-trial 99.9 per cent confidence limit near $z = 1.88$. [We require only one bin for the 99 per cent confidence limit to satisfy a minimal ‘detection’ requirement; whereas we require seven bins for the 99.9 per cent confidence limit, since this was what was actually observed near $z = 1.88$, and we wish to evaluate the likelihood of producing the observed $\chi(z)$ excess.]

The fraction of MC featureless spectra that contained at least one $z$ bin between $z = 1.88$ and $z = 2.40$ in excess of the single-$z$-trial MC probability of 99 per cent is given in Table 3. Because the observed spectrum gave $\chi(z) > 99.9$ per cent in seven consecutive $z$ bins, we also used this as our criterion to count ‘hits’ in the >99.9 per cent confidence comparison, also shown in Table 3. Of 10 000 MC spectra, between 20 and 22 per cent produced ‘hits’ for the single-$z$-trial 99.9 per cent confidence limit, and 1.3–1.7 per cent produced ‘hits’ for the single-$z$-trial 99.9 per cent confidence limit.

We note that when we search the range $z = 2.14 \pm 1.0$ (instead of $\pm 0.26$) the percentage of featureless spectra that have seven consecutive $z$ bins with $\chi(z)$ greater than the 99.9 per cent limit is 3.8–5.0 per cent. However, it is unclear if R02 would have attached equal significance to a detection at $z = 1.14$ as one at $z = 1.88$, as no limits on excess line emission as a function of assumed redshift are given, and the redshift phase space examined by R02 was not given. We therefore rely on our search of the smaller phase space; while this may underestimate the number of ‘effective trials’ used by R02, it none the less serves as the probability of producing the claimed excess line emission due to a statistical fluctuation within the $\Delta z = 0.26$ observed. If the redshift space examined by R02 were 1.14–3.14 ($\Delta z = 1.0$), then the probability of finding an excess equal to or greater than that observed would be 3.8–5.0 per cent. If the full redshift space of 0–5 was in fact examined by R02, then the probability of a false detection is $>3.8$–5.0 per cent.

5 A PRACTICAL CHALLENGE WITH MONTE CARLO $\Delta \chi^2$ TESTS FOR MULTIPARAMETER MODELS

An MC $\Delta \chi^2$ test as employed by R02 is not fundamentally flawed as is the analytic $\Delta \chi^2$ test (that is, the $F$-test) for the application of spectral emission-line discovery. In the $F$-test, the reference $\Delta \chi^2$ distribution was derived under the assumption that the null hypothesis lies on the border of the acceptable parameter space (Protassov et al. 2002), which is not true in a search for emission lines; however,
this assumption is not made in the MC $\Delta \chi^2$ test. Thus, the simulated $\Delta \chi^2$ distribution can, in principle, provide a reliable reference distribution with which the value of $\Delta \chi^2$ from application to real data can be compared to determine the false positive rate.

However, as we show below, the MC $\Delta \chi^2$ test as employed by R02 (and described more fully by R02b) suffers from a practical problem which makes it an inferior approach to the one we have applied. Specifically, to apply the $\Delta \chi^2$ statistic using the MC approach, one must assuredly find the global minimum $\chi^2$ for the applied model for every single MC realization; the description of the analysis performed by R02 (and R02b) does not assure that this has occurred.

In the case of $\chi^2$ minimization through local mapping of the $\chi^2$ surface, as in the modified Levenberg–Marquart method employed in XSPEC (Arnaud 1996; modified from the CURFIT algorithm as described by Bevington 1969; see also Press et al. 1995), one finds the vector in multiparameter space along this surface which provides the most negative derivative, follows along this vector a short way, and iterates, until one reaches a point where there are no negative derivatives in any direction along the $\chi^2$ surface (that is, when one has reached a minimum point). This approach suffers from the well-known problem of local minima, where the true global minimum can lie at a completely different set of parameter values (see, for example Press et al. 1995, p. 394). On simple $\chi^2$ surfaces, where the second partial derivatives of the $\chi^2$ surface are everywhere small – certainly in the case of the two-parameter power-law spectrum – it is rare that local minima different from the global minimum are found. However, on complex $\chi^2$ surfaces (those which contain large second partial derivatives of $\chi^2$) – as will be the case when fitting a six-parameter model of three emission lines of the spectrum from the EPIC/pn detector edge during the observation, which would have been included in the GRB spectrum from the first 5 ks, when the source was near the detector edge, but not afterwards, after the source had been moved away from the detector edge. In our own analysis, we cannot confirm this result unless we adopt non-standard event selection criteria, which differ from the ones used by R02. R02 removed events near the CCD detector edge (FLAG = 0) and selected only single and double events (PATTERN < 4) (J. N. Reeves, private communication). These selections result in a smooth, featureless background spectrum with no bright line-like feature near ~0.7 keV (as seen in fig. 4f of Borozdin & Trudolyubov 2003) as well as a reduction of the count rate by a factor of $\geq 2$ in the range $E = 0.2–3$ keV (see Fig. 4). Therefore, we are not able to confirm the applicability of Borozdin & Trudolyubov (2003) to the analysis of R02.
Figure 4. The background spectrum near the region near the CCD chip edge using two different event selection criteria: (a) no explicit selection of PATTERN and FLAG as adopted by Borozdin & Trudolyubov (2003) (circles) binned at a minimum of five counts per bin; and (b) using only PATTERN $<$ 4 and FLAG = 0 events as adopted by R02 (stars) with bin sizes identical to those of (a).

An alternative approach to the one we have taken is employed using XSPEC, in which one fits a featureless spectrum to the data, and then a spectrum that includes emission lines, to determine if the change in $\chi^2$ is significant, as according to an $F$-test; this is the approach taken by R02. However, this approach for the detection of emission or absorption lines is formally incorrect, and gives false statistical results (Protassov et al. 2002) particularly so when the true continuum is not well constrained, as in the present case. We therefore prefer our approach of applying a matched energy response filter for line detection at arbitrary energies, and to compare this with application of the matched filter to MC realizations of featureless spectra. It is a trivial statistical exercise to demonstrate that matched filtering maximizes the signal-to-noise ratio (and thus detectability) for detection of infinitely narrow emission lines.

In estimating the model-dependent confidence limit for the detection of the line complex (98.7 per cent), we accounted only for searching the redshift phase space between $z = 1.88$ and $z = 2.40$, symmetric about the host galaxy redshift – an extremely minimal requirement. We did not account for the full redshift phase space searched by R02, as such was not given in that reference; if the redshift phase space searched by R02 covered $z = 1.14$–3.14 ($z = 2.14 \pm 1.0$), then the detection significance of the two strongest lines (Si XIV and S XVI) together decreases from 98.3–98.7 to 95–96.2 per cent confidence. Finally, we did not include in this confidence limit the number of trials implicit in searching five X-ray spectra for emission lines, which was performed by R02 for different time periods (0–5, 5–10, 10–15, 15–20 and 20–27 ks). If we presume the same search was made on all five spectra, as seems a reasonable a priori search to perform, then the detection confidence for the Si XIV and S XVI lines together decreases to 0.95–0.962 $= 77$–82 per cent. We regard 98.7 per cent to be a conservative (in the sense of permitting a higher significance) upper limit to the confidence of detecting the Si XIV and S XVI lines together, while a more accurate accounting of the number of trials and phase space searched by R02 produces a 77–82 per cent confidence limit.

We consider neither a 90 per cent confidence detection in a model-independent interpretation, nor a 98.3–98.7 per cent confidence detection in a model-dependent interpretation, to be sufficient to justify the detection claims and subsequent interpretation put forth by R02. The 77–82 per cent confidence limit, which accounts for the wide redshift phase space and number of spectra examined by R02, is well below any comfortable detection confidence. If the redshift phase space actually searched by R02 is larger, the number of implicit trials is greater, and our estimate of the confidence level for the detected line complex would decrease.

Moreover, if one concludes that the marginal detection of the two lines (Si and S) near $z = 1.88$ is significant, then one must also conclude that the detection of all five lines near $z = 2.75$ is equally significant. In addition, if one concludes that the marginal detection of the five lines near $z = 1.88$ is significant, then one must also conclude that the detection of two lines (Si and S) near $z = 1.2$ is equally significant.

Therefore, one cannot conclude simply that a complex of Ke V line emission is detected near $z = 1.88$; these data permit alternative interpretations of such complexes near $z = 1.2$ and $z = 2.75$. As the statistical excesses are due to the same ‘features’ in the observed spectrum, the interpretation of the statistical excess in the context of the model presented by R02 is degenerate and cannot be resolved with these data alone.

Prospects for confirmation of line features in GRBs are very good, considering that the X-ray spectral integration for GRB 011211 was begun 11 h after the GRB was initially detected, and required 1.4 h of integration to obtain. Decreasing the reaction time would permit a longer integration, while the afterglow is brighter in the X-rays, and the marginal results found here may well be improved upon.

ACKNOWLEDGMENTS

We are grateful to J. N. Reeves, who generously made his observed spectrum of the first 5 ks of the XMM–Newton observation of GRB 011211 available to us, so that we might independently analyse it. We gratefully acknowledge useful conversations with A. MacFadyen, R. Blandford and D. Fox. The authors are grateful to F. Harrison, F. Paerels and an anonymous referee for useful comments on the manuscript. MS was supported by NASA through Chandra Postdoctoral Fellowship Award Number PF1-20016 issued by the Chandra X-ray Observatory Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-39073.

REFERENCES

Arnaud K. A., 1996, in Jacoby G., Barnes J., eds, ASP Conf. Ser. Vol. 101, Astronomical Data Analysis Software and Systems V. Astron. Soc. Pac., San Francisco, p. 17
Bevington P. R., 1969, Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, New York
Borozdin K. N., Trudolyubov S. P., 2003, ApJ, 583, L57
Dahlem M., 1999, XMM Users’ Handbook, Issue 1.1, distributed by the XMM–Newton Science Operations Center, Vilspa
Kumar P., Narayan R., 2002, ApJ, submitted, astro-ph/0205488
Lazzati D., Ramirez-Ruiz E., Rees M. J., 2002, ApJ, 572, L57
Press W., Flannery B., Teukolsky S., Vetterling W., 1995, Numerical Recipes in C. Cambridge Univ. Press, Cambridge
Protassov R., van Dyk D. A., Connors A., Kashyap V. L., Siemiginowska A., 2002, ApJ, 571, 545
Reeves J. N. et al., 2002a, Nat, 416, 512 (R02)
Reeves J. N. et al., 2002b, A&A, submitted (astro-ph/0206480) (R02b)
Strüder L. et al., 2001, A&A, 365, L18

This paper has been typeset from a TeX/LaTeX file prepared by the author.