

EXPLORING THE INTEGRAL ARCHIVE

MATTHEW POWER

Period Determination of HMXBs

Physics and Astronomy
University of Southampton

April 2013

ABSTRACT

INTEGRAL, the International Gamma Ray Astrophysics Laboratory, is a satellite launched in 2002. Containing four primary instruments, INTEGRAL looks at sources in energy ranges approximately between 5 keV and 10 MeV, from hard X-rays to Gamma rays.

The INTEGRAL archive contains data for many sources that have not been fully explored. This project focuses on a sample of High Mass X-ray Binary sources thought to have eclipses. Lomb-Scargle spectral analysis and Phase Dispersion Minimisation are both used to find periodicities in the light curves of these sources.

CONTENTS

1	INTRODUCTION	1
1.1	The INTEGRAL Satellite	1
1.2	IBIS/ISGRI	1
1.3	Other Instruments on-board INTEGRAL	1
1.4	Coded Masks	2
2	THEORETICAL BACKGROUND	3
2.1	X-ray Binaries	3
2.2	Uhuru/SAS-A	3
2.3	Mass Transfer and the Roche Lobe	3
2.3.1	High Mass X-ray Binaries	3
2.3.2	Low Mass X-ray Binaries	4
2.4	Classes of HXMB	4
2.4.1	SGXBs	4
2.4.2	BeXBs	4
2.4.3	SFXTs	5
3	METHODOLOGY	6
3.1	Project Objectives	6
3.2	Test Source	6
3.3	Aims	6
3.4	Reading in Data	7
3.5	Filtering the Data	7
3.6	Re-binning the Data	8
3.7	Finding Periodicities	8
3.7.1	Lomb-Scargle Method	9
3.8	Folding the Lightcurve	9
3.8.1	Phase Dispersion Method	10
3.9	Error calculations of the Lomb-Scargle Period	11

INTRODUCTION

1.1 THE INTEGRAL SATELLITE

INTEGRAL, the International Gamma Ray Astrophysics Laboratory, is a satellite launched in 2002. Containing four primary instruments, INTEGRAL looks at sources in energy ranges approximately between 5 keV and 10 MeV, from hard X-rays to Gamma rays. The mission was launched by ESA, with support from NASA and the Russian Space Agency. The satellite was placed into a highly elliptical orbit around the Earth, with an orbital period 72 hours. This allows the instruments significant time making observations above the Earth's radiation belts, which would otherwise contribute to the noise of the signal. The satellite has been in operation for over ten years, providing some of the best observations available to astrophysicists.

1.2 IBIS/ISGRI

IBIS, the Imager on-board the INTEGRAL Satellite, provides the majority of the data used in this project. The instrument has two sets of detecting tiles placed one above the other, the first to detect hard X-rays and low energy gamma photons, the second to detect higher energy gamma photons. This project focussed on observations in X-rays, which utilises the lower range of light frequency sensitivity of IBIS from the first detector, ISGRI, INTEGRAL Soft Gamma Ray Imager.

The designs of ISGRI and PICsIT, its higher energy counterpart, are very similar. ISGRI uses an array of Cadmium Telluride tiles in a 128 by 128 array, and PICsIT has an array of Caesium Iodide tiles arranged 64 by 64. Using this dual detector design, IBIS is sensitive between 20 keV to 10 MeV. One additional benefit of this setup is that photons which register in both layers can be identified, which allows an improvement in the signal to noise ratio by discounting those unlikely to have come from a target source.

1.3 OTHER INSTRUMENTS ON-BOARD INTEGRAL

The Spectrometer on INTEGRAL, abbreviated SPI, is used to provide information on the energy incoming photons that IBIS cannot. SPI has an energy resolution of 2.2 keV at 1.33 MeV, approximately 0.2%, whilst INTEGRAL has a spectral resolution of between 8% to 10%. SPI uses a hexagonal arrangement of 19 germanium tiles, cooled down to 85 K. INTEGRAL is also equipped with

two additional X-ray imagers, called the Joint European X-Ray Monitor, or JEM-X. The two imagers are identical, and use a gas scintillator detector to measure photons in the 3 - 35 keV region. Finally, INTEGRAL is also equipped with an optical telescope, called the Optical Monitoring Camera, OMC, to provide measurements in tandem with the other instrumentation.

1.4 CODED MASKS

X-rays cannot be focused using conventional optical lenses, so the high energy instruments on-board INTEGRAL all make use of coded masks to form an image. Coded masks use a pattern of transparent and opaque elements in front of the sensor, and could be considered a sophisticated pinhole camera. Light entering a pinhole camera casts an image on the sensor, but greatly the pinhole greatly reduces the amount of light that can enter. Most coded masks have approximately two thirds of the mask opaque. As expected, light entering through a coded mask does not create an image, but creates a shadow based on the arrangement of the opaque and transparent elements. This shadow is the combination of all the images formed from each element.

The design of a coded mask allows the usage of mathematical algorithms applied to the sensor output, to construct an image, a process called deconvolution. Furthermore the design of coded masks is such that a source in the instrument's field of view will cast a unique shadow based on the orientation to the instrument. This means that multiple sources in the field of view can be separated, at a very fine angular resolution for an instrument of its type.

THEORETICAL BACKGROUND

2.1 X-RAY BINARIES

X-ray astronomy was confined to the study of our own Sun until the discovery of the first extra-solar in 1962 by Giacconi et. al. An X-ray detector was mounted in the head of a rocket, and launched into a high parabolic arc. The experimental team had discovered what would later be designated Scorpius X-1. Initially this discovery, and the discovery of other similar X-ray sources was puzzling to astronomers, who could account for how these stars were very luminous in X-rays, yet dim in optical wavelengths. If these were stars similar to our own, they should be significantly brighter in optical telescopes.

2.2 UHURU/SAS-A

The Uhuru/SAS-A satellite was the first mission to offer the insight that led to the modern field of X-ray astronomy. Operating in the X-ray band, it detected eclipses in the lightcurves of two sources, Her X-1 and Cen X-3. Whilst both sources have their differences, this result indicated that the systems are binary. Furthermore the orbital period and profile of the eclipses suggested that the stars were orbiting very close together. This then implied that the two stars could be interacting, with the possibility of mass transfer between the two. Also, one of the objects had to be compact. In the case of Her X-1 and Cen X-3, the compact object is a neutron star. We now understand that mass is transferred from a donor star to the compact object, accreting onto it. The initial mystery of how these sources were so luminous in X-rays is solved when one considers the energy transfer that is possible as the accreting material exchanges its gravitational potential energy for heat.

2.3 MASS TRANSFER AND THE ROCHE LOBE

X-ray Binary stars are broadly divided into two categories, those of a high mass companion star, and those of a low mass companion star.

2.3.1 *High Mass X-ray Binaries*

High Mass X-ray Binaries, HXMBs, have their mass transfer from donor star to compact object driven by a powerful stellar wind. Typically the donor star is an O or B type. The compact object accretes by simple infall as it ploughs

through the relatively dense stellar wind. HMXBs do have accretion disks, but tend to be smaller than those of low mass systems.

2.3.2 *Low Mass X-ray Binaries*

Conversely, Low Mass X-ray Binaries, LMXBs, have donor stars that are not sufficiently large or luminous to drive a powerful stellar wind. Thus the compact object has to accrete by a different mechanism. Consider the Binary system in a co-rotating frame of reference, such that the two stars appear stationary. Taking account of both centrifugal force and gravitational potential, the Roche Lobe is the local minimum or potential well where matter is gravitationally bound. Some stars extent is greater than their Roche Lobe, which allows matter to be transferred to the compact object by flowing over the L1 Lagrangian point. LMXBs tend to exhibit larger accretion disks than HMXBs and, due to the dynamics of the mass transfer, form a “figure of eight” shape.

2.4 CLASSES OF HXMB

This project focusses on a sample of HMXBs, which astrophysicists subdivide into three broad categories.

2.4.1 *SGXBs*

SGXBs, Supergiant X-ray Binaries have compact objects that spend their orbit in the strong stellar wind of an OB Supergiant. They are typically the most luminous subclass of HXMBs, often showing short drops in the X-ray flux observed where the compact object is eclipsed by its donor companion. Whilst the orbital eccentricity of the accreting object is lower than the other classes of HMXB, variations in the X-ray flux observed is affected by the orbital period of the accreting object, as it moves closer and further away from the supergiant.

2.4.2 *BeXBs*

Be star X-ray Binaries are, as the name suggests, associated with spectral type Be donor stars. Of significance to the behaviour of the X-ray flux is the fact that Be stars typically form equatorial disks of matter. This is thought to be aided by the tendency for Be stars to rotate rapidly. Furthermore, BeXB systems tend to have the neutron star in an eccentric orbit. As a result, the X-ray luminosity of these systems increases rapidly as the neutron star approaches it's periastron, when it intersects with this equatorial region of material surrounding the Be star, and is relatively quiet for the rest of its orbit.

2.4.3 *SFXTs*

Finally, Superfast X-ray Transients are a class that share common features with both SGXBs and BeXBs. SFXTs show outbursts similar BeXBs, but on a shorter timescale, and in addition are associated with the OB supergiants. One proposed explanation to explain the extremely short lived outbursts of these sources compared to the classical and persistent SGXBs is that the wind of these stars is inhomogeneous, and that the outbursts are caused by sudden accretion onto the neutron star as it moves through a “clumpy” patch of solar wind. This may not be a totally sufficient explanation, since SFXTs exhibit some eccentricity in the orbit of the neutron star, and it is thought that it may be a combination of these factors that leads to the large and short outbursts.

METHODOLOGY

3.1 PROJECT OBJECTIVES

The project aims to:

- Plot and visualise the light curves from the INTEGRAL data for the sample of sources.
- Visualise the orbital light curves for the sources.
- Compare the Superfast X-ray Transient sources with both Be star X-ray Binaries and Supergiant X-ray Binaries.
- Examine the eclipse profiles of those sources exhibiting them.

In order to achieve these aims it is necessary to have the tools to be able to produce a selection of results from the data. Chosen for familiarity and ease of use, Python was used as the programming language to produce the scripts and tools to analyse the data. The project used a standard build of Python 2.7, as well as additional libraries. Matplotlib was used to produce plots and graphs. Numpy and Scipy were used for their broad selection of tools that aid scientific computing.

3.2 TEST SOURCE

A large focus of the project was producing and testing the code for data analysis. Approximately half of the time spend during the year was spent testing and bug fixing. IGR J18027-2016 was selected as a test source. This source is a SGXB where the compact object is a neutron star that is accreting from a B giant. It is a particularly bright object, and also has a regular eclipse, which makes the source ideal for testing Lomb-Scargle period analysis, a key tool in the project.

3.3 AIMS

The code needs to perform the following:

- Read in the data from a text file.
- Filter the data.

- Re-bin the data into larger bins. Typically these will be approximately a day to a month. This then needs to be plotted.
- Find periodicities in the data using the Lomb-Scargle method and produce a periodogram.
- Additionally use Phase Dispersion Minimisation as a method of finding periods, and produce a graph of the result.
- Fold the light curve on a given period, and re-bin. Then plot a folded lightcurve for a source.
- Quantify errors for all these calculations.

3.4 READING IN DATA

The starting point for the workflow is a text file containing the data. Data points are lines in the file. The Numpy module provides a convenient function for reading from text files, `loadtxt`, which can be used to read a file into an array. The function can also “unpack” the data, outputting the data column-wise. It is changed to have the data as a list of rows, so that the entire data set can be iterated over by simply looping over all rows.

3.5 FILTERING THE DATA

When viewing data collected by the IBIS/ISGRI instrument, the data is typically very noisy. The INTEGRAL satellite has been operational for ten years, and so light curves for some sources are particularly good if the source is in an area of the sky that the satellite frequently points to. However some observations, science windows, have particularly large standard errors, often a magnitude larger than the flux reading. This can be caused by two features of the INTEGRAL satellite.

Firstly, as the INTEGRAL satellite sweeps through the plane of the sky, sources are observed that the satellite does not directly point at. When a source is observed where it is at the edge of the field of view of the instrumentation, it typically has a much higher standard error. This is a feature of the coded mask aperture that is used for imaging; they perform best when pointing directly at a source. Science windows where the source is measured with an off axis angle greater than 12° are removed.

Secondly, the exposure time for a typical science window is around 20 minutes. Sometimes the instruments will record an exposure that was much shorter, and these measurements also have a much higher standard error. Data points with an exposure time shorter than 500 s are removed.

Initially during testing and developing of the project code, when filtering was not implemented, the Lomb-Scargle period analysis had difficulty finding

the strong periodicity in the test source. This is because although other parts of the code output binned data, where an average weighted by the standard error on the flux has been used, the Lomb-Scargle routine does not use this, instead running on the raw lightcurve, and so extreme data points lower the visibility of the peak in the periodogram. For sources which show a weaker signal in their periodogram, this is crucial to being able to identify the correct peak to fold on.

3.6 RE-BINNING THE DATA

A simple tool for visualising the data is to be able to compile a graph of the entire light curve. However most sources from the archive have many thousands of individual science windows; the test source used had approximately 10,000. In order so that a sensible graph can be produced, the data shown needs to be cut down in a way that throws away as little useful information as possible. Re-binning collects groups of data points into a single data point.

In the project code, a uniform bin size is selected, typically between 1 day to 1 month. Data points are grouped into the bins, and a flux measurement for each bin is calculated by taking an average of the members, weighted by the standard error. Furthermore, a new standard error is calculated for each bin by propagating errors through the calculation of the binned flux.

Re-binning the data is useful for visualising the data. However, a graph produced by re-binning the data will not show any variation in the source on a time period smaller than the bin size. Binning into one month bins loses all information that happens in the timescale of the orbital period of the object, which may be only a few days. This makes re-binning only a preliminary step in analysing a source.

[fig1] [fig2]

Figure 1 shows a raw lightcurve for the test source. Note the large errors on many data points. Figure 2 shows a re-binned light curve for the test source, with a 1 month bin size. Note that when only a few data points fall into a single bin, the bin can appear to have a large error. The points in the figure with the smallest error bars are typically calculated from many data points in the raw lightcurve.

3.7 FINDING PERIODICITIES

One of the key aims of the project is to find periodicities in the sample of sources, so it is necessary to have robust method to detect this within the data. One common method employed is to perform Fourier analysis. In breaking down the data into a series of sine waves, the periodic variations in flux will show up as the sine waves with the largest coefficients. However, Fourier

analysis requires that the data points are evenly spaced, a criterion not often fulfilled in astronomy, and also not true of the INTEGRAL data.

3.7.1 *Lomb-Scargle Method*

One alternative method, that this project uses, is the Lomb-Scargle method. This method fits trial sine waves to the data and then by minimising the squares of the residuals, produces a spectrum of the data. Crucially, the Lomb-Scargle Method can work on unevenly sampled data. There are two easily available implementations of the Lomb-Scargle method in python. One is available in Scipy 0.11 and another has been implemented by a contributor to astropython, which this project uses [ref]. Both implementations utilise Numpy, which has routines that are pre-compiled in C and Fortran, making the routine efficient.

Two useful outputs are produced from this section of the code. The first is the periodogram, which is a spectrum of the data. The test source is known to have a strong periodicity associated with an eclipse. This shows up clearly in the periodogram. Second, the Lomb-Scargle routine returns the highest peak in the periodogram, which corresponds to the strongest period. This is passed to the next section of the code for producing the folded light curves.

Whilst both routines perform the same overall function, the inner workings are a “black box”. Experience showed that the astropython routine appeared to find a stronger peak, however there were exceptions where the Scipy implementation performed better. In general, the astropython implementation is shown, unless stated otherwise.

[Fig 3] [Fig 4]

Figure 3 shows a periodogram for the test source, using the astropython implementation. It shows a strong peak in power at approximately 4.6 days, which is due to the eclipse. Figure 4 shows a periodogram for the test source, using the astropython implementation.

3.8 FOLDING THE LIGHTCURVE

Using the period found from the Lomb-Scargle routine, the next step of the code is to produce a folded light curve. The light curve produced initially cannot reveal any significant amount of information about an object, since all variation is lost by re-binning the data. In the project code, by dividing each time coordinate by the period, and keeping only the remainder, the data is “folded” on top of itself. This means that each identical point in the object’s period will line up over each other.

The next step is to define a set of bins, similar to the previous section, and to re-bin the folded data. Once again, a weighted average of the flux is taken, and errors calculated through. The number of bins is a trade-off. With a smaller

number of bins, the averaging reduces the error on each measurement, however time resolution is lost. This can be seen in the test source, as each point of the eclipse occurs at the same place on the graph. Once averaged, the eclipse profile can be seen in the data, as the averaging process causes the points to converge to the true profile, and reduces the error greatly.

[Fig5]

Figure 5 shows the folded and binned light curve, using the strongest peak from the periodogram. Two cycles are shown for clarity. The eclipse shows through clearly, and the effect of the large data set reduces the size of the error bars significantly.

3.8.1 Phase Dispersion Method

Another means to detect periodicities in data is the Phase Dispersion Method. The Lomb-Scargle routine makes the implicit assumption that the signal in the data is sinusoidal, since all it fits are sine waves. This is a reasonable assumption when dealing with SGXBs such as the test source, and Cen X-3, since the eclipse is regular and the time of the eclipse forms a significant fraction of the folded light curve. However, for sources such as BeXBs where the “on time” of the sources is a very small fraction of the orbital period of the compact object, then this assumption breaks down. Sources that also have a high range, that flare particularly brightly also reduce the effectiveness of the Lomb-Scargle method. The Phase Dispersion Method makes no assumptions about the shape of the lightcurve. It works by attempting many trial folds of the data, and then binning the data down, where the process is the same that is used to produce the plot above. The key part of the method is that it compares the variance, or spread of points, within each bin to the variance of the entire data set as a whole. If a trial period is tested that is not a periodicity of the data, then the variance of points within a bin will on average be the same as the variance of the data as a whole. However if a trial period is tested that is a periodicity of the data, then the points within a bin will tend to cluster more tightly. This can be represented as a statistic for each trial fold, which tends to be ≈ 1 for noise, and then reduces when a periodicity is found.

[Fig 6]

Figure 6 shows a PDM plot for the same test source. Notice where the drop in the statistic coincides with the peak in the Lomb-Scargle periodogram.

Unfortunately, no PDM method already exists in any available python library, so this was coded from scratch. PAPER was used as a reference to code the method. This proved to be a technical challenge during the project. Furthermore, since the method is coded almost entirely in python, and due to the inherent nature of having to perform many trial folds of the data, the method is considerably slower to produce the same results as the Lomb-Scargle method.

3.9 ERROR CALCULATIONS OF THE LOMB-SCARGLE PERIOD

When dealing with more challenging sources, it's useful to be able to quantify how accurate the period found by the Lomb-Scargle routine is. It's difficult to calculate errors conventionally, and also challenging to implement this in code. Two solutions were coded to be used in the project to find an error measurement.

A Monte-Carlo simulation randomises the data. Each flux value is randomised according to a normal distribution with mean given by the flux and standard deviation given by its error. This generates a modified light curve which is processed by the Lomb-Scargle routine. The period is recorded, and the process repeated with a new randomised light curve generated from the original. Typically ten thousand runs are made as a minimum. A histogram of the periods is plotted, which should approximate a normal distribution with standard deviation that is the error on the original period.

Bootstrapping is another similar method. Rather than randomising the data, bootstrapping randomly discards data from the set. For N data points, N random integers between 1 and N are generated. Each integer corresponds to a data point, so that if the number X is generated, the X th data point is copied to a new light curve. This allows for duplicate random integers, however the data point is only copied once. This reduced light curve is run through the routine, the period recorded, and a new light curve generated to repeat the process.