Technische Universität München

Max-Planck-Institut für extraterrestrische Physik

High-Resolution Spectroscopy of Astrophysical Gamma-Ray Lines

Karsten Alexander Kretschmer

Kurzfassung

Hochaufgelöste Spektroskopie der Gamma-Linienemission radioaktiver Kerne erlaubt es, die Struktur und Kinematik des Zentralbereichs der Milchstraße zu untersuchen. Messungen des INTEGRAL-Satellitenobservatoriums der vom Zerfall des radioaktiven ^{26}Al erzeugten Gammastrahlung ergeben die Radialgeschwindigkeit des mit frischen Nukleosyntheseprodukten angereicherten interstellaren Mediums. Das von massereichen Sternen ausgestoßene Gas bewegt sich mit deutlich höheren Geschwindigkeiten als die kälteren Komponenten des ISM. Der Vergleich meiner Ergebnisse für die Ortsabhängigkeit der Geschwindigkeit entlang der galaktischen Ebene mit anderen Beobachtungen und theoretischen Vorhersagen deutet darauf hin, daß dieses heiße Gas aufgrund der Balkenstruktur unserer Galaxis diese ungewöhnliche Kinematik zeigt.

Abstract

High-resolution spectroscopy of gamma-ray line emission from radioactive nuclei allows studying the structure and kinematics of the inner regions of the Milky Way Galaxy. Measurements of the gamma rays produced in the deacy of radioactive ^{26}Al with the INTEGRAL space observatory yield the radial velocity of the interstellar medium enriched with fresh nucleosynthesis products. The gas ejected by massive stars moves with considerably higher velocities than the colder components of the ISM. Comparing my results for velocity as a function of position along the Galactic plane with other observations and theory suggests that this hot gas shows this ununsusal kinematics due to the barred structure of our Galaxy.

Contents

Contents

1 Astrophysics issues in our Galaxy

The Milky Way has been a familiar celestial feature since before antiquity. In those times, its visual appearance was probably even more familiar to the average person than today, when a large fraction of the population lives in cities where artificial illumination makes the features of the night sky ever harder to see with the naked eye.

Since more than two thousand years, the hypothesis that the Milky Way or "the Galaxy" is composed of a large number of individual stars has been speculated upon. It took the invention of the telescope, however for Galileo Galilei to be able to resolve the light from the band of the Milky Way into a large number of faint stars in 1609. This led to the idea, first published in 1750 by Thomas Wright and later expanded by Immanuel Kant, that it was a disc-shaped system held together by gravity, similar to the solar system but on a larger scale. William Herschel attempted to determine the extent of the disc by counting the stars brighter than a certain limiting apparent brightness visible in dependence on the direction in the plane of the Galaxy.

The effects of light absorption by interstellar dust severely restricted the effectiveness of this method. This changed when Harlow Shapley was able to measure the distances to the Milky Way's globular clusters in 1919. Located mostly outside the plane of the Galaxy and therefore only weakly affected by the interstellar dust, they allowed a much less biased determination of their spatial distribution. The position of the solar system located in the disc of the Milky Way, quite far from the centre, became apparent.

The first evidence that stars move was obtained by Edmund Halley in 1718 by comparing the positions of stars as seen in his time with historical accounts. Since then, both number and precision of known stellar proper motions have been increasing, particularly after the beginning of astrophotography allowed the side-by-side comparison of observations carried out at widely separated points in time. After Joseph Fraunhofer's discovery of absorption lines in the solar spectrum, absorption lines were also observed in the spectra of stars. It was then suggested to use the Doppler shift of these spectral lines as a function of the radial velocity to measure this motion along the line of sight. First results from this method were obtained around the turn of the twentieth century (Campbell 1901).

When it became possible to measure both the radial as well as the transverse velocities of stars with increasing precision, Jan Hendrik Oort was able to determine the kinematics of stars in the vicinity of the solar system and developed a theory for their dynamics (Oort 1928).

1.1 Structure of the Galaxy

Discerning the structure of our Galaxy is a complicated topic. Somewhat counter-intuitively, this is a result of our position inside the Galaxy, participating in its dynamics. On one hand, this position affords us a close-up view of events if they occur in a favourable location close to us, but on the other hand, getting an overview of the whole structure becomes more complicated because we are unable to get a bird's eye view of the whole structure. If the solar system was located in the halo of the Galaxy, far from the plane, we could observe the large-scale structure of the disk in one glance, but all of the smaller building blocks of the disk itself, such as star clusters and gas clouds would be at large distances.

What makes it difficult to discern the structure of the Milky way from within also makes it a worthy challenge that attracts the interest of many astronomers. This can be attributed to the linking of a diverse set of topics that need to be addressed in order to solve this puzzle.

Factors responsible for these complications are the ambiguities introduced by lines-of-sight close to the plane of the Galaxy and the effects of light absorption by intervening interstellar material. When we observe another spiral galaxy and the plane of its disk does not happen to be perpendicular (or nearly so) to the plane of the sky, then the dimensions of the region where the line of sight intersects the disk are small compared to the extent of the disk and the length scales relevant for its large scale structure. In our own Galaxy, a large fraction of interesting objects are located in the disk or in the central bulge, where observations are strongly affected by absorption of light by interstellar gas and dust. The attenuation is strongest at visual and ultraviolet wavelengths, where it can reach many orders of magnitude. This limitation has become less of an issue with the increased availability of near- and mid-infrared observations and sky surveys such as 2MASS and the GLIMPSE project using the Spitzer Space Telescope. At these longer wavelengths, the interstellar absorption is low enough to permit the study of the inner regions of the Galaxy which are completely obscured in the visible light regime.

The current picture of the structure of our own Galaxy is that of a barred spiral galaxy (Gerhard 2002, Benjamin et al. 2005). The outer part of the Galaxy consists of a disk composed mainly of gas and relatively young stars arranged into a spiral structure with either four main spiral arms or two major and two minor spiral arms. The inner part contains the bar, an elongated ellipsoid-like shape composed mainly of comparatively old stars and little gas compared to the disk. The spiral arms are distinguished from the intervening regions of space by having a higher density of stars and gas, the higher gas density leading to an enhanced rate of star formation within the arms. Because the angular velocity of the orbits of stars and gas in the Galactic potential is not constant as a function of radial distance from the centre, the spiral arms are not static features, but can be more accurately described as wave phenomena.

1.2 Galactic rotation

Every constituent part of the Galaxy can be characterised by a position in phase space consisting of a spatial position and a velocity vector. The position is usually stated in spherical coordinates relative to the solar system as a pair of angles such as Galactic longitude and latitude, and a distance. Correspondingly, the velocity vector decomposes into an angular proper motion perpendicular to the line of sight and a linear radial velocity along it. This customary stating of observational data has the advantage that it corresponds quite directly to the measured data such as the pointing direction of a telescope or the Doppler shift of a spectral line. As a result, the uncertainties in measured values are generally only weakly correlated — surveys almost always measure only a subset of the six phase space coordinates. A very influential kind of such surveys is the radial velocity survey, which measures either the radial velocity of objects as a function of the position on the sky or — if there are multiple target objects in a single line of sight, such as is the case when observing diffuse gas — the density of the target as a function of angular position and radial velocity. Another significant distinction between the different surveys that have been performed is whether a survey is targeted or not. In a targeted survey, a preselected set of objects is observed and their parameters measured. Otherwise, a selected area on the sky is sampled, preferably with a sufficient density of the sampled points to satisfy the Nyquist criterion, i.e. the separation between neighbouring sky samples should be less than half of the width of the angular response function of the observing instrument.

1.2.1 H$_\mathrm{I}$ surveys

Most of the interstellar gas in the Milky Way consists of hydrogen (about 74 % by mass) and helium (about 24 % by mass), the rest being made up by heavier elements. Depending on the local conditions such as density, temperature or radiation environment, this gas occurs in a variety of phases, wherein the majority of the interstellar gas mass is either in the form of neutral atomic hydrogen or in the form of the diatomic molecule H$_2$. H$_\mathrm{I}$ surveys are intended to measure the distribution of neutral, atomic hydrogen. This is achieved by measuring the intensity of the 1 420 MHz (corresponding to 21 cm) radio line associated with the hyperfine structure transition in the ground state of the hydrogen atom. Fig. 1.1 shows an all-sky map of this emission recorded with two radio telescopes, the 25 m Dwingeloo telescope in the Netherlands and the 30 m Villa Elisa telescope in Argentina. The beam size of a radio telescope of this size at a wavelength of 21 cm is about 0.5°. Because of the limited observation time available, the observations were performed with a pointing separation of 0.5°, somewhat more than would have been desirable from a Nyquist sampling perspective. For each of the points observed a spectrum was recorded that covered a 5 MHz wide frequency interval corresponding to a range of radial velocities from -450 km s^{-1} to $+400$ km s^{-1}. The map in 1.1 shows a frequency-integrated intensity implying that H$_\mathrm{I}$ emission is detected from all places
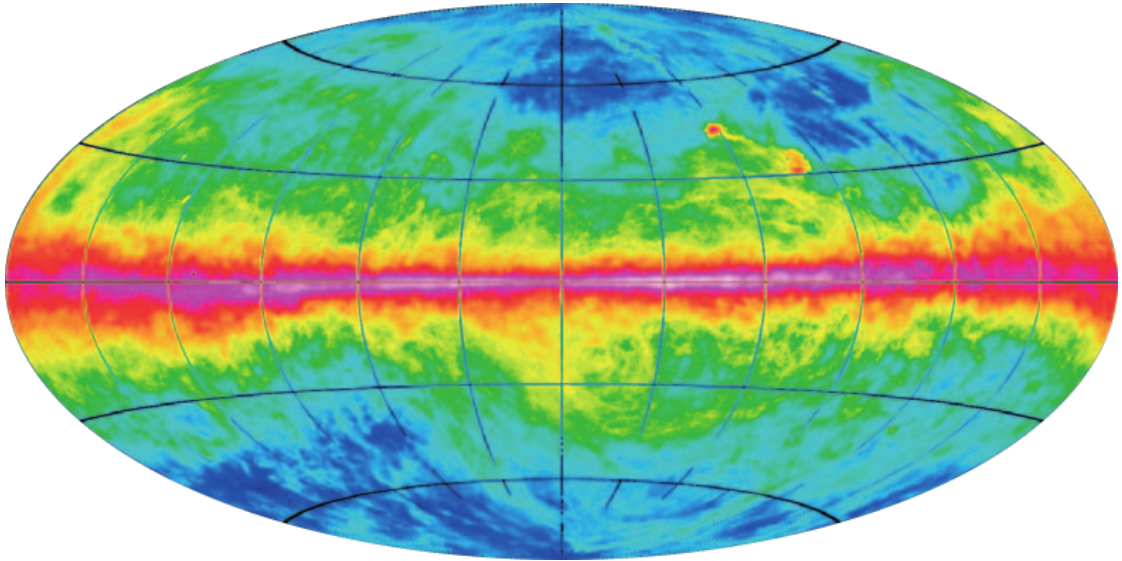
Figure 1.1: An all-sky map of the H_I emission in Galactic coordinates from the Leiden/Argentine/Bonn survey (Kalberla et al. 2005)

on the sky, predominantly however from the Galactic plane.

1.2.2 CO surveys
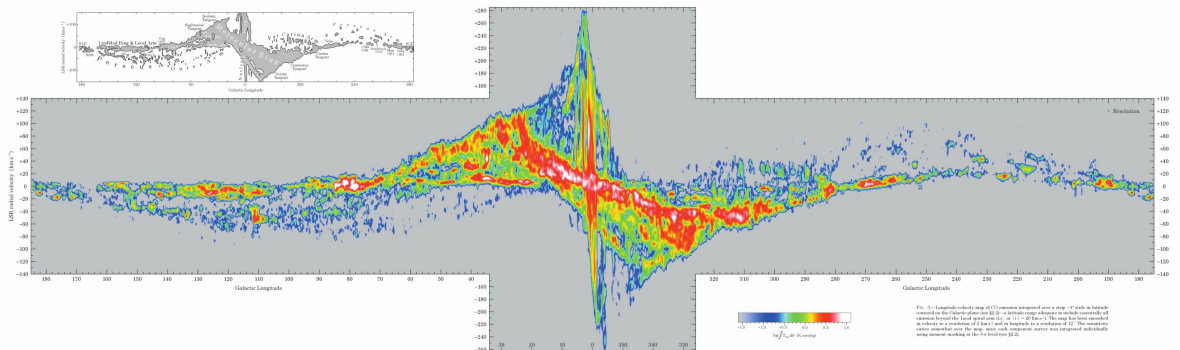
Figure 1.2: Diagram from the carbon monoxide survey of Dame et al. (2001) showing the distribution of molecular gas in radial velocity space as a function of its position along the Galactic plane

To measure the regions where stars are about to form in the near future or where the process of star formation is currently ongoing, it is necessary to look for clouds of comparatively cold gas. This gas consists largely of hydrogen in its molecular form H_2. However, the hydrogen molecule's spectral lines are impractical to observe.

As a symmetric molecule, it does not have an electric dipole moment and therefore no radiative transitions between rotational levels in the radio range. But molecular clouds in the Galaxy always contain an admixture of carbon monoxide, which has a rotational state transition in the millimetre wavelength regime, at a frequency of 115 GHz. By observing the trace gas CO and extrapolating to the bulk of H_2, the large-scale distribution of molecular gas in the Galaxy is surveyed.

1.2.3 H$_\alpha$ surveys
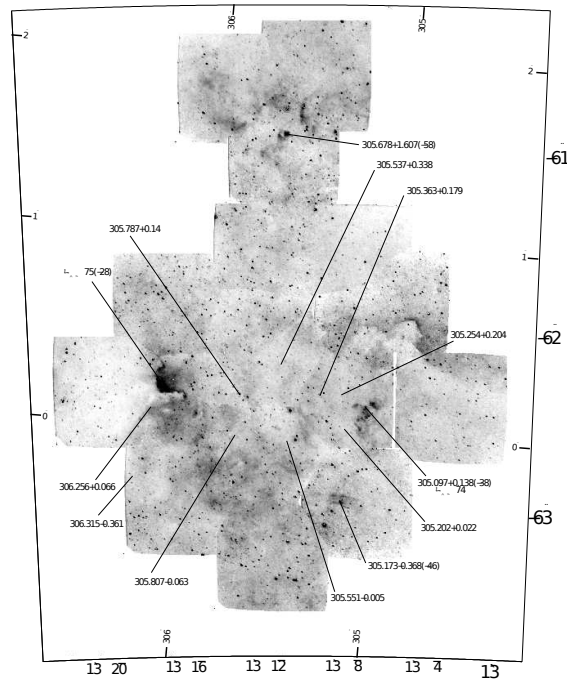
Figure 1.3: A composite of H$_\alpha$ observations in a region straddling the Galactic plane at a longitude of $\approx 305°$ (Russeil et al. 1998)

In addition to its atomic and molecular modifications, hydrogen is also detected in the Galaxy in ionised form, through the line emission associated with its recombination. Because of recombination, ionised hydrogen occurs only in the vicinity of intense enough energy sources capable of supplying the 13.6 eV per atom of ionisation energy. These conditions occur predominantly near hot stars, which ionise their surrounding gas via their extreme ultraviolet radiation. Other than energetic photons, which are also emitted by some accreting compact binary systems, collisional ionisation can occur in shock waves, e.g. in those generated by supernova explosions.

During recombination, the captured electron typically cascades through a number of energy levels of the hydrogen atom before reaching the ground state, emitting several of a large collection of possible lines in the process. Depending on their wavelength,

these lines lend themselves to different kinds of surveys. The well-known H_α line of the Balmer series at 656.3 nm lies in the red part of the visible spectrum and can be observed with optical telescopes. Surveys that are designed to cover a large solid angle on the sky and resolve the H_α line in radial velocity work with relatively small telescopes that have a large focal ratio, combined with an adjustable narrow-band filter such as a Fabry-Pérot interferometer (Russeil et al. 1998). This way, each single observation can cover a relatively large area on the sky at the expense of angular resolution, and be taken in a short exposure time.

1.2.4 Stellar surveys

Surveys of the motions of stars are generally targeted surveys aiming at a class of stars chosen by a particular set of selection criteria. The measurement of radial velocity requires the presence of a suitable spectral line in all desired targets and in a wavelength range that is accessible to observations. A type of source that has very useful characteristics for our goal of measuring the location and motion of sources in the inner part of the Galaxy are masers in the vicinity of red giant stars. A maser is a radio source whose emission mechanism is stimulated emission enabled by the presence of a population inversion in the upper and lower energy states of the transition involved. In astrophysical scenarios, masers are thought to be pumped either by radiation mainly in the infrared spectral range or by collisional excitation in shocks (Elitzur 1992).

Seed photons created by spontaneous emission can then, under the right conditions, stimulate the emission of more photons along their direction of propagation. "Right conditions" in this context means that the area where a population inversion occurs also has to have a sufficiently homogeneous velocity distribution, i.e. within the thermal broadening width of the line, to allow photons emitted in one part of the area to still be in resonance in a different part. Because of the amplification in a homogeneous medium would be exponential to the distance of propagation inside the medium, those sight lines with the longest distance can reach the highest intensities. Maser emission is therefore beamed and not isotropic. Maser activity has been observed from a number of different interstellar molecules such as the hydroxyl radical OH, ammonia NH_3, silicon monoxide SiO or methanol CH_3OH. The sources typically have narrow spectral lines and high intensities, making them well suited for radial velocity measurements.
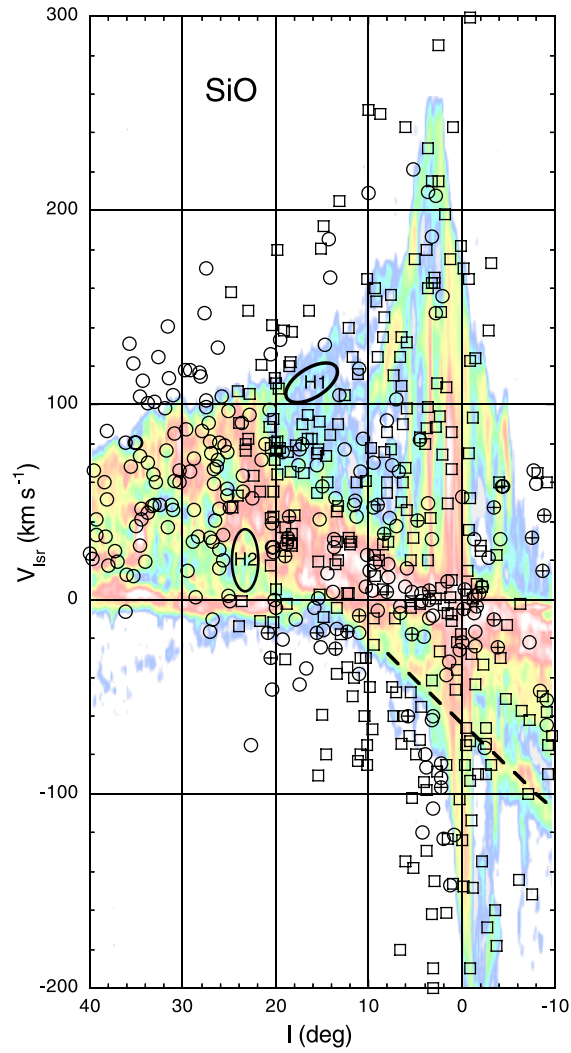
Figure 1.4: Longitudinal positions and radial velocities of silicon monoxide (SiO) maser sources in the galactic disk (Deguchi et al. 2004)

2 Sources of astrophysical gamma-ray lines

Because observations of the Galaxy, especially of the Galactic plane are strongly affected by the absorption of light in the interstellar medium, it is desirable to have access to wavelength bands where this interstellar absorption is low. Gamma rays in the MeV energy range are one of those bands, and therefore offer an opportunity to observe throughout the Galactic plane without the limitations incurred by having an inhomogeneous and poorly known "filter" between source and observer.

Gamma rays that reach the Earth from cosmic sources can be grouped into two categories: Continuum sources emit gamma rays with a spectrum that spans an energy interval with a width comparable to the average emitted photon energy and line sources, where emission is concentrated in a narrow energy interval. Sources of the first category are dominated by processes without sharply defined energy levels in the emitting system, such as interactions between free particles. This can be for example the emission of synchrotron radiation by fast moving charged particles in magnetic fields such as those occurring in the vicinity of neutron stars or interactions between free electrons and photons via the Compton or inverse Compton effects in the vicinity of black holes and the relativistic jets they can produce. Because neither the number of the free particles nor their interaction probabilities vary strongly over small energy intervals, this results in an equally flat continuum spectrum with extents of typically several orders of magnitude.

On the other hand, the second category mentioned above, the spectra with line features result from processes involving strongly bound systems such as nuclei, which have level transitions that can be extremely sharp. The sharpness of these transitions is illustrated by the precision measurements possible using the Mößbauer effect, where for example the 14.4 keV line of ^{57}Fe has a natural line width of $4.6 \cdot 10^{-9}$ eV, giving a ratio of line width to line energy of about $3 \cdot 10^{-12}$. Line emission or absorption can occur in nuclei, as well as in processes involving resonances in hadrons. Excited states of nuclei are either the result of a radioactive decay or of nuclear collisions. Nuclear collisions occur in regions where particle acceleration takes place, for example in solar flares. Gamma-ray line emission from solar flares has been measured by the CGRO/OSSE and INTEGRAL/SPI instruments, the RHESSI mission has spectroscopic observations of the sun in hard X-rays and gamma rays as its primary design goal. High resolution spectra of these lines have provided information about particle dynamics in solar flares.

2.1 Production of radioactive isotopes

Because their intensity is high enough to permit good measurements, the spectral lines in the gamma-ray regime that are most important for astronomical observations originate in the decay of radioactive nuclei and the annihilation of positrons. None of these seen today could have been produced in the early universe, on account of their atomic numbers being higher than those of any isotopes produced in big bang nucleosynthesis and their lifetimes being much shorter than the age of the universe. As a result, the radioactive isotopes that are currently observed through their decays must have been produced by recent processes. These involve either interactions with cosmic rays, such as the production of ^{14}C in the Earth's atmosphere or the high-temperature conditions found in the interiors of stars or the surface of accreting compact objects such as white dwarfs or neutron stars.

I will concentrate on the processes relevant to the production of ^{26}Al , an isotope whose presence in the Galactic interstellar medium was first detected via its gamma-ray emission with the HEAO-3 satellite (Mahoney et al. 1982). The isotope ^{26}Al, having one less neutron than the common stable isotope of Aluminium, is produced mainly by proton capture on magnesium in the reaction ^{25}Mg$(p, \gamma)^{26}$Al. For this reaction to occur, at a rate proportional to the concentrations of hydrogen and magnesium, the temperature has to be sufficiently high to allow proton capture by tunnelling of the Coulomb barrier in the energy range of the Gamow peak. For its decay to be observable, it must not happen inside a star, where the gamma-ray photon would be immediately absorbed. An understanding of the structure and development of stars is necessary to appreciate what conditions had to have been present in order to account for the ^{26}Al that is observed.

Briefly, ^{26}Al is produced when hydrogen-rich matter containing pre-seeded magnesium is exposed to temperatures on the order of 10^8 K. At too low temperatures, the production rate is too low, at too high temperatures, the rate of ^{26}Al consumption in further reactions is too high to allow significant amounts to be present. Besides allowing for production, the conditions also have to allow for the produced ^{26}Al to be transported from the production site to the interstellar medium at a time scale less than or comparable to the decay time constant of $\tau = 1$ Ma of the isotope. These conditions are fulfilled in massive stars, defined as being massive enough to have a supernova explosion at the end of their development process, translating to a lower mass limit of $\approx 8\ M_\odot$. This lower mass limit corresponds to an upper limit on the duration of the evolution of one of these stars of ≈ 40 Ma, a time scale that is important for the process of feedback. Feedback describes the series of events experienced by gas that is converted from an interstellar cloud into a star, undergoing a change in composition there, being released from the star in the form of stellar winds or the ejecta of a supernova explosion and then being mixed into the interstellar medium where the cycle then starts again. The rest of this chapter is devoted to examining these processes and the conditions necessary for them to happen, in greater detail.

2.2 Stellar evolution

2.2.1 Star formation

For a star to form, a cloud of interstellar gas must be compressed and heated so that the energy released by nuclear fusion reactions in its core can stabilise it against its self-gravity. This happens by gravitational contraction of gas clouds that have become unstable, i.e. their gravitational acceleration is larger than their stabilising pressure. The case where the supporting forces are solely due to thermal pressure is the easiest to model and was also the first studied. The mass that needs to be exceeded for a gas cloud of given temperature and density to become unstable toward gravitational collapse is termed *Jeans mass* after its discoverer J. H. Jeans, who proposed this relation more than a hundred years ago.

$$M > M_{\mathrm{J}} \approx \left(\frac{T}{50 \text{ K}} \right)^{3/2} \cdot \left(\frac{\rho}{10^{-23} \text{ g cm}^{-3}} \right)^{-1/2} \cdot 10^3 \ M_{\odot}$$

A cloud more massive than the Jeans mass was thought to contract while radiating away the potential energy liberated, essentially remaining isothermal. This process would complete on a free-fall time scale which however would predict a much larger rate of star formation than what is observed.

Since the 1950s, it has started to become clear that the pressure stabilising interstellar gas clouds is not purely thermal. Measurements of the polarisation of background starlight have shown the presence of magnetic fields (references in Mac Low and Klessen 2004, III.B.). The additional support provided by the magnetic field increases the critical mass necessary for gravitational collapse.

$$M > M_{\mathrm{cr}} \approx \left(\frac{n}{1\text{cm}^{-3}} \right)^{-2} \cdot \left(\frac{B}{0.3\text{nT}} \right)^{3} \cdot 4 \cdot 10^6 \ M_{\odot}$$

It seemed implausible that such a high cloud mass could be reached, meaning that the clouds were supported magnetically. Collapse would then occur on the time scale of diffusion of neutral gas through the ionised fraction supported by the magnetic field. The speed of this process termed *ambipolar diffusion* is determined by the fraction of gas that is ionised, as more ions generate a higher drag force opposing the collapse of the neutral gas. The ionisation fraction is in turn determined by the gas density which influences the amount of ultraviolet radiation that can penetrate the cloud, as well as the intensity of cosmic rays.

Further problems with the classical theory come from additional kinetic energy present that is not accounted for in the purely thermal picture. One aspect of this is the discrepancy between the angular momentum a cloud inherits from its participation in Galactic rotation and the observed rotation speeds of stars. Angular momentum conservation would preclude the collapse of a cloud to typical stellar densities as contraction would increase the rotation velocities to unphysically high values. It follows

that an efficient mechanism of angular momentum transfer must be active, slowing down the forming star. This has been supported by the observation of bipolar outflows from young stars from the 1980s on. If angular momentum gets transferred from the accreted matter to the ejected matter in the outflow, the excessive centrifugal support can be avoided. Another aspect is the observational evidence for random hypersonic motion inside molecular clouds seen with millimetre-wave measurements of the width of molecular lines (Zuckerman and Palmer 1974) which require an explanation for their persistence despite energy being dissipated by the shocks forming due to their interactions.

As radio and sub-millimetre wavelength observations improved, it became clear that there was another discrepancy. Theoretical predictions stated that molecular cloud cores should slowly contract until reaching critical density, then quickly collapse to a protostar inside a strongly peaked r^{-2} density profile. As a result, one expected relatively few cloud cores to contain protostars. The observations showed instead a large fraction of clouds with protostars embedded in them (Beichman et al. 1986). This means that processes faster than ambipolar diffusion must be at work during the early phases of star formation. Mac Low and Klessen (2004) argue that the turbulence inside the clouds is the determining factor. The turbulent processes inside molecular clouds are still not well understood, but there is agreement that the kinetic energy given off by stars – especially massive stars with their strong and fast stellar winds and the high-velocity ejecta of their supernova explosions – plays an important feedback role in determining the rate of star formation.

Once initiated, collapse of the cloud core proceeds rapidly on free-fall time-scales because most of the liberated gravitational energy is radiated away, keeping temperature and pressure low. With rising gas density in the centre of the collapsing cloud, the optical depth of the core increases, resulting in a reduction of the radiative energy loss and increased heating. The central object is now called a protostar. It is still increasing in mass because of the infall of gas onto its surface. This accretion process further compresses and heats the protostar from outside, as the accreted gas is decelerated in a shock where most of its kinetic energy is converted to heat and radiated away.

2.2.2 Main sequence

If the protostar accretes enough mass, meaning more than about 0.075 solar masses or $1.5 \cdot 10^{29}$ kg (Baraffe et al. 2002), the central temperature rises to a point where the fusion of hydrogen to helium releases enough heat to offset the radiative energy loss. Once the accretion of further gas from the collapsing cloud stops, either because all available gas has been accreted or because the rising luminosity of the young star has pushed the remaining gas away, the protostar settles into a nearly steady state and becomes a main-sequence star. In this state, changes to the stellar structure happen no longer on a time scale determined by the balance of gravitational contraction releasing binding energy and radiative cooling but on the much longer time scale of nuclear

energy inventory available to the burning processes to radiated power. The time scale determined by the quotient of luminosity L and gravitational potential energy E_P is called *Kelvin-Helmholtz* time scale.

$$\tau_\mathrm{KH} = -\frac{E_\mathrm{P}}{L} = \frac{G}{L} \int\limits_0^M \frac{M_r}{r}\, \mathrm{d}M_r$$

A rough estimate for the gravitational potential energy yields a duration of about $3 \cdot 10^7$ a for the sun. The time scale for nuclear energy release however is on the order of 10^{10} a. This means that once the star has reached the conditions necessary for the initiation of nuclear fusion, it changes very slowly. A single star spends most of its time in this state of quiescent hydrogen fusion, slowly converting the hydrogen available into helium and thereby tapping its nuclear energy store. As a result, most of the stars observed at any given point in time are seen while they are in their steady hydrogen burning phase. In this state, the luminosity and the surface temperature of a star are largely determined by its mass and change relatively little. As hydrogen is fused to helium in the core, the number of particles per unit mass decreases, which would reduce the pressure. A pressure reduction results in a readjustment of the core structure, leading to compression, increasing central temperature and density to maintain hydrostatic equilibrium. The increased temperature and density also increase the fusion rate and the released power. The luminosity increases roughly by a factor of 2.5 from the beginning to the end of the hydrogen burning phase, as can be seen from table 2.1. The outer layers of the star react to the increased output power of the core by expanding and possibly cooling slightly. While a sun-like star heats up slightly at its surface from 5 640 K to 5 790 K, for larger value of the stellar mass, the expansion during the main sequence results in a decrease in effective temperature.

Table 2.1: Luminosities and effective surface temperatures at the beginning and at the end of the main sequence

initial mass	luminosity		effective temperature	
	begin	end	begin	end
$M_\odot$	$L_\odot$	$L_\odot$	K	K
1	0.69	1.66	5 640	5 790
5	550	1 420	17 200	15 100
25	80 000	180 000	38 100	26 500

It is already visible from these values, that the luminosity of a star depends very strongly on the star's mass. A simple estimate of the dependence of the luminosity on the mass based on averaged values for radius, density, temperature (Weigert and Wendker 1996) leads to

$$L \sim M^3$$
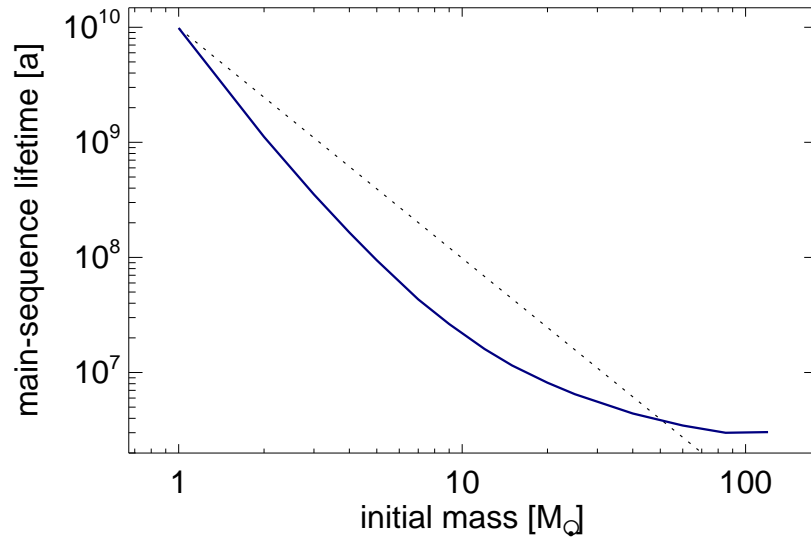
Figure 2.1: Main sequence lifetimes of stars as a function of their initial mass. (data from Schaller et al. 1992, Meynet et al. 1994) The dotted line shows a M^{-2} dependence for comparison.

As the total amount of nuclear energy that can possibly be released by fusion processes is proportional to the amount of fusion fuel available, it is proportional to the mass. This means that the time to exhaustion of the nuclear energy reservoir is approximately inversely proportional to the square of the mass. Fig. 2.1, which is based on numerical model calculations shows that in reality this relationship holds only approximately for stars with a mass equal to or larger than a solar mass, a power law being a bad approximation of the correct functional relationship. It overestimates the lifetimes of intermediate mass stars while underestimating the lifetimes of the most massive stars. Nevertheless, it is clear that the lifetime of a star is dramatically shorter the more massive the star is. While a star of a solar mass spends approximately 10^{10} a on the main sequence, this lifetime decreases to $9.5 \cdot 10^7$ a for a 5 $M_\odot$ star and to $6.5 \cdot 10^6$ a for a 25 $M_\odot$ star.

2.2.3 Post-main sequence evolution

Energy transport from the hot core with its maximum of heat production at the star's centre can be either by radiation or by convection. Convection happens if the temperature gradient necessary to transport the energy by radiation alone would be greater than the temperature gradient established by adiabatic expansion of rising gas. These conditions exist in the cores of stars more massive than 1.25–1.5 $M_\odot$ (Clayton 1983, Arnett 1996), where fusion is dominated by the CNO cycle. The proton
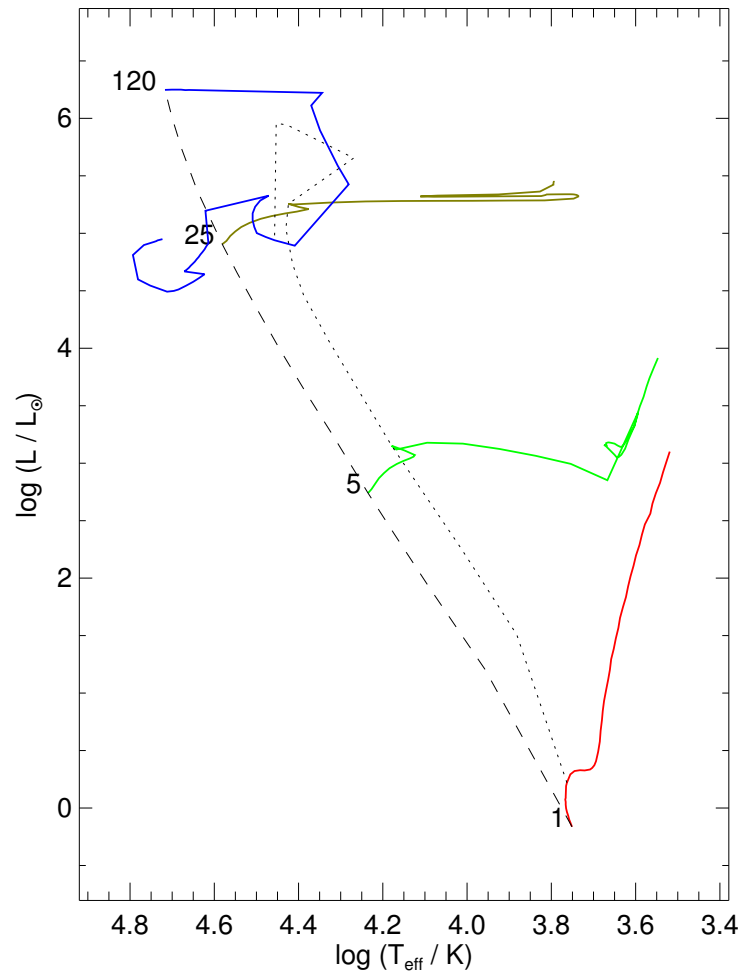
Figure 2.2: Hertzsprung-Russell diagram showing the trajectories in effective temperature vs. luminosity space of four stars with initial masses of 1, 5, 25 and 120 solar masses. The data come from the stellar evolution simulations of Schaller et al. (1992), Meynet et al. (1994)

captures on the heavier nuclei involved in the CNO cycle cause it to be more strongly temperature dependent than the reactions between hydrogen and helium nuclei that dominate energy release via the p-p chain. The strong temperature dependence causes the energy release to be more concentrated toward the core and therefore results in a high power flow per unit area. On the other hand, low-mass stars have a cooler core where, despite fusing via the p-p chain, radiative energy transport is inefficient compared to convection, causing the convection zone to extend through the whole volume of the star. The efficiency of radiative energy transport is strongly dependent on the temperature because of the Stefan-Boltzmann T^4 dependence of emitted power of a black body. As the net energy transport rate depends on the difference in emitted power from hotter towards colder layers, it is approximately proportional to T^3. The importance of the presence or absence of convection on stellar evolution comes from the mixing associated with it. The concentration of chemical species in a zone where convection is active is thereby kept very nearly homogeneous. This means that if the stellar core is convective, the concentration of hydrogen drops uniformly throughout the convective region. Despite the nuclear reaction rates being much higher in the centre, the macroscopic convective gas exchange redistributes fresh fuel from outside to the burning zone until the fuel is exhausted.

Once hydrogen in the stellar core is used up and the core therefore consists almost entirely of pure helium, energy generation in the centre stops, but it continues in the layers further out, where there is still hydrogen left. The burning of hydrogen in a shell around the helium core continues to add freshly produced helium to this core, compressing and heating it through increased density of helium over hydrogen. The increase in helium core mass added by the hydrogen burning in the shell above it causes it to shrink in radius. This shrinkage in turn causes a rise in gravitational acceleration at the core surface and with it an increase in temperature, density and therefore fusion rate of the hydrogen burning shell. As star's luminosity increases with time, its surface expands and cools as a reaction to the increased energy flow, which would otherwise require a higher temperature gradient than the original transport mechanisms can support (Clayton 1983, sec. 6-7). It becomes a red giant star, located at the low-temperature right side of the Hertzsprung-Russell diagram.

The further evolution of the star qualitatively depends on the star's mass. In very low mass stars, meaning $M < 0.5\ M_\odot$, further nuclear energy release after the fusion of hydrogen to helium might not be possible, because the the core temperature never reaches values high enough to further react helium into carbon. For these stars, the helium core becomes electron degenerate, i.e. supported mainly by Fermi pressure instead of ideal gas pressure. With no new source of heat available, the stellar core slowly contracts and cools until its internal energy consists entirely of the electron kinetic energy in the Fermi gas. This scenario is thought to be unlikely to happen in the Galaxy (D'Antona and Mazzitelli 1990). In more massive stars, the core heating through compression and heat conduction from the hydrogen burning shell eventually

reaches conditions where the rate of the triple-alpha process

$$\alpha + \alpha \rightleftharpoons {}^{8}\text{Be}$$
$$^{8}\text{Be} + \alpha \Rightarrow {}^{12}\text{C} + \gamma$$
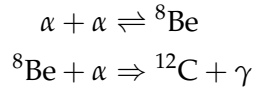
begins to become larger than the heat loss to neutrino emission. If this helium burning is initiated in an electron degenerate gas (in low-mass stars), the weak dependence of the gas pressure on temperature causes the strong positive feedback of temperature increase due to nuclear energy release without an accompanying negative feedback due to core expansion by increased pressure. The result is an extremely intense ($L \sim 10^{11}$ $L_{\odot}$) but brief *helium flash* that heats the core until it is no longer degenerate and negative feedback decreases the luminosity again and the core settles into steady helium burning. Still more massive stars ($M \leq 2.25$ $M_{\odot}$; Arnett 1996, sec. 8.2) directly move to core helium burning without forming a degenerate core first. This process is very fast, because it occurs on the thermal time scale of the core, not on the nuclear time scale of the hydrogen burning shell. Regardless of how exactly the process of helium burning was initiated, the star then again settles into steady state where helium fusion powers a core that is convecting because of the steep temperature dependence of the triple-alpha process. The convective core is surrounded by an inert helium shell that is steadily growing in mass because hydrogen burns in a shell on top of the helium core and the helium produced by this shell accumulates onto the helium core. In fact, hydrogen fusion typically accounts for 90 % of the nuclear energy release in this stage, only 10 % being provided by the helium burning core.

Helium burning proceeds at higher temperatures with a correspondingly large temperature gradient from the core outwards. The large luminosity required to support this leads to a high fuel consumption rate exacerbated by the smaller amount of energy available from helium fusion compared to hydrogen fusion. The energy available from fusing four hydrogen atoms into one helium atom is (Firestone et al. 1996)

$$
\begin{aligned}
\Delta_E &= 4 \cdot \Delta_{1_\text{H}} - \Delta_{4_\text{He}} \\
&= 4 \cdot 7\,388.969(10) \text{ keV} - 2\,424.911(10) \text{ keV} \\
&= 26\,730.965(41) \text{ keV}
\end{aligned}
$$

Here, $\Delta_{\text{A}_\text{X}}$ denotes the mass excess of isotope ${}^{\text{A}}\text{X}$ relative to the mass per nucleon in a ${}^{12}\text{C}$ atom.

$$\Delta_{\text{A}_\text{X}} = \left(m_{\text{A}_\text{X}} - \frac{A}{12} \cdot m_{12_\text{C}} \right) \cdot c^2$$

Table 2.2: Comparison of hydrogen- and helium-burning life times of stars depending on their initial mass. Data from Schaller et al. (1992), Meynet et al. (1994)

initial mass $[M_\odot]$	H-burning [a]	He-burning [a]
1	$9.84 \cdot 10^9$	$2.41 \cdot 10^9$
3	$3.53 \cdot 10^8$	$8.09 \cdot 10^7$
9	$2.64 \cdot 10^7$	$2.72 \cdot 10^6$
20	$8.16 \cdot 10^6$	$8.28 \cdot 10^5$
60	$3.46 \cdot 10^6$	$5.06 \cdot 10^5$
120	$3.04 \cdot 10^6$	$8.44 \cdot 10^5$

Not all of the nuclear binding energy is available as thermal energy in the star, a variable fraction of typically less than 1 MeV is carried away by neutrinos emitted in beta decays or electron captures in the different parts of the p-p chain or the CNO cycle. The exact amount depends on the rates of the various neutrino-producing reactions and therefore on temperature and density. In contrast, helium burning releases

$$\Delta_E = 3 \cdot \Delta_{4\text{He}} - \Delta_{12\text{C}}$$
$$= 3 \cdot 2\,424.911(10) \text{ keV} - 0 \text{ keV}$$
$$= 7\,274.733(30) \text{ keV}$$

This makes the energy release per nucleon and by extension per unit mass of fuel consumed

$$\frac{\Delta_E(\text{H} \to \text{He})}{\Delta_E(\text{He} \to \text{C})} = \frac{3 \cdot 26\,730.965(41) \text{ keV}}{7\,274.733(30) \text{ keV}} \approx 11$$

about an order of magnitude less for helium burning. As a result, this phase of stellar evolution is much shorter.

In contrast to hydrogen burning, helium burning does not produce a pure carbon core, a result of the triple-alpha process being a *three*-particle reaction. Its rate is strongly dependent on the concentration of ^{4}He. As this concentration dwindles to ever lower values towards helium exhaustion, the competing reaction

$$^{12}\text{C} + {}^4\text{He} \to {}^{16}\text{O} + \gamma$$
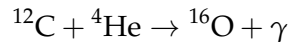
becomes more probable. The ^{12}C concentration peaks before the end of helium burning, afterwards, more carbon is converted into oxygen than is produced in the triple-alpha reaction. The specific amounts of ^{12}C and ^{16}O produced depend on the cross section of the ^{12}C(α, γ) reaction, which is difficult to measure under laboratory conditions. As more and more oxygen builds up, ^{16}O can itself react with another alpha particle into neon and magnesium.
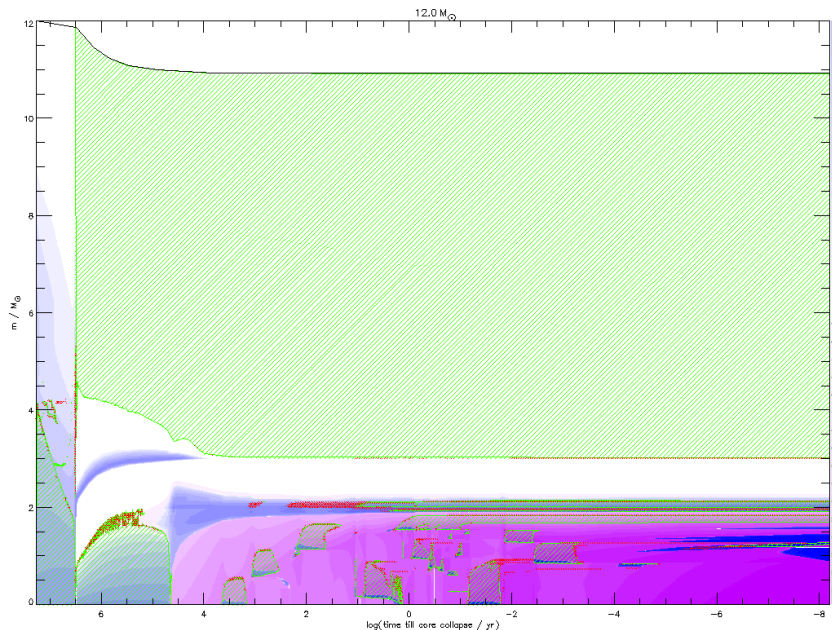
Figure 2.3: Kippenhahn diagram showing the evolution of the internal structure of a 12 $M_{\odot}$ star, from Woosley et al. (2002), Heger (2006)

Once helium is exhausted in the core, further evolution proceeds similarly to that following the end of hydrogen core burning. The core region that is without a heat source loses thermal energy, resulting in contraction with a corresponding rise of temperature and density. Helium continues to burn in a shell on top of the helium-exhausted core. However, an important difference is that after the end of helium burning, energy loss due to the emission of neutrinos is no longer a small effect as it is after hydrogen burning but it dominates the energy loss from the core (Arnett 1996, chapter 10). This is illustrated in figure 2.3, where areas with net energy gain are shown in shades of blue and those with net energy loss are shown in shades of purple. This effectively decouples the core thermally from the overlying layers. Again, the mass of the core decides the future evolution. Stars with low-mass carbon/oxygen cores do not reach the temperatures of $> 0.6 \cdot 10^9$ K required for carbon burning because they become supported by electron degeneracy pressure at lower central temperatures. They eventually lose most of their remaining hydrogen in the stellar envelope to the combined effects of hydrogen shell fusion and mass loss in stellar winds. They turn from red giants located on the asymptotic giant branch on the Hertzsprung-Russell diagram into white dwarfs consisting of an electron degenerate carbon/oxygen core surrounded by a thin hydrogen/helium atmosphere. More massive stars ignite carbon burning after the growth of the electron degenerate core and the most massive ones, with their initially lower central densities, can achieve carbon ignition without a significant contribution of degeneracy pressure.

Due to the lower Coulomb barrier for reactions involving the lowest-Z element present, carbon, reactions between two carbon-12 nuclei

$$^{12}\text{C} + {}^{12}\text{C} \rightarrow {}^{20}\text{Ne} + \alpha$$
$$^{12}\text{C} + {}^{12}\text{C} \rightarrow {}^{23}\text{Na} + p$$
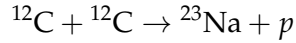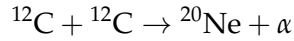
dominate. A reaction producing ^{24}Mg and a photon is also possible, but has low probability. The freed alpha particles react with oxygen,

$$^{16}\text{O} + \alpha \rightarrow {}^{20}\text{Ne} + \gamma$$
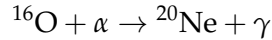
so that oxygen also contributes to neon production. After the exhaustion of carbon, the next nuclear reactions able to supply energy are neon-, oxygen- and silicon burning. Neon burning sets in at temperatures around $1.5 \cdot 10^9$ K and consumes the neon produced during carbon burning while producing ^{16}O, ^{24}Mg and ^{28}Si, increasing the amount of ^{16}O to values slightly higher than they were before carbon burning. Oxygen burning (at $T > 2 \cdot 10^9$ K) reacts ^{16}O into predominantly ^{28}Si and ^{32}S, but in addition, a significant number of beta decays or electron captures occur, which increase the neutron abundance. The fractional neutron excess $\eta = (N - A)/A$ increases by a factor of more than two, whereas it remained almost constant during carbon and neon burning. This is a higher neutron abundance than is measured in the isotopic make-up of solar system material. It is therefore unlikely that matter that went through oxygen burning is ejected into the interstellar medium, but probably immobilised inside a compact remnant at the end of stellar evolution. During silicon burning, which occurs at temperatures of $3 \cdot 10^9$–$4 \cdot 10^9$ K, even more β^+ decays and electron captures increase the neutron excess further. The bulk of the matter is converted into iron-peak isotopes, those located in the regions in the $Z - N$-plane with the largest binding energy per nucleon. The increases in density and temperature increase the neutrino heat loss rate strongly while the energy extracted by the nuclear fusion reactions per nucleon diminishes the more the iron peak is approached. As a consequence, the time spent in each successive burning phase decreases sharply from three centuries for carbon burning in the core of a 25 $M_\odot$ star over one year for neon burning, eight months for oxygen burning down to silicon burning, which requires only four days (Arnett 1996, chapter 10).

2.2.4 Core collapse

When further nuclear energy release in the core has become impossible, the heat lost to neutrino emission can only be supplied by gravitational energy from contraction. As the density increases, the pressure is dominated by the Fermi pressure of the degenerate electrons. An increase in Fermi pressure corresponds to an increase in the Fermi energy,

the energy of the highest populated quantum state in the approximation that thermal energy is much smaller than the Fermi energy. As the electrons with the highest energies become relativistic, the exponent γ of the equation of state

$$p \sim \rho^{\gamma}$$

describing the dependence of the pressure on temperature changes from its value of about 5/3 for a non-relativistic electron gas and approaches 4/3, the value valid for an ultra-relativistic Fermi gas. I this regime, the dependence of Fermi pressure on density approaches that of gravitational pressure. Were both exponents exactly equal, no amount of compression could increase the pressure sufficiently to counteract the gravitational attraction. This means that a star with a mildly relativistic degenerate core with an electron Fermi energy of ≈ 5 MeV (Arnett 1996, chapter 12) is only marginally stable against further contraction. As the Fermi energy rises with increasing density, it becomes more and more energetically favourable for a nucleus or a free proton to capture an electron from the Fermi sea, producing a more neutron-rich nucleus or a free neutron and an electron neutrino. The neutrinos have a very low interaction cross section with the surrounding matter and can leave the core nearly unimpeded. This free neutrino emission results in a low neutrino density and a correspondingly low Fermi energy. Once electron capture starts, it reduces the number of electrons present and accordingly causes a pressure drop that results in less resistance to further collapse, which accelerates as a result. A contributing factor is energy taken out of the thermal budget going into photo-dissociation of nuclei into free nucleons that sets in due to the increasing temperatures. Despite the electron gas being highly degenerate, the ion gas composed of the nuclei and free nucleons is not degenerate and is heated by compression. Once initiated, the collapse proceeds on a dynamical time scale, which is on the order of 100 ms, because the relative contribution of electron Fermi pressure becomes unimportant for the dynamics soon after the onset of collapse. This collapse can only be halted when the pressure is increased by contributions other than electron Fermi pressure becoming dominant. According to current understanding of the behaviour or matter at these conditions, it is a combination of neutron Fermi pressure and the repulsive component of the inter-nucleon interaction becoming dominant when the density approaches nuclear density. During the collapse, the inner part of the core is still partly in hydrostatic equilibrium, meaning the sound speed in these regions is fast enough to allow information about the changes in pressure to propagate across it in less time than the collapse requires. The radius of this part reduces from about 1 000 km to about 10 km. This releases an enormous amount of gravitational binding energy, which is on the order of 10^{46} J, corresponding to the energy equivalent of 5% of the mass of the sun. This energy is initially present predominantly in the form of kinetic energy of the bulk motion of matter. When the equation of state becomes stiffer, i.e. when the pressure increases faster with increasing density as nuclear densities are approached, the kinetic energy is converted into internal energy consisting of thermal

energy and a large amount of lepton Fermi energy. The Fermi energy of the leptons is still very large (> 100 MeV), because of a very important effect: neutrino trapping.

Whereas neutrinos can leave the core of a normal star with only an infinitesimally small probability of interaction, this changes drastically in the dense medium of a proto-neutron star. Under these conditions, the mean free path of a neutrino becomes much smaller than the radius of the collapsed core. As a result, neutrino energy diffuses outward via frequent scattering as well as absorption and re-emission, a process that takes much longer than it would take the neutrinos to escape directly at ultra-relativistic speed if there was no scattering. Once the neutrinos have diffused so far outwards that their mean free path, which rises with distance from the centre, becomes large enough to allow them to escape, they carry off most of the gravitational energy that became stored in the proto-neutron star. This cooling process by neutrino emission happens with a time scale of about 10 s, the theoretical prediction having been confirmed by the measurement of the neutrinos emitted during the supernova 1987A.

While the majority of the energy trapped in the core is radiated away by the neutrinos, not everything is. Around 1 % of the energy is converted into kinetic energy of the supernova ejecta. The mechanism driving this process that transports energy trapped in the core into the envelope is presently not yet conclusively identified. The candidates are neutrino scattering, acoustic waves and magnetic field interactions. Irrespective of the details of the process involved, the observations of supernova explosions imply that enough energy is transferred to the envelope to accelerate the fraction of the envelope outside a certain layer termed *mass cut*, to escape velocity.

2.3 Feedback

Supernova explosions return large amounts of matter that has undergone changes in its composition back to the interstellar medium. Stellar winds have a comparable effect, especially in luminous, massive stars. Besides the emission of light, stars also emit matter from their surfaces. These stellar winds are responsible for a mass loss that can have a large impact on the evolution of the star, compared to the evolution in the absence of mass loss. The mechanism powering the stellar winds depends on a variety of factors such as surface temperature, surface gravitational acceleration and chemical composition, Lamers (1998) provides an overview of the possible processes. In massive stars, the loss of gas in the form of stellar winds is due to light pressure on the gas atoms due to the many strong absorption lines of higher-Z elements, predominantly iron. This mechanism operates on the difference in momentum between light being absorbed by and later re-emitted by atoms or ions in the stellar vicinity. When an atom absorbs stellar radiation, the momentum of the absorbed photon gets transferred to the atom. Because the momentum distribution of the photon field is not isotropic, being that it carries away energy from the star, the absorption process imparts on average a radially outward directed momentum on the absorbing atom or ion. The re-emission

process, on the other hand, proceeds isotropically, imparting no net recoil momentum on the emitting atom. As a consequence, the scattering process transfers momentum from the photon field to the gas surrounding the star. The efficiency of this process depends on the number of resonance lines available for this process, which is usually dominated by the very large number of lines that iron ions have in the ultraviolet and far-ultraviolet spectral range. A result of the fact that line absorption is responsible for the acceleration of the wind is that the fraction of the starlight taking part in the wind acceleration process is dependent of the velocity range the wind covers. In a constant speed wind, absorption in lower layers would first deplete the photon field reaching the higher layers of those photons for which the absorption coefficient is largest. Because of Doppler shift of the absorption lines with increasing outflow velocity, in the rest frame of gas in higher layers, starlight is red-shifted more than that in lower layers. Light whose energy is just below resonance line in the gas at a given radius and so passes through unaffected, can be absorbed by an ion at a larger radius, thereby making a larger fraction of the star's optical output available to conversion into wind kinetic energy.

In a massive star, which has both a large convective core and a high mass loss rate through its stellar wind, matter that has undergone nuclear reactions in the core can be transported to the outer layers of the star via convection. The stellar wind can then transport material from these outer layers away from the star into the interstellar medium. This can be seen in Fig. 2.3: material in the mass coordinate range from 3 $M_\odot$ to 4 $M_\odot$ is part of the convective core in the early main sequence phase and part of the convective envelope during the red giant phase, where one solar mass of material is returned to the interstellar medium via the stellar wind.

In addition to the chemical enrichment that the material contained in stellar winds and supernova ejecta causes, a large amount of kinetic energy is carried in this gas. When high-velocity ejecta from stars encounter the surrounding interstellar medium, the dynamics of the ISM is strongly affected. Depending on the particular conditions, the pressure exerted when ejecta collide with molecular clouds can trigger the formation of stars or increase the amount of turbulence present in the cloud, stabilising it against gravitational collapse. Both effects probably occur in variable proportion, a process termed "feedback" that determines the pace of star formation.

3 Gamma-ray line shapes and source kinematics

The gamma rays we observe from the decay of a radioactive isotope are not mono-energetic but have a finite line width. Several processes contribute to the shape of the lines. A fundamental lower limit on the width is given by the uncertainty relation with respect to the lifetime of the excited state of the emitting nucleus.

$$\Delta E = \frac{\hbar}{\tau}$$

With typical values for the lifetimes on the order of picoseconds (see Table 3.1),

Table 3.1: Selected excitation states of nuclei, their energies and lifetimes relevant for gamma-ray lines of astrophysical interest. Values are from Firestone et al. (1996)

isotope	state	energy [keV]	lifetime	
^{26}Mg	2^+	1 808.7	476	fs
^{44}Sc	0^-	146.2	51.1	μs
	1^-	67.9	126.1	ns
^{44}Ca	2^+	1 157.0	2.61	ps
^{60}Co	2^+	58.6	10.47	min
^{60}Ni	4^+	2 505.8	0.30	ps
	2^+	1 332.5	0.713	ps

This corresponds to intrinsic line widths of 0.7 meV in the case of the lowest lifetime in the table, 476 fs. For longer lifetimes of the excited energy levels, the intrinsic width is still lower. This is so much below the energy resolution of all existing gamma-ray detectors that it can be neglected in practice.

Compton scattering can modify the energy of gamma ray photons, but for it to play a significant role, high column densities of intervening matter are required (see fig. 4.1). This means that in practice, Compton scattering is unimportant outside of high-density environments like a supernova envelope. Furthermore, for a Compton-scattered line to still be recognisable as a line instead of a continuum feature, the scattering angles have to be small, so that the energy transfer to the Compton electron is also small, further limiting the relevant interaction cross sections.

3.1 Doppler Effect

The Doppler effect is the most important process responsible for a measurable modification of line energies compared to the laboratory value. If the gamma-ray emitting nucleus is in motion relative to the observer, the Doppler effect will lead to a change in the observed energy and, correspondingly, the wavelength. If there is a range of particle velocities in the emitting material, for example appearing in our frame of reference as originating from a point source or orbiting around a common centre of mass, this will in general result in a variation in radial velocity and therefore in observed Doppler shifted frequency. In this section, I examine some specific scenarios and describe their distinguishing features. These scenarios are idealised model situations intended to give an overview of the most important individual processes that contribute to the overall spectral shape, a composite result of several different processes at work simultaneously.

3.1.1 Spherically symmetric expansion

In astrophysics, frequently structures dominated by a central mass or energy source are observed. These structures often show a spherically symmetric distribution of mass and velocity of the material around the centre, prominent examples for this behaviour are stellar winds and supernova explosions. If this material emits mono-energetic radiation, the Doppler shift caused by its velocity component along the line of sight of an observer will make the observer see a spectrum different from that of a stationary source.

The goal of this section is to analyse the spectrum observed from a source having both a spherically symmetric matter and velocity distribution. Symmetry requires a purely radial flow. Such a structure can be thought of as an assembly of infinitely many infinitesimally thin spherical shells having a homogeneous surface density and composition and a constant radial velocity with respect to the source centre. Each shell's mass and radial velocity are determined by the radial matter and velocity distributions of the source.

The first step to understanding the observed spectrum is therefore to determine the spectrum resulting from such a thin shell. I will begin with a distant source, where "distant" means that the shell's distance is much larger than its radius. Then, I will discuss the changes happening when the source has such a large angular extension on the sky that this approximation is no longer valid. Finally, I will show how the results obtained for a thin shell can be combined to provide the spectrum observed from a fully three-dimensional source.

Distant shell $(R \ll d)$

If the shell's diameter is much smaller than its distance from the observer, the viewing directions from the observer to different parts of the shell can be assumed to be parallel. Also, the distances from the observer to the front and back of the shell are nearly equal.

These approximations will be used in the following calculation.

Prerequisites

We choose a coordinate system so that the shell's centre is at the origin of the coordinate system $(0,0,0)$ and the observer's position is at $(d,0,0)$. The shell's radius as a function of time is $R = R(t)$, this yields the expansion velocity $\dot{R}$. Because $d \gg R$, parallel lines of sight can be assumed. If $\dot{R} \ll c$, the shell expansion during the light travel time through the shell can be neglected.
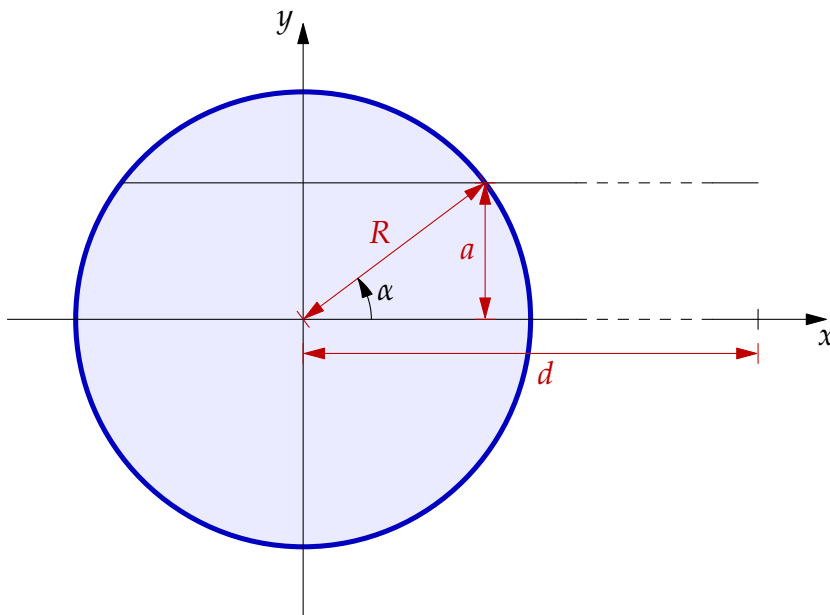

Figure 3.1: Geometry of a thin expanding shell $(d \gg R)$

Calculation of the line shape

We want to know the differential flux as a function of energy or frequency. To determine this, we have to calculate which fraction of the shell contributes to the emission at a given frequency interval. The frequency interval corresponds – via the Doppler effect – to an interval of radial velocity values. From the radial velocities we can determine the matching interval of angular or radial separation from the viewing axis. This separation in turn tells us the relevant shell surface area, which determines the flux in our frequency interval.

From this geometry, it follows for the relationship between the angle α and the radial

velocity v_R that

$$v_R = -\dot{R} \cos \alpha \quad \Rightarrow \cos \alpha = \frac{-v_R}{\dot{R}}$$

$$a = R \sin \alpha = R \sqrt{1 - \cos^2 \alpha}$$

$$= R \sqrt{1 - \left(\frac{v_R}{\dot{R}}\right)^2}$$

The Doppler effect yields, if the approximation $\dot{R} \ll c$ is justified,

$$\nu = \nu_0 \left(1 - \frac{v_R}{c}\right) \quad \Rightarrow v_R = c \left(1 - \frac{\nu}{\nu_0}\right)$$

$$\Rightarrow a = R \sqrt{1 - \frac{c^2 \left(1 - \frac{\nu}{\nu_0}\right)^2}{\dot{R}^2}}$$

$$\Rightarrow \frac{da}{d\nu} = \frac{R \cdot \frac{-c^2}{\dot{R}^2} \cdot 2\left(1 - \frac{\nu}{\nu_0}\right) \cdot \frac{-1}{\nu_0}}{2\sqrt{1 - \left(\frac{v_R}{\dot{R}}\right)^2}} = \frac{Rc^2 \left(1 - \frac{\nu}{\nu_0}\right)}{\dot{R}^2 \nu_0 \sqrt{1 - \left(\frac{v_R}{\dot{R}}\right)^2}}$$

This yields for dA, the surface area of the ring corresponding to da and $d\nu$

$$dA = \frac{A}{4\pi R^2} 2\pi a \frac{R}{\sqrt{R^2 - a^2}} da$$

$$\frac{R}{\sqrt{R^2 - a^2}} = \frac{R}{\sqrt{R^2 - R^2 \left(1 - \frac{c^2 \left(1 - \frac{\nu}{\nu_0}\right)^2}{\dot{R}^2}\right)}} = \frac{1}{\sqrt{\frac{c^2 \left(1 - \frac{\nu}{\nu_0}\right)^2}{\dot{R}^2}}} = \frac{\dot{R}}{c \left(1 - \frac{\nu}{\nu_0}\right)}$$

$$\Rightarrow dA = \frac{A}{4\pi R^2} 2\pi R \sqrt{} \frac{\dot{R}}{c \left(1 - \frac{\nu}{\nu_0}\right)} \frac{Rc^2 \left(1 - \frac{\nu}{\nu_0}\right)}{\dot{R}^2 \nu_0 \sqrt{}} = \frac{Ac}{2\dot{R}\nu_0} d\nu$$

$$dA = \Theta \left(|\dot{R}| - \left|c \left(\frac{\nu}{\nu_0} - 1\right)\right|\right) \cdot \frac{Ac}{2\dot{R}\nu_0} d\nu$$

$$\int_{\nu_0 \left(1 - \frac{\dot{R}}{c}\right)}^{\nu_0 \left(1 + \frac{\dot{R}}{c}\right)} d\nu \left(\frac{Ac}{2\dot{R}\nu_0}\right) = \frac{Ac}{2\dot{R}\nu_0} \left[\nu_0 \left(1 + \frac{\dot{R}}{c}\right) - \nu_0 \left(1 - \frac{\dot{R}}{c}\right)\right] = \frac{Ac}{2\dot{R}\nu_0} \left[\frac{2\nu_0 \dot{R}}{c}\right] = A$$

$$\boxed{dA = \Theta \left(|\dot{R}| - \left|c \left(\frac{\nu}{\nu_0} - 1\right)\right|\right) \cdot \frac{Ac}{2\dot{R}\nu_0} d\nu}$$

Conclusion: The line profile observed from a thin, spherical shell is rectangular, i.e. the spectral flux density is constant from the low energy end corresponding to the far side of the shell to the high energy end received from the point on the shell closest to the observer.

It is interesting to understand how this behaviour can result despite the fact that the high energy and low energy ends of the distribution come from those parts of the shell close to the poles, where little mass is located, compared to the sphere's equator, which is the source of unshifted radiation. It can be explained by looking at the radiation emitted by a ring on the shell at a given angular separation α from the axis and with a given width $d\alpha$. The ring's surface area is $2\pi R^2 \sin\alpha \, d\alpha$, it grows with α rising from 0 to $\pi/2$, increasing the emitted flux. However, with larger α, the energy is distributed over a larger frequency interval because the projection of the ring on the axis which is $d\alpha \sin\alpha$. The ratio of the flux to the width of the frequency interval is therefore independent of α.
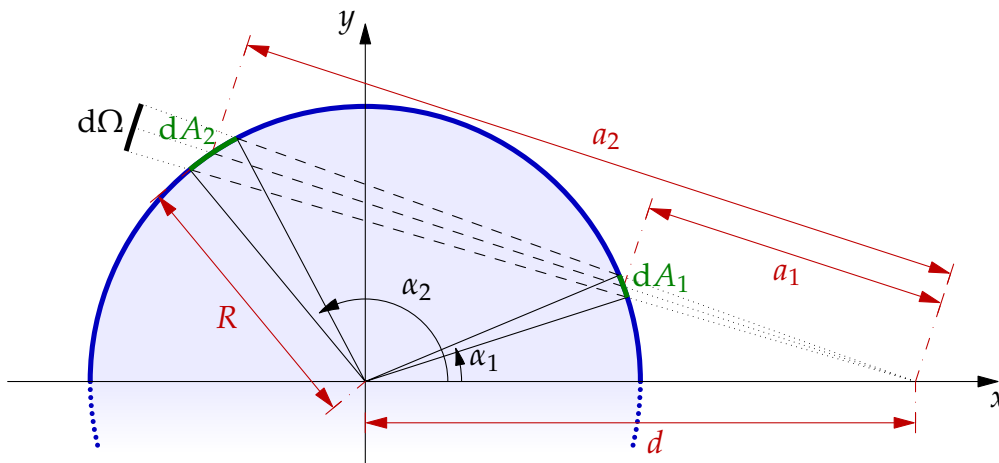
3.1.2 Close shell ($R \leq d$)

Figure 3.2: Geometry of a close shell

Qualitative behaviour

If we drop the assumption made for the distant shell, distance to the shell is no longer very large compared to its radius. This means that the side of the shell facing toward the observer and the side facing away from the observer no longer appear symmetric to each other.

From the cosine rule, we know that for the distance a between the observer and the shell, the following relationship holds

$$a^2 = d^2 + R^2 - 2dR\cos\alpha$$

therefore the radial velocity is

$$
\begin{aligned}
v_R = \dot{a} &= \frac{\mathrm{d}}{\mathrm{d}t}\sqrt{d^2 + R^2 - 2dR\cos\alpha} \\
&= \dot{R}\cdot\frac{R - d\cos\alpha}{\sqrt{d^2 + R^2 - 2dR\cos\alpha}}
\end{aligned}
$$

this gives the radial velocity interval corresponding to $\mathrm{d}\alpha$ as a function of α as

$$
\begin{aligned}
\frac{\mathrm{d}v_R}{\mathrm{d}\alpha} &= \dot{R}\cdot\frac{\mathrm{d}}{\mathrm{d}\alpha}\frac{R - d\cos\alpha}{\sqrt{d^2 + R^2 - 2dR\cos\alpha}} \\
&= \dot{R}\cdot\frac{(d\sin\alpha)\cdot\sqrt{X} - \frac{1}{2\sqrt{X}}(2dR\sin\alpha)(R - d\cos\alpha)}{X} \\
&= \dot{R}\cdot\frac{d^2\sin\alpha\cdot(d - R\cos\alpha)}{(d^2 + R^2 - 2dR\cos\alpha)^{3/2}}
\end{aligned}
$$

where $X = a^2 = d^2 + R^2 - 2dR\cos\alpha$. With $\dot{R} \ll c$, we obtain

$$
\begin{aligned}
\frac{\mathrm{d}\nu}{\mathrm{d}\alpha} &= \frac{\mathrm{d}}{\mathrm{d}\alpha}\left[\nu_0\cdot\left(1 - \frac{v_R}{c}\right)\right] \\
&= -\frac{\dot{R}\nu_0}{c}\cdot\frac{d^2\sin\alpha\cdot(d - R\cos\alpha)}{(d^2 + R^2 - 2dR\cos\alpha)^{3/2}}
\end{aligned}
$$

The flux density the observer sees from the surface element corresponding to $\mathrm{d}\alpha$ is

$$
\begin{aligned}
\mathrm{d}\Phi &= \frac{I}{4\pi R^2}\cdot(2\pi\cdot R\sin\alpha)\cdot R\,\mathrm{d}\alpha\cdot\frac{1}{4\pi a^2} \\
&= I\cdot\frac{\sin\alpha}{8\pi\cdot(d^2 + R^2 - 2dR\cos\alpha)}\,\mathrm{d}\alpha
\end{aligned}
$$

Therefore, the intensity per unit frequency interval as a function of α is

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\nu} = \frac{I \cdot \dfrac{\sin\alpha}{8\pi \cdot (d^2 + R^2 - 2dR\cos\alpha)}}{-\dfrac{\dot{R}\nu_0}{c} \cdot \dfrac{d^2 \sin\alpha \cdot (d - R\cos\alpha)}{(d^2 + R^2 - 2dR\cos\alpha)^{3/2}}}$$

$$= I \cdot \frac{\sin\alpha}{8\pi \cdot (d^2 + R^2 - 2dR\cos\alpha)} \cdot \frac{-c}{\dot{R}\nu_0} \cdot \frac{(d^2 + R^2 - 2dR\cos\alpha)^{3/2}}{d^2 \sin\alpha \cdot (d - R\cos\alpha)}$$

$$= \frac{-Ic}{8\pi\dot{R}\nu_0} \cdot \frac{\sqrt{d^2 + R^2 - 2dR\cos\alpha}}{d^2 \cdot (d - R\cos\alpha)}$$

$$\cos\alpha(v_\mathrm{R}) = \frac{-Rv_\mathrm{R}{}^2 + \dot{R}^2 R \pm \sqrt{-v_\mathrm{R}{}^2(-R^2 v_\mathrm{R}{}^2 + \dot{R}^2 R^2 - \dot{R}^2 d^2)}}{\dot{R}^2 d}$$

Let us consider some limiting cases of this result: $\alpha = 0$, $\alpha = \pi$ and $\alpha = \arccos R/d$, namely the points closest to and farthest away from us and the circle we see tangentially to the shell surface. For $\alpha = 0$ or $\alpha = \pi$, $\cos\alpha = \pm 1$, so

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\nu} = \frac{I}{4\pi d^2} \cdot \frac{c}{2\dot{R}} \cdot \frac{1}{\nu_0}$$

which is the same value as for a distant source. For $\cos\alpha = R/d$, we have

$$\frac{\mathrm{d}\Phi}{\mathrm{d}\nu} = \frac{Ic}{8\pi\dot{R}\nu_0} \cdot \frac{\sqrt{d^2 + R^2 - 2R^2}}{d \cdot (d^2 - R^2)}$$

$$= \frac{Ic}{8\pi\dot{R}\nu_0} \cdot \frac{1}{d\sqrt{d^2 - R^2}}$$

$$= \frac{Ic}{8\pi\dot{R}\nu_0} \cdot \frac{1}{da}$$

This flux value, which is observed at zero frequency shift, because of the purely tangential motion seen by the observer, is larger than the previous value by a factor of $d/a > 0$.

The symmetry of the resulting line shape can be understood qualitatively by examining figure 3.2. Along a given line of sight, a given solid angle element $\mathrm{d}\Omega$ intersects the shell surface in general in two places. Because $\mathrm{d}\Omega$ is small, the corresponding surface elements $\mathrm{d}A_1$ and $\mathrm{d}A_2$ can be approximated by two plane surfaces. Symmetry demands that these two surfaces have the same absolute inclination angle with respect to the line of sight, merely the sign of this angle is opposite. Because of this, the surface areas

$$\mathrm{d}A_i = a_i^2 \,\mathrm{d}\Omega$$

scale in such a way that the received intensity at the observer's position

$$\mathrm{d}I_i \sim \mathrm{d}A_i \cdot \frac{1}{a_i^2} = \text{const.}$$

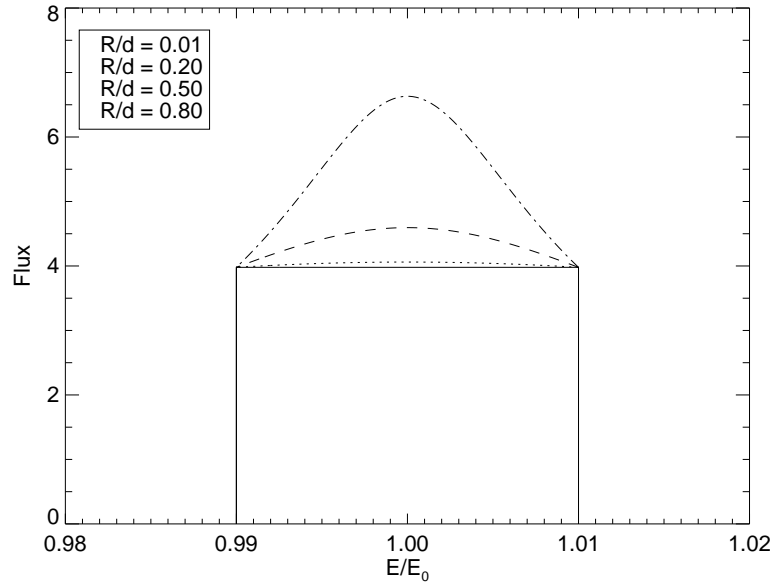is the same for both surface elements.

Figure 3.3: Spectra of shells for different values of R/d, $\dot{R} = 0.1\,c$

3.2 Doppler Effect from thermal motion

If a radioactive nucleus is part of a gas or plasma phase of the interstellar medium that is in thermal equilibrium, the particle velocities due to Brownian motion contribute to the formation of the line shape. In the case of typical interstellar medium conditions, with very large mean free paths and correspondingly unimportant inter-particle interactions, the nuclei behave like an ideal gas. In the absence of interactions, however, no thermal equilibrium state would be reached. For an equilibrium description of the situation to be applicable, a sufficient amount of particle interactions, e.g. via a sufficiently high density as found in star formation regions is necessary.

In thermal equilibrium, the particle energies are distributed according to Boltzmann's law: $N(E) \propto \exp(-E/k_\mathrm{B}T)$. In the absence of significant inter-particle interaction, the energy E is just the kinetic energy of the particle $E_\mathrm{kin} = 1/2 m \bar{v}^2$. Because E can be separated into three independent velocity components, $E_\mathrm{kin} = 1/2 m (v_x^2 + v_y^2 + v_z^2)$, $N(E)$ is a product of the three components:

$$N(E) \propto \prod_{i \in \{x,y,z\}} \exp\left(-\frac{v_i^2}{2mk_\mathrm{B}T}\right)$$

Because the velocity components perpendicular to the line of sight have (to the first approximation) no influence on the Doppler shift, it is sufficient to consider the distribution of the velocity component along the line of sight, v_x. The particle number as a function of v_x is therefore proportional to $\exp(-v_x^2/2mk_\mathrm{B}T)$, which is a Gaussian

distribution centred around zero velocity along the line of sight, with a standard deviation of:

$$\sigma_{v_x} = \sqrt{\frac{k_B T}{m}}$$

The observed frequency ν of radiation emitted by a particle at velocity v_x with a laboratory frequency ν_0 is

$$\nu = \nu_0 \cdot \left(1 - \frac{v_x}{c}\right)$$

this is linear in v_x, causing the shape of the particle number as a function of observed frequency and consequently also the observed spectrum to be Gaussian. The Gaussian is centred around ν_0 and has a standard deviation of

$$\sigma_\nu = \nu_0 \cdot \frac{\sigma_{v_x}}{c} = \frac{\nu_0}{c} \cdot \sqrt{\frac{k_B T}{m}}$$

which corresponds to a full with at half maximum of

$$\text{FWHM}_\nu = \sigma_\nu \cdot 2\sqrt{2\ln 2} = \nu_0 \cdot \frac{2}{c}\sqrt{\frac{2k_B T \ln 2}{m}}$$

These relations allow the comparison of the conditions necessary to give rise to a certain measured line width. A Gaussian line profile is not only useful because it is the spectral shape of the emission of an ideal gas, but it is also useful in comparisons because the Gaussian probability distribution arises from a maximum entropy argument when only the first and second moments, that is the position and the width of a distribution are known (e.g. Jaynes 2003, ch. 7). Solving for the gas temperature given a measured value of the full width at half maximum of the observed line, we obtain

$$T = \frac{(\text{FWHM}_E / E_0 \cdot c)^2 \cdot m}{8 k_B \ln 2}$$

For getting an impression of the temperatures required to account for line widths observed to the ^{26}Al emission from the inner galaxy, inserting numbers yields

$$T \approx 15.5 \cdot 10^5 \text{ K} \left(\frac{\text{FWHM}_E}{1 \text{ keV}}\right)^2 \left(\frac{m}{26\,u}\right)$$

Gas in the interstellar medium of Hɪɪ regions is heated by the absorption of extreme ultraviolet radiation from hot stars until an equilibrium is reached with the cooling due to emission mainly by C, N, O and Ne atoms excited by inelastic collisions in the gas (Weigert and Wendker 1996, Spitzer 1978).

The temperature of around 15 MK required to account for a 1 keV broadening of the ^{26}Al line at 1 808.67 keV is very high when compared to typical temperatures in the interstellar medium, which is around 10^4 K in Hɪɪ regions and around 10^2 K in Hɪ regions. For making comparisons with non-thermal environments, the particle

velocity is more interesting than the temperature. In an ideal gas, the kinetic velocity corresponding to the average particle energy is the square root of the average square of the individual particle velocities:

$$\sqrt{\overline{v^2}} = \sqrt{\frac{3k_{\mathrm{B}}T}{m}} = \frac{\sqrt{3}\,c \cdot \mathrm{FWHM}_E}{2\sqrt{2\ln 2} \cdot E_0}$$

When the comparison value of 1 keV FWHM is used, the root-mean-square velocity as a function of line width is

$$\sqrt{\overline{v^2}} \approx 122 \text{ km s}^{-1} \left(\frac{\mathrm{FWHM}_E}{1 \text{ keV}} \right)$$

3.3 Doppler Effect from Galactic rotation

3.3.1 Geometry

We observe the Galaxy, with the overall velocity distribution of its constituents stars, gas, etc. from the inside. This means that the motion of an object within the Galaxy relative to us depend on the object's trajectory, for example a star's orbit in the Galactic gravitational potential as well as on the orbital motion of the solar system which is of course itself a component of the Galaxy. The orbits of stars within the Galaxy are mainly determined by the large-scale shape of the gravitational potential, occasional close encounters between individual stars or stars and gas clouds cause scattering of orbits. The large-scale potential in this case is not dominated by a central point-like mass but by the combined gravitational field of the dark matter halo of the Galaxy, its gas and its stars. The shape of this potential is not of the Coulomb $1/r$ type or the harmonic r^2 type, the only ones with closed orbits in the general case. This means that orbits in the Galactic potential are not closed elliptic curves, but precess quickly around the axis of the Galaxy.

   Measurements of the motion of stars in the vicinity of the solar system (Holmberg et al. 2007) have shown that the majority of them have very similar velocity vectors and the density of stars in velocity space has a clearly defined maximum that is close to the circular orbit velocity. This means it is justified to talk about a characteristic velocity of the local stars. Assuming this is also the case over most of the Galactic disc, the concept of rotational velocity as a function of galactocentric radius is sensible.

   The rotation curve, i.e. the average orbital velocity as a function of galactocentric radius is relatively straightforward to measure when observing another galaxy , provided the line of sight is not perpendicular to the galaxy's disc. In our own Galaxy, however this is by far more difficult because we observe from inside the rotating structure from a vantage point that is itself in orbit in the Galactic gravitational potential. This is further complicated by the fact that observations in the optical range are limited by dust absorption in the Galactic plane and that the strongest atomic and molecular transition

lines in the radio wavelength range become optically thick in parts of the Galactic disc. The two most important lines in this category are the 21 cm, 1 420 MHz H I hyperfine transition of atomic hydrogen and the 2.6 mm, 115 GHz CO(1-0) rotational transition of the carbon monoxide molecule. Dame et al. (2001) show how the optical thickness limitation can be sidestepped to a certain degree by observing the transitions of the rare and therefore much less optically thick ^{13}CO molecule instead of the predominant ^{12}CO.

Sofue et al. (2009) combined the available observational data, also including the 656 nm H_α line emitted by ionised hydrogen in H II regions surrounding luminous hot stars to model the rotation curve of the Galaxy by varying a parametric model of the mass distribution from which the rotational velocity of a circular orbit as a function of galactocentric radius follows through Newton's laws. The data and the corresponding models are plotted in Fig. 3.4.
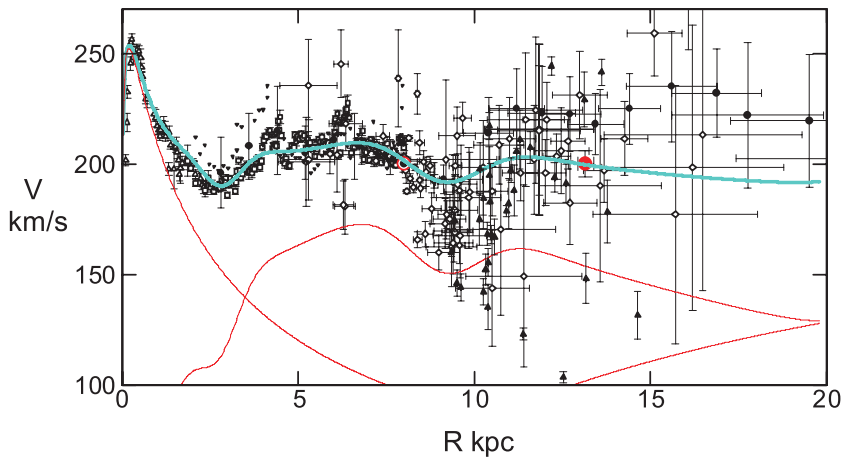
Figure 3.4: Rotation curve of the Galaxy from Sofue et al. (2009)

The rotation velocity as a function of galactocentric radius together with the velocity of the solar system with respect to the mean motion of the stars in its vicinity, the local standard of rest (LSR, Francis and Anderson 2009) allows calculating the radial velocity of a source as a function of source position in the Galactic plane. The result is shown in Fig. 3.5. The radial velocity values span an interval of $v_r \in [-240; 220]$ km s^{-1}, where the highest values of $|v_r| > 150$ km s^{-1} are reached for $|l| < 30°$.

The effect this kinematic behaviour has on the observed energy distribution of a line-emitting source depends on whether individual sources can be observed or not. In the case of individual sources, the situation is comparable to the maser sources mentioned in Sec. 1.2.4. In the case of unresolved sources, the observation consists of a continuous function of intensity as a function of both angular position and radial velocity. This continuous joint distribution in a 3-D parameter space is then sampled by the observing process. Figure 3.6 shows the result of a Monte-Carlo simulation reproducing the
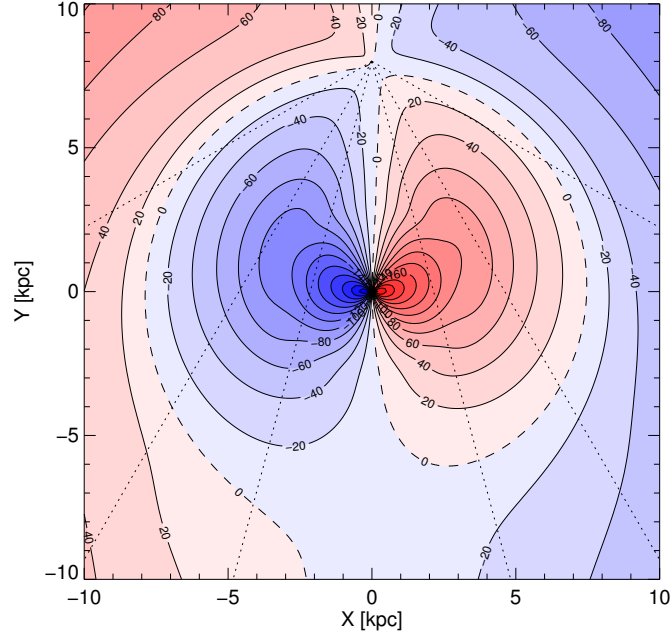
Figure 3.5: Radial velocity in km s^{-1} for circular orbits following the rotation curve of Sofue et al. (2009) and the LSR from Francis and Anderson (2009). The dotted lines mark galactic longitudes of $\pm15°$, $\pm30°$ and $\pm60°$.

spectral energy distribution of a sample of sources distributed spatially according to an exponential disk density distribution ($R_0 = 4$ kpc, $z_0 = 180$ pc). The sources are assumed to orbit on circular trajectories and consist of a spherically expanding shell each (see Sec. 3.1.1), modelling the expansion behaviour of the dominant ^{26}Al sources, stellar winds and supernova explosions. The resulting diagram is structurally similar to the CO *l-v* diagram (Fig. 1.2), showing blue-shift at negative Galactic longitudes in the fourth quadrant and red-shift in the first quadrant ($\Delta_E = +1$ keV corresponds to $v_r = -166$ km s^{-1}).

When the resulting intensity distribution is observed by integrating the emission over a macroscopic solid angle on the sky, spectra as shown in Fig. 3.7 are obtained. This figure shows the spectra corresponding to three rectangular areas towards the Galactic centre and extending to the east as well as to the west along the Galactic plane, for comparison, the dash-dotted lines show the model spectra after being convolved with a Gaussian of 3.0 keV FWHM, corresponding to the typical energy resolution of a germanium detector at the energy of the ^{26}Al line. The next chapter then deals with the implication this has for the measurement of the Galactic ^{26}Al emission in practice.
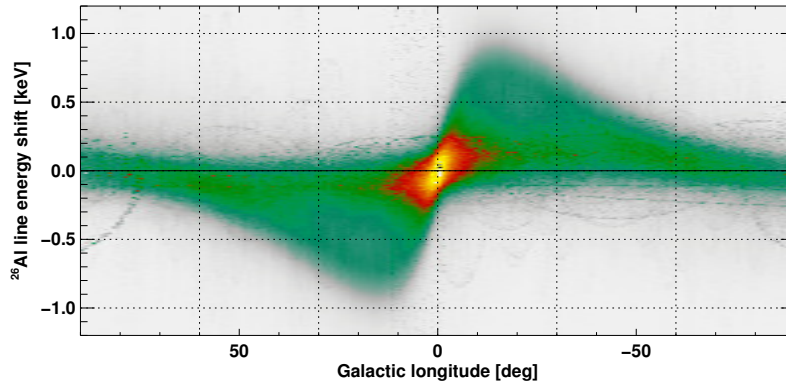
Figure 3.6: Doppler shift vs. longitude diagram of an exponential disk model of the Galaxy with parameters $R_0 = 4$ kpc, $z_0 = 180$ pc, $E_0 = 1\,808.63$ keV
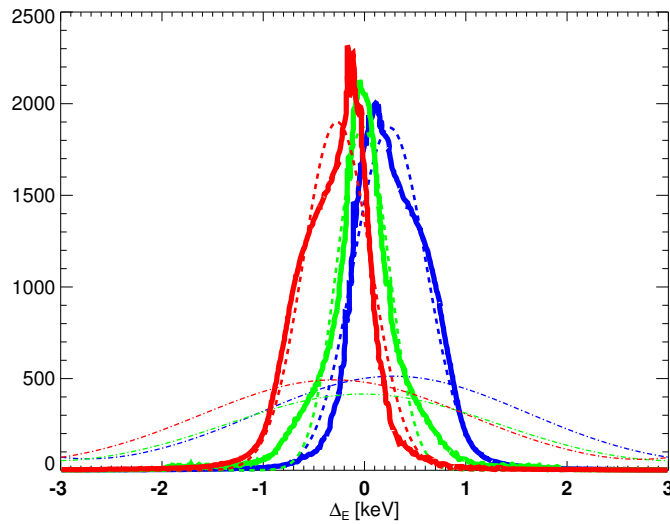

Figure 3.7: Spectra from an exponential disk model (see Fig. 3.6) integrated over three sky areas: $b \in [-10°, 10°]$, $l \in [-40°, -10°]$ (blue), $l \in [-10°, -10°]$ (green), $l \in [10°, 40°]$ (red). The different line styles show: model spectrum (solid), best Gaussian approximation (dashed), convolution with 3.0 keV FWHM (dash-dotted).

4 Detection of gamma rays with SPI

4.1 Physical limitations

Any attempt to detect gamma rays from astrophysical sources in the energy range where nuclear lines occur, i.e. roughly from 50 keV to 10 MeV, corresponding to wavelengths from about 25 pm to 120 fm has to deal with the peculiarities of the interaction between electromagnetic radiation and matter in this part of the spectrum.
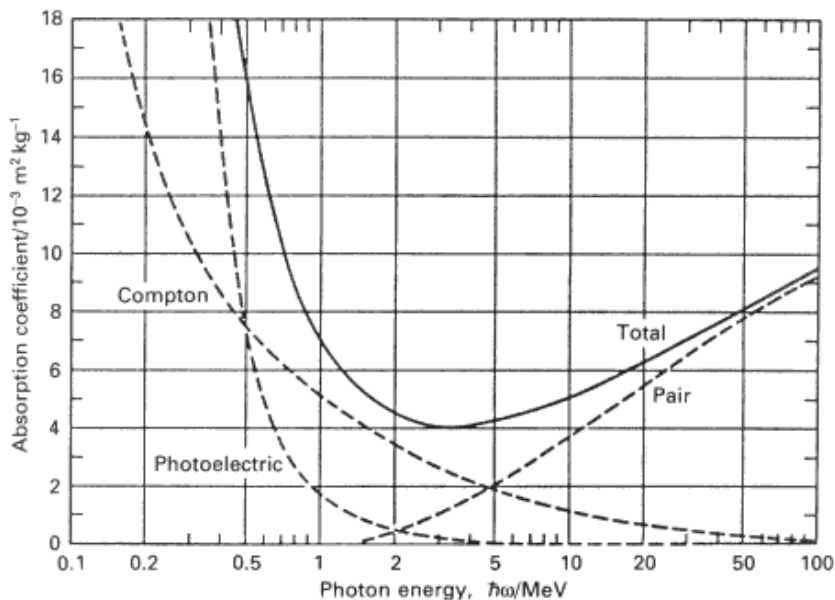

Figure 4.1: Absorption coefficient of lead in the energy range 100 keV-10 MeV (from Longair 1994)

As visible in fig. 4.1, the absorption coefficient of electromagnetic radiation reaches a minimum in the MeV region of the spectrum. The lowest value attained is around $4 \cdot 10^{-3}$ $m^{-2}kg^{-1}$. Together with the density of lead, $1.134 \cdot 10^4$ kg m^{-3} at standard conditions, $\kappa\rho \approx 45$ m^{-1}, which means that in order to absorb 90 % of the incident gamma flux, about 5 cm of lead are needed. The detectors need to be fairly thick compared to those in other wavelength ranges. This high material requirement is compounded by the unavailability of focusing optics, which exist at the time of writing only on paper (Skinner et al. 2004) or as prototypes (von Ballmoos et al. 2004). All long-

duration gamma-ray telescopes to date have had to rely on collimation of the incident radiation or on the tracking of the reaction partners in particle interactions like the Compton and pair creation effects. These methods have in common that the effective area of the instrument cannot exceed the geometric area of the detector and is usually much less. The large area combined with the large thickness directly results in a large mass of detector material. These, in turn, present large targets for the interaction with cosmic rays that are encountered above the Earth's atmosphere, where the observations must be done to avoid absorbing the gamma rays in air. The interactions with energetic particles in the detector material itself as well as in the surrounding spacecraft cause prompt signals in the detector and also activate nuclei leading to later radioactive decays again causing background. A key to interpreting the data from a gamma-ray observatory is therefore the ability to account for this instrumental background.

4.2 Using coded masks

Collimation is the way that detectors on early gamma-ray satellite missions like the third High Energy Astrophysics Observatory (HEAO-3) and Solar Maximum Mission (SMM) operated. The method however limits the spatial resolution achievable for a given target solid angle on the sky and available exposure time. A high spatial resolution requires a small field of view for the collimator which, in turn, means that in a fixed amount of time only a proportionally smaller target solid angle can be observed.
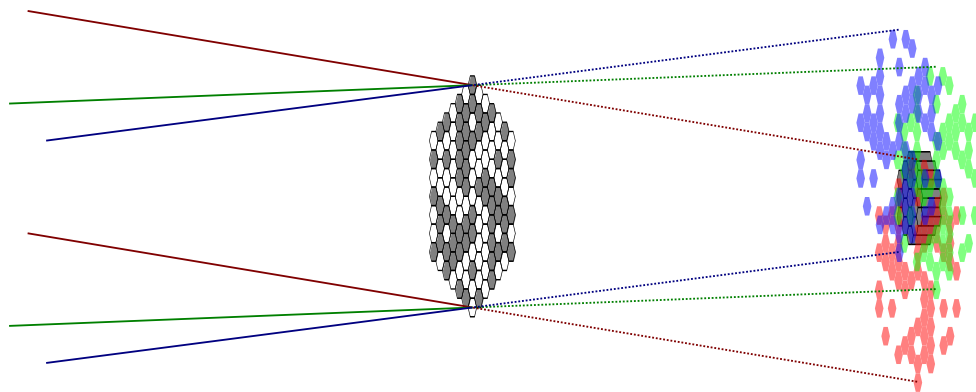
Figure 4.2: SPI shadow patterns: The radiation from three point sources (red, green and blue) entering from the left encounters the coded mask (centre). The modulated radiation creates a combination of three shifted mask patterns in the detection plane on the right.

One way around this limitation is the coded mask, a generalisation of the collimator or pinhole aperture system. A coded mask instrument consists of two parts: the mask

and the detector plane (fig. 4.2). The mask is an arrangement of either transmissive or opaque elements that modulates the incident radiation. The modulation causes a shadow pattern on the detector plane where it is sampled by a position sensitive detector, for example one consisting of an array of indepently operating pixels. Such an arrangement can allow a high spatial resolution by employing small elements in the mask and in the detector and at the same time offer a much larger field of view as a collimator with the same aperture size as the mask elements. The presence of multiple transparent mask elements means that an individual position on the detector plane can receive radiation from multiple locations on the sky at the same time. The detection of a single event is then not sufficient to determine the source position. The combination of multiple events in different detectors and/or different instrument orientations is necessary to resolve the ambiguity caused by the shape of the mask.
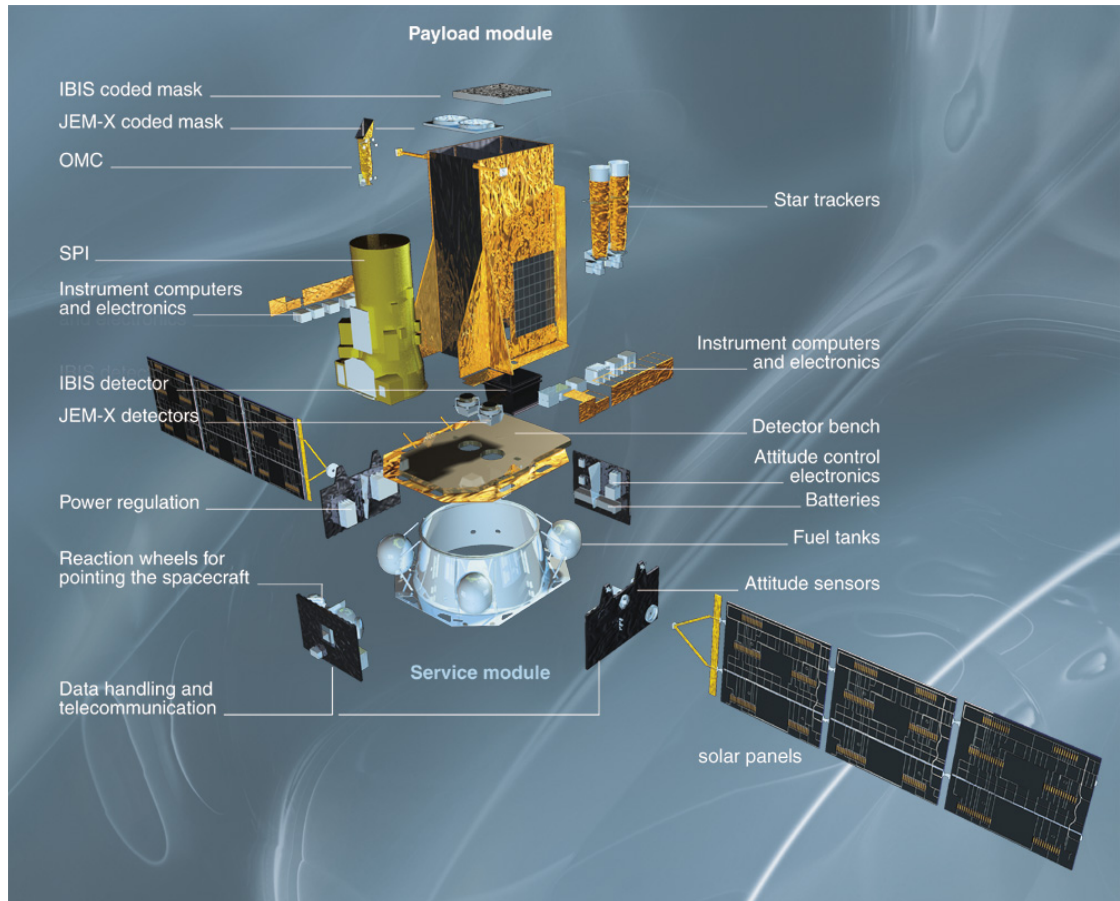
4.3 INTEGRAL

SPI, the *SPectrometer on INTEGRAL*, is a coded mask telescope optimised for the high-resolution spectroscopy of astronomical hard-X-ray and gamma-ray sources. Its platform is the *INTEGRAL* spacecraft, the Int_ernational G_amma-R_ay A_strophysics L_aboratory. It is a satellite observatory project led by the European Space Agency ESA in collaboration with the United States, Russia, the Czech Republic, and Poland.

   INTEGRAL was designed to perform follow-up observations of the hard X-ray and gamma-ray sources discovered with instruments such as CGRO/COMPTEL, BeppoSAX as well as to perform survey and monitoring functions for unknown or variable sources located in the Galactic plane. It is equipped with four instruments, which are co-axially aligned (fig. 4.3) only one of which, the Optical Monitoring Camera OMC, does not utilise the coded mask principle for its operation as it is a refracting optical telescope. The other three instruments are SPI, the *Imager on Board the INTEGRAL Satellite* (IBIS) and the *Joint European Monitor for X-rays* (JEM-X). It is already denoted by the names of the instruments that they all are optimised towards different goals and complement each other with their respective capabilities. The design of each instrument is the result of a trade-off between spatial resolution, energy resolution, energy range and other factors. The imager IBIS uses a finely patterned coded mask combined with detectors having closely spaced pixels in order to achieve a high spatial resolution. SPI, on the other hand, has a much coarser mask and correspondingly large size detectors in order to be able to measure the photon energy to high accuracy.

4.4 The design of SPI

SPI's design goal is the ability to perform high spectral resolution measurements up to photon energies of 8 MeV. The upper end of the energy range, where photons are the most penetrating, defines the thickness of the detector and mask elements required.

**Figure 4.3:** Exploded view of INTEGRAL showing its instruments and major components. Image: ESA
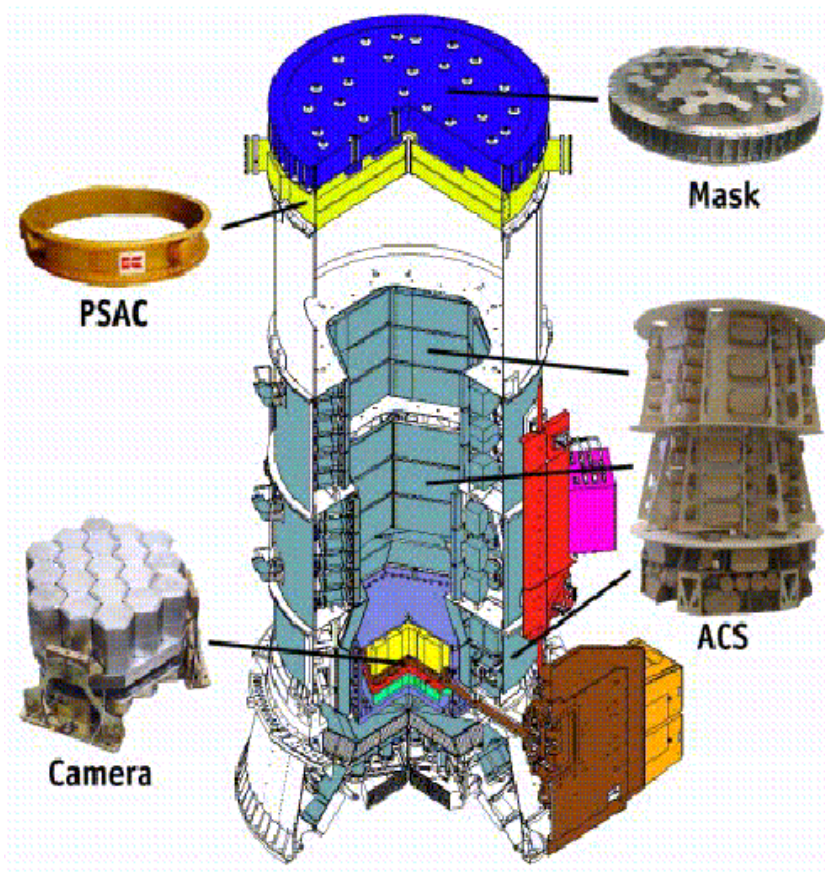
Figure 4.4: A cut-away view of the SPI instrument. The mask, plastic scintillator, camera and ACS subsystems are highlighted.

The requirement that the detectors used must be able to absorb and measure the full energy of the received photons with high probability, implies that the SPI detectors need a comparatively large thickness, 7 cm of germanium. A thinner layer of active material would result in a high probability of only a Compton scattering to occur inside the detector and the scattered photon leaving the active volume, thus losing the information about the initial photon energy. Similarly, the mask elements must have a sufficient thickness to be absorbent enough. In the case of SPI, they consist of tungsten alloy with a thickness of 30 mm, which is sufficient to absorb with an efficiency of > 95% at energies of 1 MeV. The thickness of the mask elements also affects the lateral extent of the mask structure that is practical. Consider the case where the incident radiation originates from a source located off the axis of the telescope: an incident photon will have to pass through a transparent mask element to reach the detector. Were the mask thickness larger than the lateral size of the mask elements, the transparent mask elements would have a substantially reduced cross-section for off-axis photons, limiting the effective area of the instrument.

4.5 Data interpretation

Comparing a coded mask instrument with a more conventional and familiar one, such as an optical telescope with a CCD camera, makes it clear that the interpretation of the data collected by the detectors is far less straightforward. In the case of the telescope, the simple first step of plotting an image with the colour or lightness of the pixels corresponding to the charge collected by each detector pixel already goes a long way. Of course, these images are nearly always post-processed carefully, by subtracting the dark current, compensating for vignetting, defective pixels and cosmic ray traces. Often, contrast enhancement or deconvolution with respect to the instrument's point spread function is done to increase the visibility of small-scale features in the image, but in principle, the raw image itself would already be usable. In contrast, the data returned by a coded mask instrument or one measuring the parameters of Compton interactions is not human-readable at all, maybe with the exception of a select few special cases, without computer-based data processing.

The particular difficulty experienced in gamma-ray observations is the often complicated and not "compact" structure of the function relating the celestial source emission to the data returned by detectors. Fig. 4.2 schematically shows the situation when SPI is observing a region with three point sources in the field of view. Radiation coming from the left hits the mask (centre), where it either gets absorbed in the opaque mask elements or transmitted through the mask openings and its support structure. The transmitted radiation coming from a particular direction forms a shadow pattern in the plane at the surface of the camera array which is quite similar to the shape of the mask element arrangement. To determine the direction of the incoming radiation, the location of this shadow pattern in relation to the instrument's orientation must be
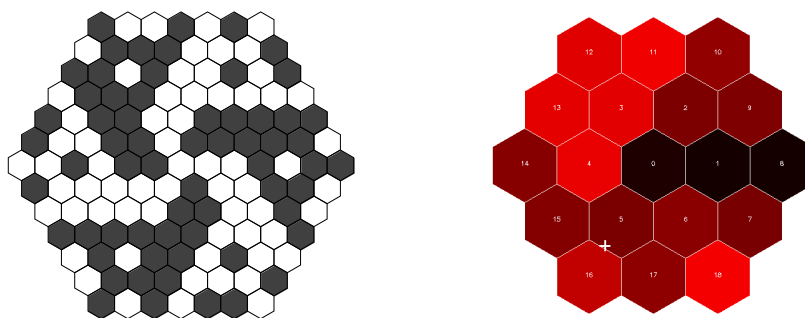
Figure 4.5: The SPI mask pattern (left) gets projected on the detector array by a point source, resulting in an exposure shadow pattern (right). In this example, the line of sight to the source is pointed 1.5° left and 2.5° down from the centre of the detector array. The plus sign marks the projection of the mask pattern's centre on the detector plane.
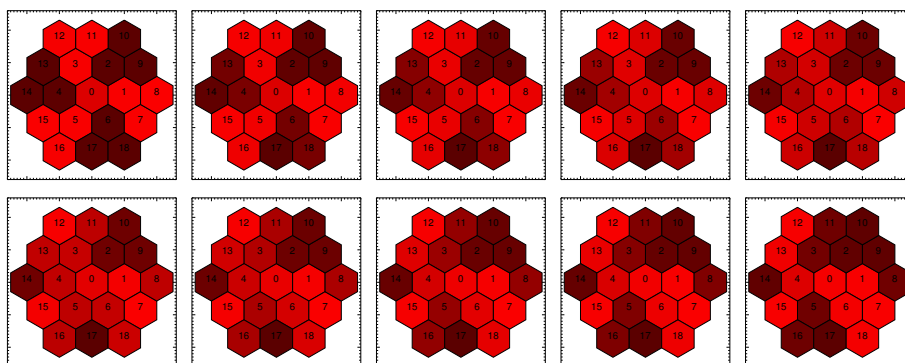
Figure 4.6: Exposure of the 19 SPI detectors in response to a point source that is moved in 0.2° steps along the horizontal axis.

determined.

Fig. 4.5 compares the SPI mask pattern with the exposure pattern resulting from its projection on to the detector plane by the radiation from a point source. The figure shows the "single events", i.e. those which interact solely within a single detector. No information about where in a particular detector the interaction took place is available. The data just provides the observed intensities in the 19 detectors.

4.6 Handling instrumental background

As an instrument sensitive to gamma rays in the MeVenergy range, SPI is affected by a number of constraints determining its performance which are unavoidable with

current technology. Probably the most important of these is the necessity to perform observations from outside the Earth's atmosphere. A direct consequence is the exposure to energetic particles from cosmic rays as well as from the sun. Interactions between these particles and the structure of the spacecraft carrying the instrument, the structure of the instrument and the detector material itself lead to nuclear reactions, some of which produce gamma photons which are, by their nature, not distinguishable from the photons one wants to measure. Practical methods for shielding the detector against energetic particle bombardment can only go so far as they are limited by the tight mass constraints of satellite construction and launch. This means that the distinction between the possibilities "source photon" and "background photon" can often not be done with a high degree of confidence when considering a single event. Only by interpreting a large number of detected events within the framework of probability theory can this problem be solved.

4.6.1 Background for the ^{26}Al 1809 keV line
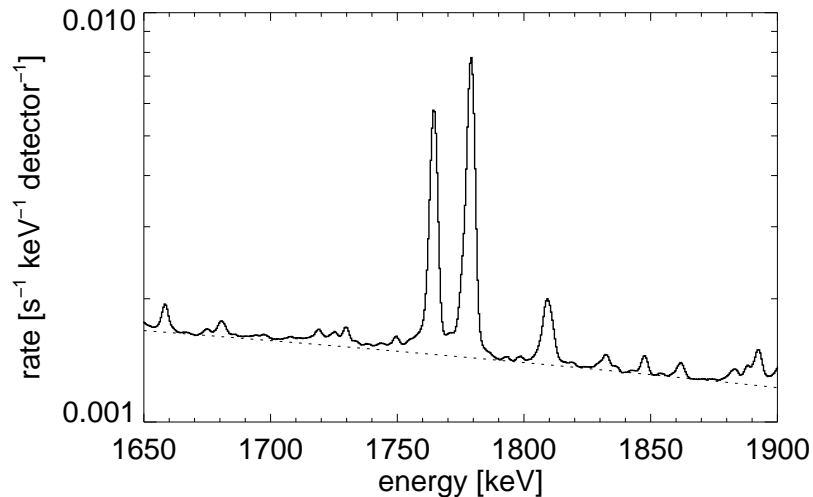
Figure 4.7: An excerpt from the SPI data showing the spectral features in the vicinity of the ^{26}Al line. The rate shown is the average over all SPI detectors over the first 250 orbits of INTEGRAL.

SPI spectra of the energy region around the position of the 1 809 keV line from ^{26}Al, as seen in fig. 4.7, show as general characteristics a continuum that is approximately linear and falls off from around $1.7 \cdot 10^{-3}$ s^{-1} keV^{-1} detector^{-1} at 1 650 keV to about $1.2 \cdot 10^{-3}$ s^{-1} keV^{-1} detector^{-1} at 1 900 keV. This continuum, hinted at by the dotted line in the figure, is overlaid by a number of line features, of which the one at 1 764 keV and the one at 1 779 keV are the most prominent. The feature at an energy of around

1 810 keV that includes the celestial ^{26}Al decay photons is significantly broader than the instrumental resolution of the germanium detectors. This larger width is due to the superposition of the sky emission with multiple background features of instrumental origin. Because of the interaction of cosmic rays with the satellite structure of INTEGRAL and the components of SPI, there is a component of the instrumental background generated by the decay of the same 1 808.70 keV excited state of ^{26}Mg that is populated during the ^{26}Al decay. Another contribution is from ^{56}Fe* at 1 810.772 keV, which can be populated by the beta decay of cosmic-ray generated ^{56}Mn having a half-life of 2.578 5 h, contrasting with the 2.241 4 min half-life of ^{28}Al whose beta decay produces the strong 1 779 keV background line.

4.6.2 Background models

The different time variability of the sky signal modulated with the mask pattern compared with the background is the feature that determines the ability to measure a signal in the presence of a variable background. This would not be possible if the behaviour of the background as a function of time was completely unknown. With more knowledge of the behaviour of the background as a function of time, it becomes possible to disentangle the contributions of signal and background.

A lot of work was put into searching for good predictors of SPI's instrumental background in different applications. A good background predictor correlates closely with the actual background intensity in the energy range of interest and is also known with a sufficient precision as not to introduce too much statistical uncertainty. It has turned out that the rate of "saturated germanium events" is a good tracer of the intensity of energetic particle interactions in the SPI instrument. This is the rate of events that generate a charge pulse larger than the maximum level of the analogue-to-digital converter's range, i.e. ones which deposit more than 8 MeV in a single detector. Because these interactions are the dominant source of instrumental background, their rate correlates closely with the background, confirmed on observations of empty fields. Their rate (typically 300 s^{-1} for one detector) is high enough to permit a low statistical uncertainty.

Because the background level in a given energy bin is not always directly proportional to the saturated germanium event rate, but may have a non-linear relationship to it, other observables also serve a useful purpose in creating a background model. In the case of line measurements, the level of the signal in adjacent energy bins to the lines of interest is a measure of the continuum in the line energy range, which is considered background for line intensity measurement purposes. Compared to the saturated germanium events, these adjacent energy ranges have much lower count rates and a correspondingly higher level of statistical uncertainty. The typical rate of events that are not rejected by the anti-coincidence veto system is < 100 s^{-1} per detector over the whole energy range, and only a small fraction of these lie in the immediate vicinity of a given spectral line of interest (see Fig. 4.7).

5 Spectroscopy with SPI

5.1 Flux determination by model fitting

The background model is one input model for the fitting procedure, the next step in the SPI data processing chain. This step uses the module `spimodfit`, which is designed to fit a set of spatial models of the source distribution and a set of background models to the SPI data. The method works by comparing the measured data to the expected data

$$d_k = \sum_j R_{jk} I_j + b_k$$

$$I_j = \sum_i \theta_i M_{ij}$$

where b_k is the background model and I_j is a set of spatial models that are transformed from the space of celestial coordinates into the space of data elements by the matrix R_{jk}, the instrument response. The image models I_j in turn are linear combinations of the spatial models M_{ij}, scaled by the model parameters θ_i. `spimodfit` employs a maximum likelihood algorithm to determine the parameters θ_i, separately for every energy interval in the energy range of interest. The output of this `spimodfit` fitting procedure is a tuple of scale parameters for every energy interval, which can then be turned into a spectrum. Fig. 5.1 shows such a spectrum obtained by plotting the scaling factor returned by `spimodfit` for the input sky model together with its uncertainty for the 30 independent 0.5 keV wide energy bins fitted.

5.2 Spectral analysis

The above procedure of fitting a spatial model plus a background model to the measured SPI data leaves us with a spectrum consisting of maximum-likelihood values and uncertainties for the intensities for every energy bin and spatial model component. But often, the question we are really interested in answering requires further analysis summarising the information contained in the typically 20–40 intensity and uncertainty values. Typically the interest will lie in knowing the probability distributions of the line intensity, its position and some simplifying summary of its shape. A highly detailed model for the shape of the instrumental response with a correspondingly large set of parameters will be disfavoured by Occam's razor unless there is strong evidence in the data for it, requiring high-precision data to push its probability higher than that
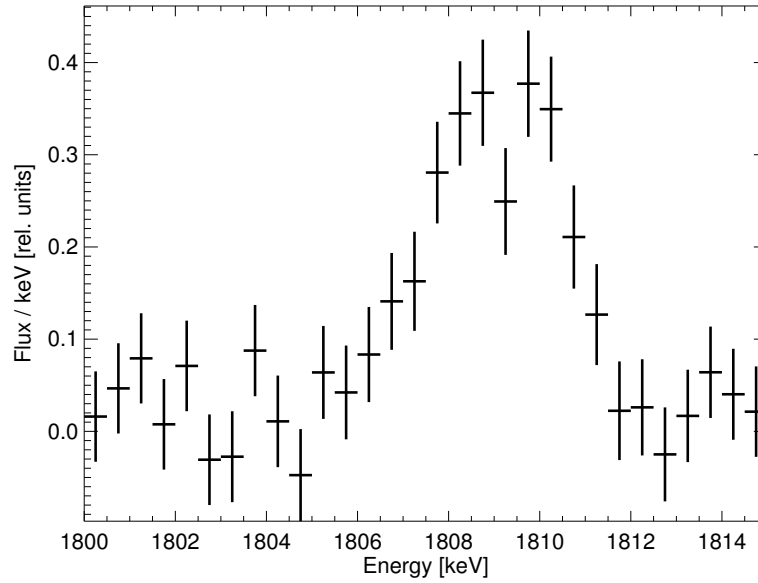
Figure 5.1: Spectrum made from `spimodfit` output values.

of a simpler model employing just one or two parameters such as the mean and the standard deviation or alternatively the FWHM of its intrinsic energy distribution.

For this type of data consolidation, we use a Bayesian parameter estimation procedure that determines the probability distribution of a parameterised model for the ^{26}Al 1808.63 keV line and the nearby continuum. The parameter space of this model is spanned by the average level of the continuum and its slope, additionally the line is specified by its intensity, its position and the FWHM of the line as it hits the instrument. A Bayesian parameter estimation for this model requires the prior probability distribution of the model parameters and the likelihood function specifying the probability of obtaining the data at hand given a particular set of model parameters. Given only the expectation value and the standard deviation of the intensity in each energy bin, the principle of maximum entropy leads to the assumption of a Gaussian distribution for the bin intensities. This makes the likelihood function the product of N_{bins} Gaussians with the mean being the expectation value of the flux and the standard deviation being the uncertainty of the measurement as returned by the model fitting procedure in `spimodfit`. What is further required for the determination of the likelihood function is therefore the expected spectrum given the model parameters.

The function we employ for the task of providing a spectral model to be used in the calculation of the likelihood function is composed of a linear continuum plus a single line. When applied to a relatively narrow energy interval, typically from 1800 keV to 1820 keV in the case of the ^{26}Al line, a linear continuum behaviour has proven to be an adequate approximation. Most of the model complexity is contained in the model

for the line itself. The shape of the line as returned by `spimodfit` is a combination of the real shape of the intensity as a function of energy as it hits the satellite with the response of the instrument itself. This response function is not known a priori and has to be determined from the instrumental data itself.

5.3 Bayesian probability

When trying to infer statements about astrophysical hypotheses from spectra obtained with SPI, Bayesian probability theory has proved to be a very useful approach. Bayesian probability theory is distinct from the probability theory familiar to most mainly by a difference in what the concept of probability refers to. Conventional probability theory, often referred to as "frequentist" by Bayesians, uses the long-term behaviour of repeated invocations of the same process that yields a random variable. The limit of the relative frequencies of different values of this random variable is the probability. In Bayesian probability theory, probability is a measure of the plausibility of statements given certain information. This concept of probability can be seen as a generalisation of the frequentist concept in the sense that in the absence of prior information about the unknown variables in question, the Bayesian and frequentist probabilities have equal values.

A great advantage of Bayesian methods compared to frequentist ones is that the different probability concept allows a Bayesian analysis to address questions not easily accessible to frequentist analysis. Because probability applies to individual statements and not only to random processes, this results in Bayesians analysing probabilities of different hypotheses given certain data, in contrast to regarding the probabilities of different data given a particular hypothesis. I want to introduce this approach using an example, then continue to its applications in the context of our SPI analysis.

5.3.1 Source detection with the "on-off"-method

The "on-off"-method makes use of the collimating properties of the instrument. Measurements of the region on the sky where the presence of a source is to be determined ("on") are compared with measurements of regions where it is known that no source is present ("off"). The off-measurement is required to determine the background intensity.

Let the data consist of pairs of intensities and times of measurement

$$D = \{(t_i, r_i), (t_j, s_j)\}, \quad i \in [1, N_r], \ j \in [1, N_s]$$

These data might have been obtained by fitting a spectral line, which is a sum of sky signal and instrumental background. The intensity of the background line may have long-term time variability, for example a linear increase due to activated radioactivity. This is case for the detection of ^{60}Fe due to build-up of ^{60}Co in the instrument due to cosmic ray activation.

We want to know first if a source is present or not — if it is present, we want to know its intensity. In the language of probability theory, we are interested in the probabilities of the hypotheses

- H_0: There is no source, the background rate is $A + Bt$, the measurement therefore is $r_i = A + Bt_i + e_i$, where e_i is the unknown measurement uncertainty.

- H_1: The source has constant intensity causing a detection event rate S, therefore $s_j = A + Bt_j + S + e_j$.

the likelihood of an individual measurement is

$$p(r_i|AB\sigma I) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(r_i - A - Bt_i)^2}{2\sigma^2}\right)$$

$$p(s_i|ABS\sigma I) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(r_i - A - Bt_i - S)^2}{2\sigma^2}\right)$$

giving a total likelihood of

$$p(D|ABS\sigma I) = \left(\frac{1}{2\pi\sigma^2}\right)^{(N_r+N_s)/2} \cdot \exp\left(-\frac{N_r}{2\sigma^2}Q_r(A,B) - \frac{N_s}{2\sigma^2}Q_r(A,B,S)\right)$$

where the quadratic forms Q_r and Q_s are

$$
\begin{aligned}
Q_r(A,B) &= \frac{1}{N}\sum_{i=1}^{N_r}(r_i - A - Bt_i)^2 = \\
&= \frac{1}{N}\sum_{i=1}^{N_r}r_i^2 + A^2 + B^2t_i^2 - r_iA - r_iBt_i + ABt_i \\
&= \overline{r^2} + A^2 + b^2\overline{t^2} - \overline{r}A - \overline{rt}B + AB\overline{t} \\
Q_s(A,B,S) &= \frac{1}{N}\sum_{i=1}^{N_s}(s_i - A - Bt_i - S)^2 = \\
&= \frac{1}{N}\sum_{i=1}^{N_s}s_i^2 + A^2 + B^2t_i^2 + S^2 - s_iA - s_iBt_i - s_iS + ABt_i + AS + Bt_iS \\
&= \overline{s^2} + A^2 + B^2\overline{t^2} + S^2 - \overline{s}A - \overline{st}B + \overline{s}S + AB\overline{t} + AS + BS\overline{t}
\end{aligned}
$$

5.3.2 Bayesian methods for spectral fitting

While experimenting with fitting parts of SPI spectra with a model consisting of a collection of lines on top of a continuum, using Bayesian methods has been useful. A clear advantage is the straightforward way in which to include background information, such as instrumental lines always being emission lines, not absorption lines, i.e. that the amplitude of a line is always positive.

An application of Bayesian parameter estimation of high importance to the study of Doppler shifted lines is the determination of the achievable accuracy of line position measurements.

A simplified situation that is useful to determine the influence of the data points and their uncertainty on the accuracy achievable during line parameter estimation is the following: I assume the signal to be a Gauss-shaped peak on top of a continuum whose intensity is independent of energy.

$$P(abE_0\sigma|d_1\ldots d_N eI) = P(abE_0\sigma|I) \cdot \frac{P(d_1\ldots d_N|abE_0\sigma eI)}{P(d_1\ldots d_N|eI))}$$

$$= P(abE_0\sigma|I) \cdot \frac{\prod\limits_{i=1}^{N} \frac{1}{\sqrt{2\pi e}} \cdot \exp\left(-\frac{(d_i-F_i)^2}{2e^2}\right)}{\int\limits_{a,b,E_0,\sigma} P(d_1\ldots d_N|abE_0\sigma eI) \cdot P(abE_0\sigma|eI)\, \mathrm{d}a\mathrm{d}b\mathrm{d}E_0\mathrm{d}\sigma}$$

$$= P(abE_0\sigma|I) \cdot \frac{\left(\frac{1}{\sqrt{2\pi e}}\right)^N \cdot \exp\left(-\sum\limits_{i=1}^{N}\frac{(d_i-f_i)^2}{2e^2}\right)^2}{\int\limits_{a,b,E_0,\sigma} \left(\frac{1}{\sqrt{2\pi e}}\right)^N \cdot \exp\left(-\sum\limits_{i=1}^{N}\frac{(d_i-f_i)^2}{2e^2}\right)^2 \cdot P(abE_0\sigma|eI)\, \mathrm{d}a\mathrm{d}b\mathrm{d}E_0\mathrm{d}\sigma}$$

$$= P(abE_0\sigma|I) \cdot \frac{\exp\left(-\sum\limits_{i=1}^{N}(d_i-f_i)^2\right)^2}{\int\limits_{a,b,E_0,\sigma} \exp\left(-\sum\limits_{i=1}^{N}(d_i-f_i)^2\right)^2 \cdot P(abE_0\sigma|eI)\, \mathrm{d}a\mathrm{d}b\mathrm{d}E_0\mathrm{d}\sigma}$$

Here, d_i are the individual data points, related to the individual flux values f_i per energy bin by way of the errors e_i which are normally distributed around an expectation value of zero (i.e. no systematic bias) and have the standard deviation e.

$$d_i = a + \frac{b}{\sqrt{2\pi\sigma}} \cdot \exp\left(-\frac{(E_i-E_0)^2}{2\sigma^2}\right) + e_i = f_i + e_i$$

$$P(e_i \in [x, x+\Delta x]) = \frac{1}{\sqrt{2\pi e}} \exp\left(-\frac{x^2}{2e^2}\right) \Delta x$$

5.4 Instrumental line shape

An important part of the detector response function in a spectrometer is the spectral response, i.e. the relationship between the energy of a detected photon and the numerical value returned by the measurement electronics. In general, assuming we know the track and the energy of an incoming photon, we will not be able to predict what value the instrument will return. Therefore the spectral response, as a function of detection

geometry, is a probability distribution

$$P(\text{data}|\text{photon}, \text{instrument})$$

describing probability of obtaining the data, given the incidence of the specified photon with our instrument. Because we *know* the data, but want to infer information about the properties of the received photons, this probability distribution is necessary for constructing the likelihood function of our data.

The processes giving rise to the detector response function are dominated by the generation of electron-hole pairs in the semiconductor material of the detector and transport of these charge carriers to the readout electrodes on the crystal surface. If all the energy of the fast electron could be converted into excitation of electron-hole pairs requiring the band gap energy, the number of charge carriers produced as a function of photon energy would be uniquely determined as the largest integer less than photon energy divided by band gap energy. However, electronic transitions at or just above the band gap energy can access only a small phase space volume, making the process improbable. If an electron-hole pair with higher energy than the band gap is excited, it can and will lose the excess energy because the conduction band is largely empty. The energy can go into the creation of further electron-hole pairs or into the excitation of phonons. Because we measure the amount of charge collected at the electrodes coated onto the detector crystal surface, any energy converted into phonons, i.e. internal energy of the crystal lattice, can not be measured in this way. This means that the excitation of an electron-hole pair requires at least the band gap energy plus a variable additional amount of energy going into phonons.

In general, the probability distribution of the energy loss of the initial fast electron required per electron-hole pair is a complex function. However, as the central limit theorem (see e.g. Jaynes 2003, chapter 7) shows that the details of this complexity are actually not of great relevance to the final line shape measured. If the number of independent interactions leading to the final state where no further electron-hole pair excitations can occur is large, the probability distribution for this number will tend towards a Gaussian distribution.

In practice, this is not the only mechanism at play here, therefore, the line shape of the SPI germanium detector response is not Gaussian. Furthermore it is also not constant over time, so knowledge of the class of line shapes possible and of their changes is necessary in order to construct precise background models.

5.4.1 Time-variability of the response

The main cause of the change in the SPI detector response as a function of time is damage of the germanium crystal structure in response to cosmic ray interactions. This damage was known during the mission design phase and accounted for by testing different detector materials for their susceptibility to radiation damage (Borrel et al. 1999, Kandel et al. 1999). A high energy particle hitting the detector can transmit

recoil momentum to a germanium nucleus. If this recoil is large enough to overcome the potential barrier surrounding the nucleus at its equilibrium position within the crystal lattice, the nucleus may come to rest in a different potential minimum, leaving a disturbance in the otherwise regular lattice structure. This disturbance consists for example of an atom at an interstitial lattice position and a vacancy at the atom's original position. This type of point defect has consequences for the electronic structure of the semiconductor, generally causing the presence of electronic states in the band gap.

Charge carriers can become temporarily trapped in these deep centres, which form local potential minima. With a potential difference of less than a few tenths of eV, a fraction of the band gap energy of germanium, which is less than 0.75 eV (Debertin and Helmer 1988), these local minima are shallow compared to the potential difference of several thousand volts between the readout electrodes on the crystal surface. The externally applied electric field means that the electrostatic potential becomes lower than that of the trapping centre a short distance away from it. This results in a finite thickness potential barrier for the trapped charges, allowing those that do not recombine to eventually leave the trapping centres by tunnelling through the potential barrier. This process however incurs a time delay that lengthens the charge collection process and so causes a change in the shape of the charge pulse delivered to the readout electronics. The change in the shape of this pulse also causes a change in the analogue-to-digital converter output for a given photon energy. This change is not constant for all recorded events given a constant photon energy, therefore the effect on the recorded spectrum from a mono-energetic source is not just a shift of the peak energy but a more complex distortion of the line shape as well.
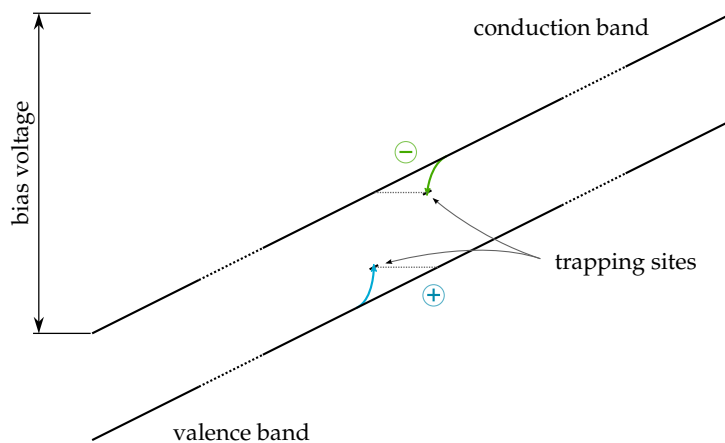
Figure 5.2: Electronic structure of a deep defect (trapping centre). The trap is able to capture electrons from the conduction band or holes from the valence band.

The literature available on this subject discusses a number of different functions designed to reproduce the shape of the recorded spectrum in the vicinity of the photo-

peak (see e.g. the collection in Debertin and Helmer 1988), determined for various different detector designs. I had to choose a model suitable for the SPI detectors and their operating environment and the methods at hand for analysing the data.

As an ideal semiconductor detector would produce instrumental lines with Gaussian shape, determined purely by the properties inherent to the detector material, an attractive approach is to model the detector degradation as a distortion of the intrinsic properties. The function used to perform this transformation would take as input the Gaussian shape of the idealised detector, which is constant over time and a set of parameters that are functions of time or of the radiation dose absorbed in the detector, the output would be the line shape that the detector produces in response to incoming mono-energetic radiation.

A model function for applying this approach that is attractive from a physical standpoint is the convolution of the intrinsic Gaussian line with a truncated exponential, i.e. a function whose value is e^{-x} for $x \leq 0$ and 0 for $x > 0$. With the definition of the convolution

$$f_1 * f_2 = \int_{-\infty}^{+\infty} f_1(\tau) f_2(t - \tau) \, d\tau$$

The intensity as a function of energy for the line is

$$I(E) = \int_0^\infty \frac{e^{-E'/\tau}}{\tau} \cdot G(E + E') \, dE'$$
$$= \frac{1}{2\tau} \cdot \exp\left(\frac{2E\tau + \sigma^2}{2\tau^2}\right) \cdot \text{erfc}\left(\frac{E\tau + \sigma^2}{\sqrt{2}\,\sigma\tau}\right)$$

The argument for the choice of this function as a model for the instrumental response is the desire to model the overall characteristics of the changing line shape in an adequate way without needing to determine too many parameters. A more detailed model would require describing detail that is difficult to measure accurately given the often low count rates in the energy range of interest, thereby sacrificing precision.

Physical motivation for the model line shape function comes from a simplified model of charge transfer in the detector crystal. This assumes that only one type of charge carriers is affected by trapping in the crystal, that the charge transfer occurs along one dimension only and that the crystal is homogeneous along this dimension, e.g. the axis of a cylinder. A cloud of charge carriers generated at a distance l from the electrode collecting the respective type of carrier will be subject to trapping of a fraction of them, the non-trapped portion being proportional to $\exp(-\kappa l)$, with the absorption coefficient κ.

Figure 5.3 shows a plot of a set of functions of this type for different values of τ, while keeping the shape of the underlying Gaussian constant at a full width at half maximum of one energy unit and a peak intensity of one intensity unit. Clearly evident to the eye is the increase in width and the shift towards lower energies with increasing degradation parameter τ.
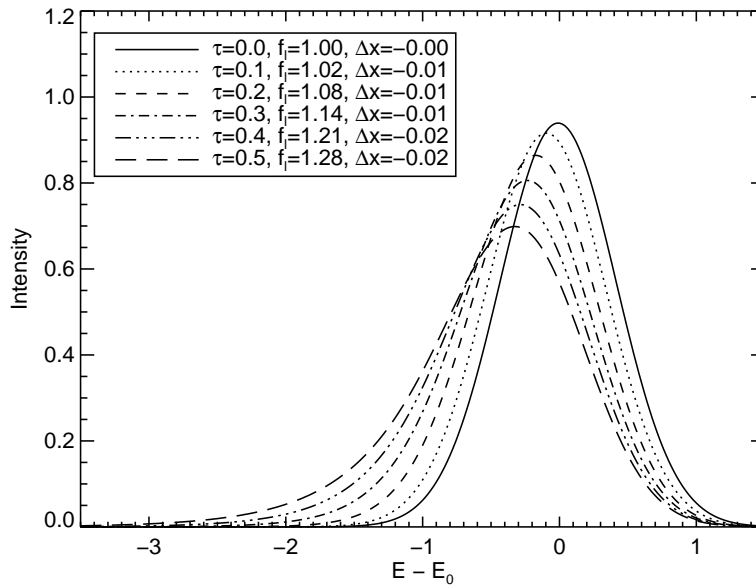
Figure 5.3: Line shape function used to model the behaviour of SPI's instrumental lines as a function of time.

5.4.2 Line shape calibration

The model for the time dependent line shape described in section 5.4.1 serves as a basis for data analysis while accounting for the change of instrumental response with time. The model is used as an input for fitting background lines in the SPI data at different points in time during the mission and for different energy ranges. These individual measurements are then consolidated into a model function allowing the determination of the intrinsic width of the observed sky emission in the face of varying instrumental response parameters.

One motivation for using the Gaussian convolved with an exponential described in section 5.4.1 was the idea that a suitable choice of the model function should allow to separate the constant properties of the detector material and geometry from the time-variable effects of degradation. In relation to the model function, this means that the width of the Gaussian component was expected to remain approximately constant as a function of time, whereas the decay constant of the exponential referring to the charge trapping processes would account for the majority of the time variability of the instrumental line shape.

To test this assumption, I developed software for automated fitting of spectral model to the measured data for a selected set of strong instrumental lines. I chose a set of a total of 18 lines in five energy regions, shown in table 5.1. In each of these five energy intervals, the algorithm fits a model composed of a linear continuum plus a set of

Table 5.1: Instrumental background lines used for calibrating the time variability of SPI's spectral response.

energy range [keV]	line energy [keV]	producing reaction
415 − 460	439	$^{69}\mathrm{Cu}(\beta^-) \rightarrow {}^{69}\mathrm{Zn}^{\mathrm{m}}$
565 − 590	575	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga}$
	585	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga} + \mathrm{K}$
860 − 930	873	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga}$
	883	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga} + \mathrm{K}$
	900	
	912	
1060 − 1175	1108	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga}$
	1110	
	1115.5	$^{65}\mathrm{Ni}(\beta^-) \rightarrow {}^{65}\mathrm{Zn}(\mathrm{EC}) \rightarrow {}^{65}\mathrm{Cu}$
	1117.2	$^{69}\mathrm{Ge}(\mathrm{EC}) \rightarrow {}^{69}\mathrm{Ga} + \mathrm{K}$
	1121	
	1125	$^{65}\mathrm{Zn}(\mathrm{EC}) \rightarrow {}^{65}\mathrm{Cu}$
1740 − 1820	1764	$^{205}\mathrm{Bi}(\mathrm{EC}) \rightarrow {}^{205}\mathrm{Pb}$
	1776	$^{205}\mathrm{Bi}(\mathrm{EC}) \rightarrow {}^{205}\mathrm{Pb}$
	1780	$^{28}\mathrm{Al}(\beta^-) \rightarrow {}^{28}\mathrm{Si}$
	1808.7	$^{26}\mathrm{Na}(\beta^-) \rightarrow {}^{26}\mathrm{Mg}$
	1810.7	$^{56}\mathrm{Mn}(\beta^-) \rightarrow {}^{56}\mathrm{Co}(\mathrm{EC}) \rightarrow {}^{56}\mathrm{Fe}$

spectral lines to the data. A two-step process is used in order to speed up the numerical calculations: first, a conventional fitting algorithm, the `curvefit` procedure provided with IDL finds a point in parameter space close to the maximum of the probability distribution. As a second step, a Bayesian parameter estimation procedure (see sec. 5.3) based on the Markov Chain Monte Carlo algorithm is performed to sample the posterior probability distribution of the model parameters. This algorithm allows taking into account prior probabilities that are available in the form of various constraints. One such information is the type of parameter, such as whether it is a location parameter or a scale parameter. In the absence of other information, a location parameter would be assigned a uniform prior, a scale parameter a Jeffreys prior, i.e. one that is proportional to $1/x$. Another constraint is when parameter ranges are known, such as the fact that a given slice of the spectrum cannot be used for inference about lines outside the energy range covered by more than a few line widths, because such lines would have negligible influence on the count rates seen inside the energy window under consideration.
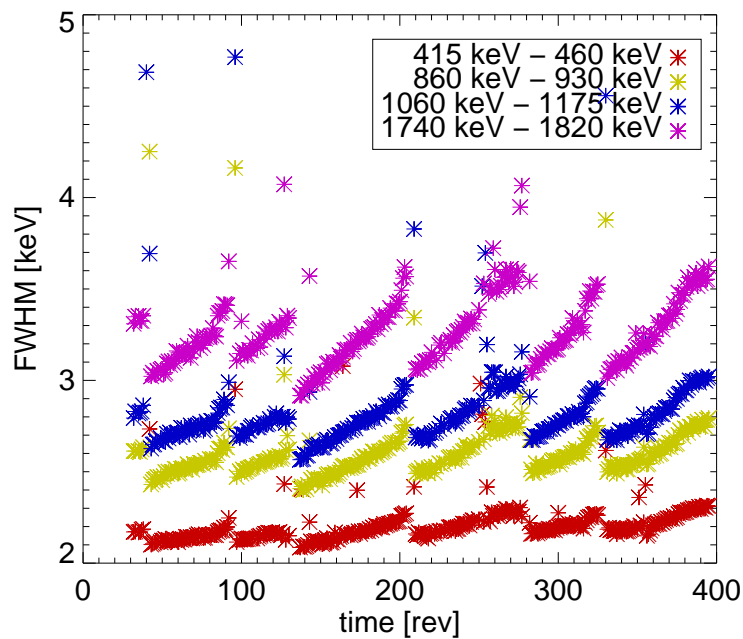
Figure 5.4: The width of the SPI response in five energy bands, unfiltered fit results.

Initially, the data set consisting of the results of this fitting procedure sported a significant number of outliers in the estimated parameters, pointing to the presence of unusual data or problems with the fitting procedure (see fig. 5.4). An investigation of the causes of these outliers showed that nearly all could be traced to one of two possible causes. The first cause was data sets with much lower than average exposure time in the respective orbit, with therefore much less counts per energy bin and a correspondingly

lower signal-to-noise ratio. This resulted in large scatter and uncertainties in the estimation of all parameters because data sets with such low event counts constrain the parameter space only weakly. The second cause found was the variability of the detector temperatures. A large amount of temperature variation during a single orbit causes, when not compensated for, a broadening of the energy response of the instrument, because the observed shape is then a superposition of slightly shifted constant-temperature instrumental responses as a result of the variation in detector gain with temperature.

In order to test the performance of the fit algorithm employed for calibrating the SPI energy response over time, Monte Carlo simulations generating synthetic spectra with known properties and then fitting these spectra are useful. I performed an experiment that first creates a spectrum consisting of a linear continuum background with a superimposed line matching the Gauss-convolved-with-exponential shape in use for SPI line shape calibration. In order to test the dependence of the line position uncertainty on the signal level, a series of model spectra with proportionally increasing continuum and line intensities were first constructed according to the fixed shape parameters and then fitted using the same Markov chain Monte Carlo algorithm that is used for calibration purposes. The proportional scaling of both continuum and line intensity ensures that the overall shape of the spectrum remains the same, while the relative uncertainty of the number of counts in each energy bin varies according to the Poisson distribution as $\sqrt{N}$.

5.4.3 Effective instrumental response

In order to be able to assess the intrinsic width of the sky emission, it is necessary to know the instrumental response for the data set used in the model fitting procedure. The spatial model fitting procedure currently in use in our group treats the energy bins of the data set to be analysed completely independently from each other. This means that the effect of the instrumental response on the observed spectrum given a particular sky spectrum have to be modelled in a separate processing step done after the energy-independent sky model fitting.

As the observations used for model fitting have been filtered in order to reduce the influence of periods with poorly known background behaviour, they do not consist only of full orbits, different from the data set used for line shape calibration. This means that the effective instrumental response has to be determined from the set of individual exposures used. To do this, the calibration data set described above (section 5.4.2) is used to obtain the averaged spectral response corresponding to the set of instrument pointings actually used for a given analysis.

The method I use for this uses the fit results from the procedure described above in section 5.4.2. However, just using the line shape parameters straight from the calibration measurements is not making enough use of the available information because the obvious correlations between subsequent INTEGRAL revolutions in a period between
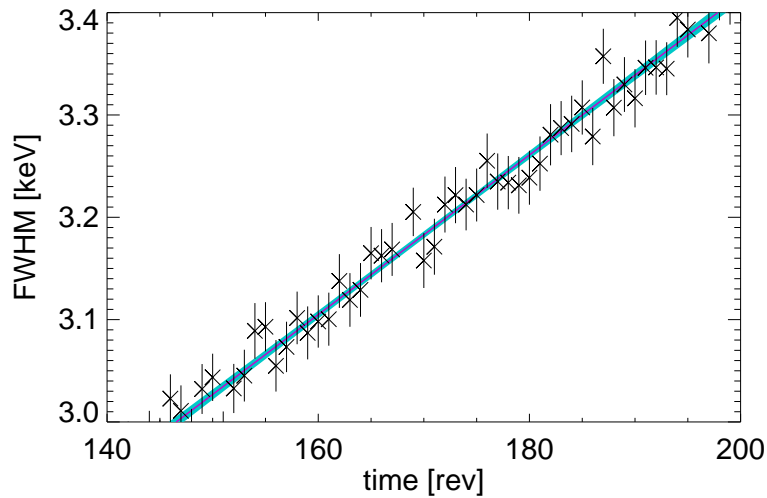
Figure 5.5: An excerpt from the calibration procedure described in section 5.4.3. The data points show the line width measured in a single revolution, averaged over the 19 detectors. The shaded area shows the uncertainty range of the piecewise linear fit.

two annealings are not taken into account. So, what is done is a consolidation of the calibration line shape data by first averaging over the 19 individual detectors and then modelling the change of the line shape parameters over time. Visual inspection of the plots showing the behaviour of these parameters suggests that a piecewise linear model should be of sufficient accuracy to reproduce the general behaviour of the parameters while averaging over the individual-detector, individual-revolution fit results. This averaging process drastically reduces the uncertainties stemming from the statistical fluctuations of the individual fit results. Figure 5.5 shows this by plotting the individual error bars of the fits of the FWHM of the lines in the energy region 1 740 keV – 1 820 keV, already having been averaged over the 19 individual detectors, thereby giving a reduction of the error bars by a factor of $\sqrt{19}$. It also shows as a shaded band the region of uncertainty of the piecewise linear fit to the same data points, which is clearly much narrower. In the time interval shown, the individual data points have an average uncertainty of 25 eV whereas the average uncertainty of the linear fit is only 4.0 eV. The measured parameters f_{line} and f_{G}, the full width at half maximum of the full response and its Gaussian component and the corresponding fits over the first eight hundred INTEGRAL orbits, are shown in Fig. 5.6. The hypothesis that the width of the Gaussian component remains constant is justified on the 6-month time scale between detector annealings, it does however show a long term increase indicating that there is a component of the radiation damage that can not be completely compensated. The
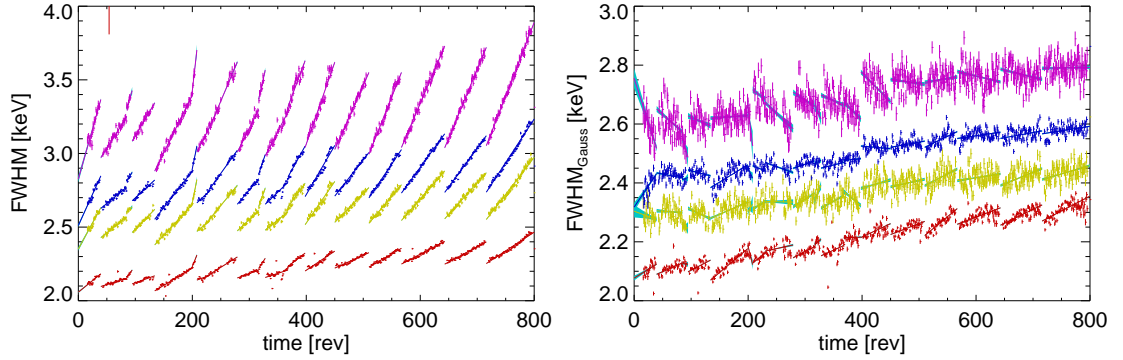
Figure 5.6: The full width at half maximum of the SPI response (left) and that of the Gaussian component (right) in five energy bands. From bottom to top: red: $415 \,\text{keV} - 460 \,\text{keV}$, yellow: $860 \,\text{keV} - 930 \,\text{keV}$, blue: $1\,060 \,\text{keV} - 1\,175 \,\text{keV}$, magenta: $1\,740 \,\text{keV} - 1\,820 \,\text{keV}$, .

periodic detector annealing cycles are however successfully restoring the SPI spectral performance close to launch levels.

The model of the line shape parameters consisting of the three values width of the Gaussian component f_G, energy scale of the truncated exponential component τ and shift of the line position Δ_E as piecewise linear functions serves as the input to the next step: determining the average instrumental response for a given observation. For every instrument pointing included in the data set to be used in the analysis, the best estimate for the response parameter tuple $(f_\text{G}, \tau, \Delta_\text{E})$ is determined from the database of piecewise linear fits to the calibration measurements. The Δ_E parameter is adjusted for the radial velocity component of the Earth's orbital motion in the solar system (see Fig. 5.9). The tuple is then used to calculate a finely resolved model spectrum using an energy bin size of 10 eV. The model spectra for each pointing are then averaged, weighted by the exposure time of the pointings. The resulting average instrumental response functions are plotted in Fig. 5.7, showing the influence of selecting different subsets of the available data based on Galactic longitude intervals, and Fig. 5.8, showing the energy dependence of the instrumental response.

5.5 Evidence determination

The Bayesian equivalent to the concept of the confidence level is the Bayes factor, which is often also called the "odds", the ratio of the probabilities of two models, given the same data and relevant background information. When applied to the detection of a spectral line this ratio is the ratio of the marginal probability of the model including the spectral line to the model without the line (Jaynes 2003, 4.2).

$$e(H_\text{line}|DI) = 10 \log_{10} O(H_\text{line}|DI) = 10 \log_{10} P(H_\text{line}|DI)/P(H_\text{no line}|DI)$$
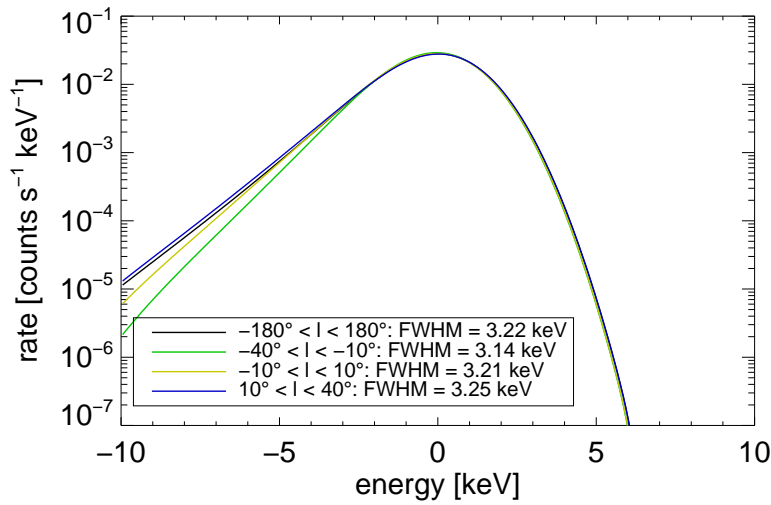
Figure 5.7: Average spectral response of SPI at the energy of 1 764.328 1 keV for the complete data set spanning the INTEGRAL revolutions 38 – 358 and subsets with pointing directions in three Galactic longitude intervals.
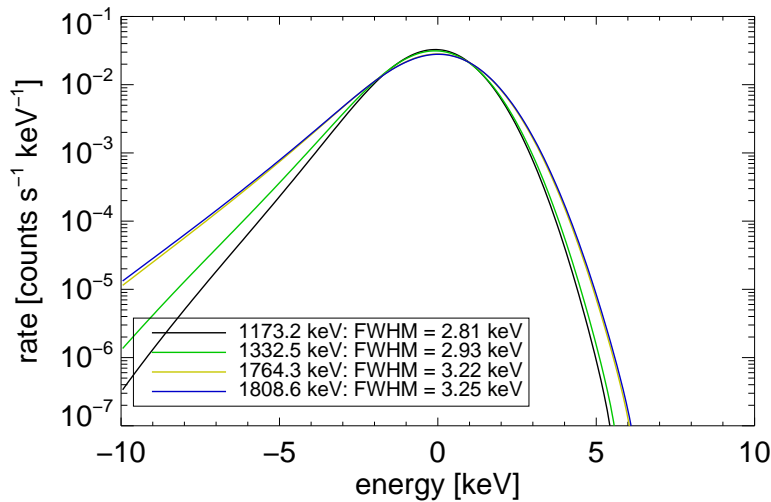
Figure 5.8: Average spectral response of SPI at the energies of the two dominant lines from the ^{60}Co decay at 1 173.237 keV and 1 332.501 keV, the 1 764.321 8 keV calibration line and the for the 1 808.63 keV line from the ^{26}Al decay for the complete data set spanning the INTEGRAL revolutions 38 – 358.
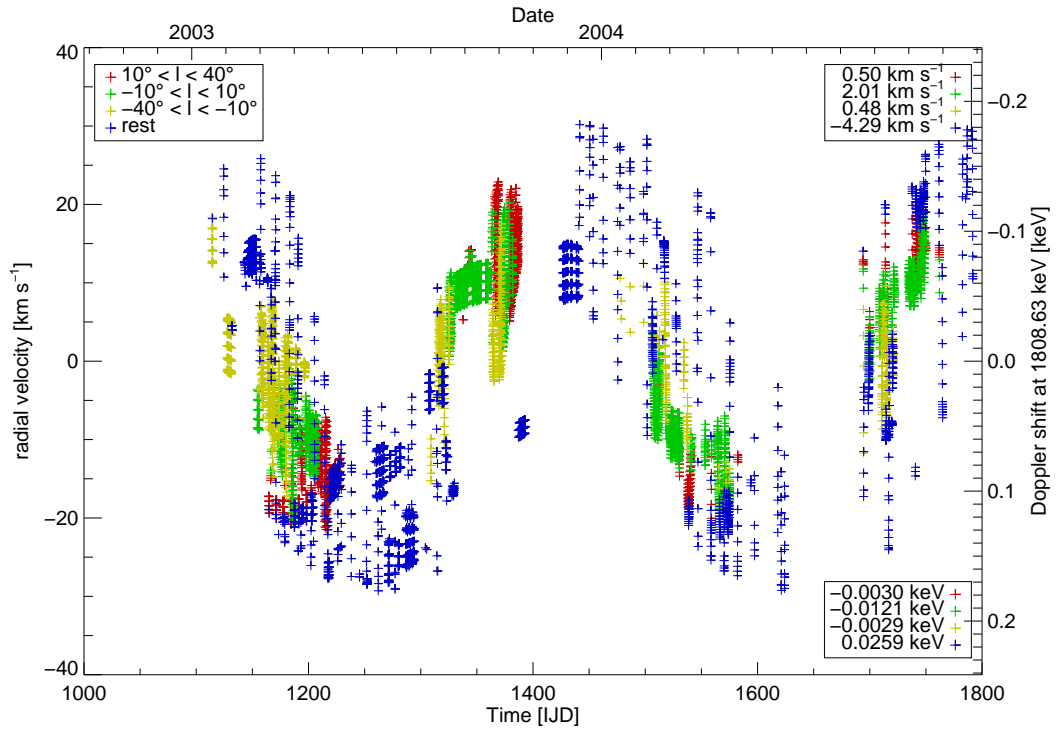
Figure 5.9: Influence of the Earth's barycentric velocity on the observed energy of the
^{26}Al line

In particular, we are interested in the probability of a line being present relative to the presence of a continuum only. In classical statistics, a maximum likelihood ratio test would be done, comparing the probabilities of obtaining the observed data under the respective models given those model parameters that maximise the likelihood. By contrast, the Bayes factor is the ratio of the marginal likelihoods of the models, averaging over the volume of parameter space specified and weighted by the prior probabilities of the parameters specifying the models. This has the effect of penalising models that are very unspecific, meaning that a large fraction of the parameter-space volume where the prior is large consists of parameters corresponding to low likelihoods. When this is the case, it means that the model in question yields on average, as specified by the prior, low likelihoods. Only for a comparatively small subregion of the parameter space is the likelihood high. This property provides an inbuilt Occam's razor mechanism that helps to avoid choosing a complex model when this is not warranted by the data. A complex model with a large number of parameters can fit the data well because of the large amount of flexibility provided by its high-dimensional parameter space but only for a highly specific choice of parameters, making it not very robust. If our knowledge and with it the prior probabilities do not restrict the model to a small subset of the parameter space around the most probable values, a simpler model is often preferable.

To see why this is the case, it is instructive to compare the simple cases of a model with a single parameter in relation to another model for the same data, but with two parameters instead. With a single parameter, close to the maximum of the posterior probability distribution, the probability can usually be approximated by a Gaussian, which is proportional to the likelihood in the absence of cogent prior information The plot of the log-likelihood is therefore a downward-opening parabola. In the case of two parameters, the situation is analogous, but with a two-dimensional Gaussian and a paraboloid for probability distribution and log-likelihood, respectively.

In the specific case of testing for the presence or absence of a spectral line, the choice is between a model containing a continuum only and a model containing continuum and line. The models have parameter spaces of different dimensionality, in this case both have two parameters for the linear continuum specifying the average continuum intensity and its slope with respect to photon energy. Our line model has three additional parameters specifying the intensity, position and shape of the spectral line, yielding a total of five dimensions for its parameter space. The effect of a higher-dimensional parameter space on the average likelihood of a model can be qualitatively understood by looking at the example of a n-dimensional Gaussian for the likelihood function. If the standard deviation of the Gaussian is identical for the 2-D and 5-D distributions, then the difference in dimensionality leads to a different weighting of the areas with lower than the maximum likelihood contributing to the average. The situation is similar to the density of states in an electron gas of different dimension, be it 3-D, where the functional dependence of the density of states is $\sim k$ or $\sim k^{-1}$ for a one-dimensional electron gas. In the 5-D case, the outskirts of the probability distribution where the likelihood is low comprise a larger fraction of the

parameter volume.

In the framework of Markov-chain Monte Carlo based parameter integration, the Bayes factor of a set of models, marginalising over their respective parameter spaces, can be calculated from the likelihood values corresponding to the samples drawn from the parameter space once the Markov chain has stabilised. By averaging the log-likelihood samples of the models in question, we get a marginal log-likelihood for each model, taking into account the prior probabilities of the model parameters. The most plausible model given the data is the one with the highest marginal log-likelihood value.

The model fitting method we use to analyse SPI data for the presence of line features provides us with a spectrum consisting of most likely flux values per energy bin with corresponding error bars. In the absence of additional information, i.e. when only the mean and the standard deviation of the individual uncertainties are known, maximum entropy arguments lead us the assumption of a normal distribution for the sampling probabilities in each bin.

5.6 Analysing the results from spatial model-fitting

When looking at the spectrum obtained from modelling the emission of the inner Galaxy with the COMPTEL 1.8 MeV map (Diehl et al. 2006), the spectral shape obtained allows to constrain the intrinsic width of the sky emission. The challenge lies in disentangling the intrinsic properties of the received radiation from detector effects. The key to this is the precise knowledge of the instrumental response during the particular observation at hand. With the knowledge of the response as described in sec. 5.4.3, we are again able to determine the likelihood function for the observed data, which are in this application the intensities per energy bin obtained from spatial model fitting, as a function of the spectral properties, in this case the shape of the intrinsic sky emission spectrum, as specified by a parameterised function. Knowledge of the likelihood function in combination with the prior probabilities of the parameters describing the intrinsic spectral shape allows, via Bayes' theorem, to update these probabilities.

The method I used for this approach employs again the Metropolis-Hastings algorithm using simulated annealing that has already been used for the instrumental line shape calibration (sec. 5.4.2). The different application requires however a different likelihood function and, due to the different parameter set, a different set of prior probabilities. In contrast to the raw detector counts modelled for the line shape calibration, where the sampling probabilities of the counts are Poissonian, in this case we are working with the output of a previous stage of probabilistic estimation, in particular one that approximated the shape of the likelihood function in the vicinity of the point of maximum likelihood by a paraboloid. Because this means that we have only the first and second moments of the intensity per spectral bin's probability distribution available, maximum entropy requires a normally distributed sampling probability.

Table 5.2: Parameters used in the spectral fitting procedure applied to the `spimodfit` output.

| #	parameter
0	continuum level
1	continuum slope
2	line intensity
3	line position
4	FWHM of the Gaussian

To build a likelihood function, this set of data points and a corresponding spectral model is needed. This model has to be as simple as possible to minimise the uncertainty coming from multiple correlated parameters. At the same time, it must include the spectral features of interest in parameterised form. These requirements are satisfied by a spectrum determined by a linear (in energy) continuum with a single superimposed spectral line. This line is a combination of an instrumental line having the average shape determined in Sec. 5.4.3 with a Gaussian representing the celestial line shape in the absence of further information about its higher moments. Fig. 5.10 shows the five marginal probability distributions of the five parameters (Table 5.2) determining this model:

It is clear from the plots that the first four parameters are well constrained by the data and their probability distributions could be closely approximated by Gaussians. In such a case there is little difference to a treatment using conventional statistical methods. However the fifth parameter, the intrinsic width of the Galactic line emission, is a different matter and shown by itself in fig. 5.11. Its probability distribution is obviously not Gaussian at all. The figure shows, in addition to the probability density, the upper limits of the 68.3%, 95.4% and 99.7% highest posterior density (HPD) regions, corresponding to the probability mass contained beneath a Gaussian in the intervals $\pm 1\sigma$, $\pm 2\sigma$ and $\pm 3\sigma$. The lower limit of the HPD regions is zero in all three cases. The mean of the probability distribution is also shown. Fig. 5.12 shows how this probability distribution changes when more data become available.

The shape of the probability distribution for the Galactic FWHM shows that in this application, approximating the full distribution by stating only its mean and standard deviation gives unsatisfactory results as they describe a Gaussian distribution that differs grossly from the full MCMC-derived one. Instead, in a case like Fig. 5.12, more useful summaries of the full distribution are the upper limits at various probability levels. Since in this particular case, the lower bound of the three shown HPD regions is zero, their upper bounds are identical to the distribution's upper limits.
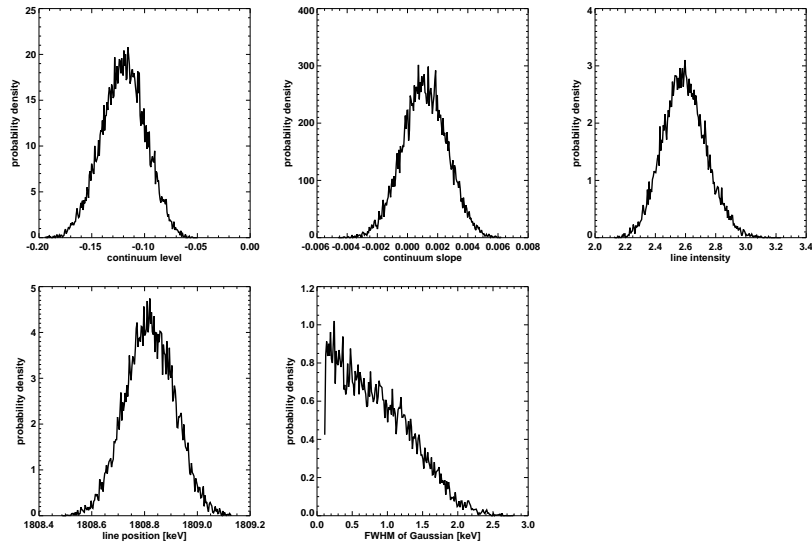
Figure 5.10: Probability distributions of the five parameters used to model the spectrum obtained from modelling the emission from the inner Galaxy with the COMPTEL 1.8 MeV map. Based on INTEGRAL orbits 15–259. See also Diehl et al. (2006)
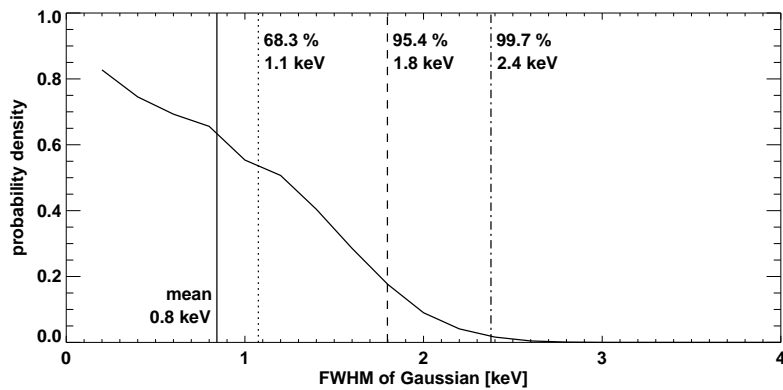
Figure 5.11: Probability distribution of the intrinsic width (full width at half maximum) of the inner Galaxy emission. This is the smoothed and enlarged version of panel 5 from fig. 5.10
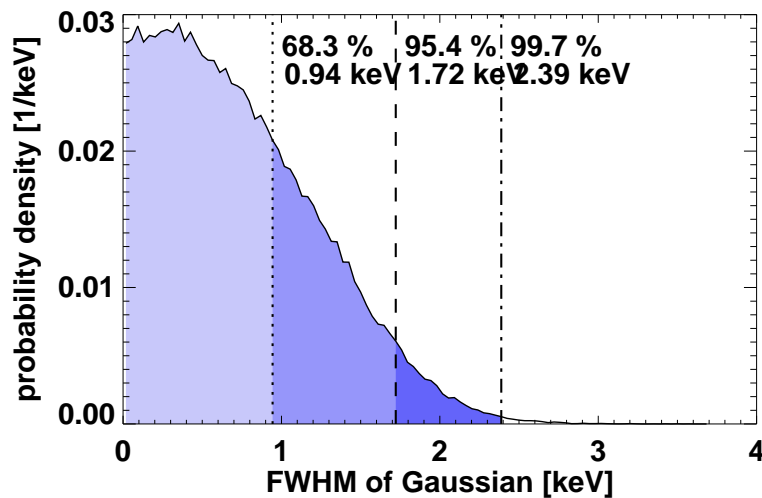
Figure 5.12: Probability distribution of the intrinsic width (full width at half maximum) of the inner Galaxy emission, based on data taken on INTEGRAL orbits 39–358

5.6.1 Comparing rotation curve models

The comparison of models predicting the ^{26}Al emission using the results from `spimodfit` model fitting can be performed using Bayes factors. These are ratios between the marginal likelihoods of the data assuming the validity of the different models.

In principle the process of comparing the different models would involve calculating the likelihood of the models as functions of the model parameters given the data. However, performing this task in a straightforward way, i.e. determining the likelihood from the model parameters by performing a Markov Chain Monte Carlo parameter estimation, would be computationally too demanding – some optimisation is required to be able to test the models on a practical amount of computer time. The approximations necessary for this stem to a large degree from the desire to be able to use the existing spatial model-fitting program `spimodfit` for the analysis. The method presently used to test models of the spatial and kinematic structure of the galactic ^{26}Al distribution is therefore a multi-step process.

In a first step, a model of the galactic emission corresponding to the chosen model parameters is built describing the intensity and velocity of the emitting matter as a function of position in 3-D space. This distribution is then projected into the field of view of the observer position of interest where it is described by intensity as a function of 2-D angular coordinates and radial velocity which corresponds directly to the observed line energy via Doppler shift. In the next step, this function is then mapped into a set of sky intensity distributions which are used as an input for a model fitting procedure that tries to reproduce the measured data with the simulated sky

models. In the first iteration only one map is used, one that integrates the whole 3-D space along the energy axis.

Besides the measurement of the width of the ^{26}Al emission from the inner Galaxy under the condition that the spectral shape of the measured spectrum is well modelled by the parametric line shape model used, the comparison of model spectra obtained from simulation is an alternative. The comparison of the measured spectrum obtained from model fitting with the reference spectra obtained from simulation is a hypothesis test without free parameters, compared to the previous case where the intrinsic width of the sky emission was a continuous parameter whose probability distribution was to be determined.

6 Probing ^{26}Al in the inner Galaxy with SPI

The SPI data on the ^{26}Al 1.8 MeV line from the inner parts of the Galaxy can be used in different ways to gain information about the distribution of ^{26}Al there. In this chapter I describe the two approaches I used. Both have in common that they are based on significant amounts of additional information put into the analysis process – therein lies the big difference when compared with the other survey methods discussed in Ch. 1, made necessary by the nature of the SPI observations.

6.1 Creating a l-v diagram from SPI data

The aim of this section is to describe how I approached the goal of creating a longitude-velocity diagram of the Galactic 1.8 MeV emission using SPI data. What we would like to obtain is a graphic comparable to those resulting from the gas surveys such as H_I and CO discussed in Sec. 1.2, i.e. to have spectra for a closely spaced set of galactic longitude intervals. There is, however a fundamental difference between the instruments involved with respect to their direction-sensitivity. When a radio telescope is performing a sky survey in the 21 cm line of neutral hydrogen, its angular response is a diffraction pattern. A large fraction of the power received comes from an area with a diameter of around 0.5° (Kalberla et al. 2005), which is of the order of the ratio between the wavelength being observed and the diameter of the radio telescope's reflector dish. This width is much less than the angular extent of the area on the sky to be surveyed. To cover the survey area, it needs to be sampled with multiple separate observations. Surveys of this type typically operate by pointing the telescope at a chosen position on the sky and tracking it to counter the earth's rotation. The resulting data point is an intensity averaged over the instrument's point spread function, also called beam pattern in radio astronomy. To be able to map the survey area with the spatial resolution the instrument permits, the separate sample locations of the survey should be spaced so that their PSFs overlap. Ideally according to the Nyquist sampling theorem, the spatial frequency of the sampling locations should be at least twice that of the PSF. In practice, because of time limitations often a coarser spacing is used, e.g. one approximately equal to PSF width, i.e. about double the spacing compared to the fully Nyquist sampled one. This procedure increases the uncertainty in the detection of compact sources because it is not possible to fully correct for the variation in detection efficiency depending on whether the source is in the central part of the PSF of a single sample or between two neighbouring samples. However, a doubling of the sample spacing means a four-fold

reduction in observing time for a given solid angle to be covered, which is worth the loss in precision for many applications. As a result, neighbouring observations on the sky are almost independent from each other in a typical radio sky survey.
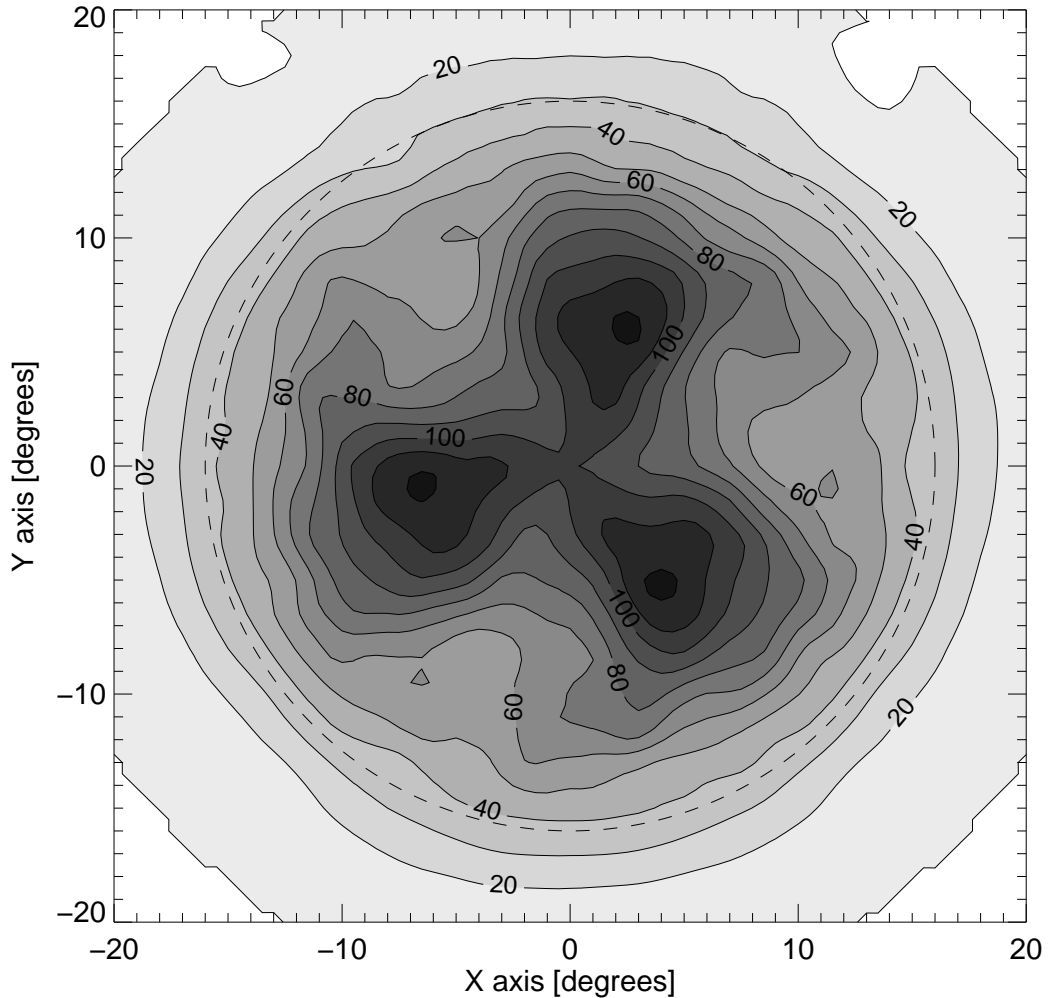
Figure 6.1: SPI angular response for the 1.8 MeV spectral region. The values of the contour labels are the effective area given in cm².

A survey using SPI observations or, coded-aperture observations in general, works differently. For a given pointing of the instrument, it is sensitive to photons coming from an area on the sky which is much larger than the angular scale of the instrument angular resolution. For SPI, this angular response fills a roughly circular area more than 30° wide, illustrated in Fig. 6.1. Neighbouring SPI pointings, typically separated

by 2°, are not independent. As discussed in Chap. 4, this is in fact required for the coded mask principle of operation, if the angular pattern of the source is not known in advance. An independent assessment of two or more distinct areas on the sky within the same field of view using data from a coded mask instrument is only possible in post-processing. It can be achieved by taking the whole of the relevant data, i.e. ideally all observations where some exposure fell into one of the target areas, and using model fitting to infer the parameters of a model describing the emission features of interest.

At this point, the choice of the model to be fitted to the data requires a trade-off to be done. To a first approximation, angular resolution has to be weighed against sensitivity. Sensitivity, in turn determines the uncertainties of the individual intensity measurements making up a spectrum, influencing the spectral resolution achievable when a spectral model is fitted to these data points. One extreme of this trade-off would be to divide the area to be surveyed into pixels with a size of half of the angular resolution of the instrument, which would lead to a pixel size of 1° for the case of SPI. In practice, such an approach has the drawback that there is a large amount of correlation between neighbouring areas on the sky, i.e. within the same field of view. This correlation, together with the uncertainty about the intensity of the instrumental background, makes the intensity determined for every individual pixel very uncertain, unless additional information is available. Such additional information can take the form of a smoothness criterion that limits the amount of freedom available for the differences in intensity between neighbouring parts of the image. Such criteria are applied together with image reconstruction methods such as the maximum entropy method or the multi-resolution expectation-maximisation (MREM) algorithm that have been used for the generation of images from SPI data. The other extreme is the case where only a single spatial distribution is used. This requires that the distribution be known in advance, which is commonly the case when observing a point source that has already been detected in other observations. There is a continuum of varying model complexity and number of model parameters between these extremes. The choice of a model that can adequately address the problem at hand is an important factor.

6.2 Model choice in model fitting

When the ^{26}Al intensity along the Galactic plane is modelled by a collection of neighbouring regions on the sky whose intensities are fitted simultaneously, the achievable precision is limited by the correlation between those areas. The fitting procedure solves the equation

$$D_j = b \cdot B_j + \sum_{i=1}^{N} a_i \cdot S_{ij} \qquad (6.1)$$

This attempts to reproduce the measured data D_j by the linear combination of the source detector patterns S_{ij} plus a background model B_j. Here i enumerates the collection of sky elements and j the data points used. j is shorthand for a combination

of pointing and detector ID. Since photon detection is a probabilistic process, this linear system cannot be inverted to yield a unique solution, but the fitting procedure gives an approximation to the joint probability distribution of the parameters $\{a_i, b\}$. The degree of precision achievable for the individual dimensions of this parameter vector is dependent on the amount of correlation between the corresponding source detector patterns S_{ij}. If these are all linearly independent, the situation is quite clear, because their only relation now is via the common scaling factor for the background model.
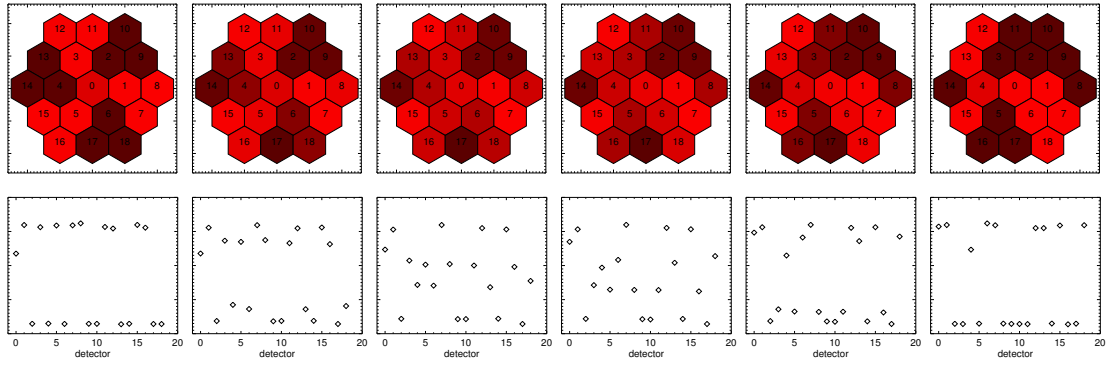
Figure 6.2: Exposure of the 19 SPI detectors in response to a point source that is moved in 0.4° steps along the horizontal axis.

Figures 6.2 and 6.3 illustrate this behaviour using the SPI response at the ^{26}Al line energy of 1 808.63 keV on a point source and on a Gaussian with a FWHM of 5°. The top rows in each figure show the detector hexagons shaded according to the incident radiative intensity, while the lower plot shows the same values in a form more suitable for quantitative reading. A 5° spatial scale is about the same scale as the DIRBE/COBE and COMPTEL surveys have as their spatial resolution.

If there is more than one source in the field of view, a procedure to disentangle the contributions of the N individual sources according to Eq. 6.1 therefore requires at least $N + 1$ measurements, providing they all sample different combinations of light sources and background intensity. Given perfect measurements, this would be sufficient to invert the equation. In practice, the measurements provide intensity values associated with an uncertainty giving a finite width probability distribution for the measured values. The influence this uncertainty in the individual intensity measurements has on the uncertainty of the reconstructed source intensities depends on the shadow patterns the sources make. In the case of the point source (Fig. 6.2), there are always several detector pairings with a large intensity contrast (≈ 10) between them. In the case of the extended source (Fig. 6.3), the intensity contrast is less than a factor of two for all detector pairings, a fivefold reduction in contrast. As a result, a given uncertainty in intensity measurement has a larger associated uncertainty in source position. A similar issue is at work when the intensity of several sources in the field of view is to
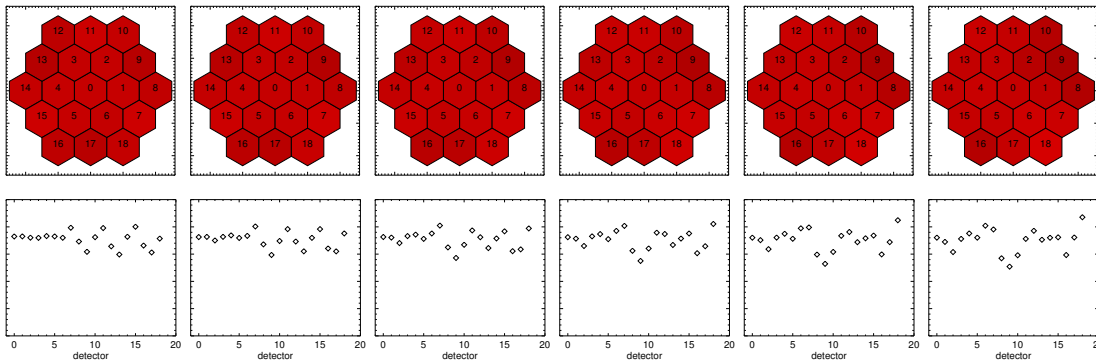
Figure 6.3: Exposure of the 19 SPI detectors in response to an extended source with a Gaussian intensity profile with a FWHM of 5° that is moved in 0.4° steps along the horizontal axis.

be reconstructed from the individual detector intensity measurements. The ability to distinguish between the contributions of the different sources depends on the difference in their detector patterns.

An example of a situation with a large amount of correlation is when two spotlights are pointed nearly at the same spot. Such a spotlight beam has a large intensity in the centre, which first slowly falls off with increasing angular distance, then more rapidly to near zero further out. As such, this behaviour is analogous to the angular response of SPI, when one averages over the effects of the mask pattern. The (azimuth-averaged) SPI response is nearly constant over the inner fully-coded field of view (FCFOV), then declines roughly linearly within the partially-coded field of view (PCFOV) to small values outside where only photons leaking through the shield can reach the camera. In such an arrangement, the intensities $F(\vec{x})$ at points near the centre of the illuminated area are only varying slowly with position and they are almost proportional to the summed intensities of the two light sources I_1 and I_2.

$$F(\vec{x}) = s_1(\vec{x})I_1 + s_2(\vec{x})I_2$$

Because of the slow variation with position close to the centre of the beam, the scaling factors s_1 and s_2 are nearly equal. In the case where we have the flux values measured at two places in this area available. Solving for I_1 and I_2 will provide us with a maximum likelihood solution, but the marginal probability distributions of the two intensity values will cover a large intensity interval because the measured fluxes are highly sensitive to the summed intensities but only weakly sensitive to their difference. This means that when looking at a set of particular flux measurements from inside the central region of the illuminated area, the flux could have been caused by each of the two spotlights with nearly equal plausibility. The measured flux at a given position can be expressed as a weighted sum of the sum of the two light source intensities and their

difference. In this example the scaling factor, i.e. the equations are nearly degenerate when solving for the intensities. In the case of model fitting analysis of SPI data, this translates into a large uncertainty in the individual model intensities when the two models used to fit the data result in very similar detector intensity patterns. This effect has to be kept in mind when choosing the spatial models to be used as model fit input.
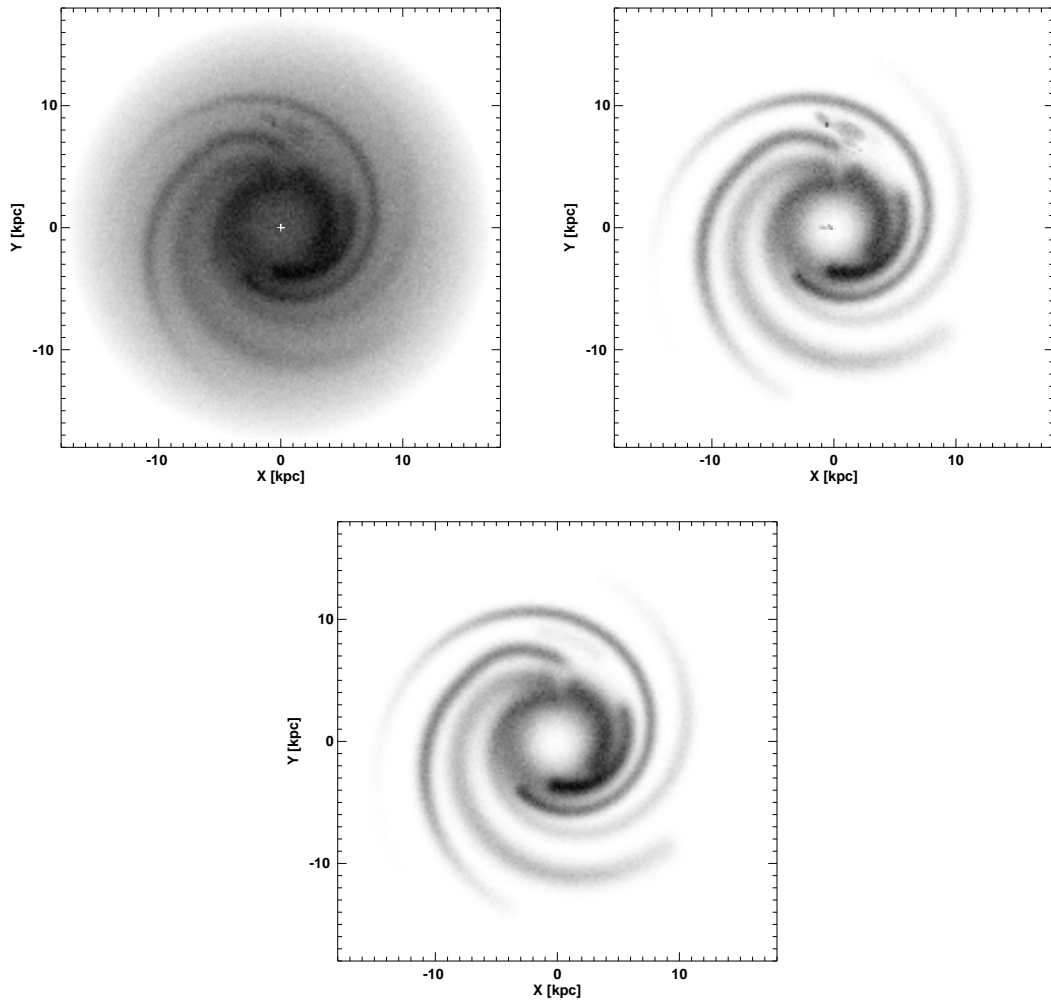
Figure 6.4: Area density of supernovae obtained from the NE2001 model by Cordes and Lazio (2002). Shown are three combinations of the seven components of the density provided: All components, all except the thick disk and the galactic centre, only thin disk and spiral arms.

6.3 The sliding-window method

Fitting a large number of separate spatial components concurrently, such as dividing the plane of the Galaxy in a large number of individual longitude bins also incurs a penalty in form of reduced precision of the individual intensity measurements. The reason for this loss is the increased level of correlation between the intensities in neighbouring sky bins as their angular size is reduced and more of them are within the field of view in any given SPI observation. It is desirable to minimise this penalty while still keeping the spatial resolution offered by the closely spaced samples in Galactic longitude.

The method applied towards this goal is based on the idea of measuring the variation of the parameter of interest with longitude not by trying to determine it for all individual longitude bins at once. Instead, the aim is to compare each one with the rest of the Galaxy sequentially, where only two models go into a fit at any given time. One takes a sky map of the area of interest, such as the COMPTEL 1.8 MeV map or one generated from a 3D model such as the NE2001 electron density model by Cordes and Lazio (2002). This is a three-dimensional model for the density of free electrons in the Galaxy that correlates well with the distribution of massive stars – they are the dominant source of ionisation. Fig. 6.4 shows maps of the density in the Galactic plane predicted by this model for three different selections from its sub-components. The method requires a 2D intensity distribution over the area on the sky to be surveyed. Specific knowledge of the 3D distribution of the source intensity is not needed, but can be used to generate a 2D model by projecting it on the 2D angular distribution as seen from the vantage point of the observer. The sky map is then divided into two maps whose summed intensity is the same as the intensity in the template map for every pixel. This can be done by making one map by taking the intensity values from the template for all pixels inside a (spherical) rectangle, and zero outside the rectangle and doing just the reverse for the other map.

These two complementary sky models are then used as input maps for the model fitting procedure from Cha. 4, using a linear combination of these models plus one or more background models. The result of the model fitting is a set of model scaling factors and the corresponding intensity values for each of the models and each energy interval. These spectra are in turn used as input for the spectral fitting procedure described in Cha. 5. This yields samples from the joint probability distribution of the parameter vector describing level and slope of the linear continuum and the area, central energy and intrinsic width of the line emission. From the samples, mean parameter values and their uncertainties are then derived. Plotting the line energy and its uncertainty as a function of the position in terms of Galactic longitude of the window then gives one data point for the position of the window used. The procedure is then repeated for different positions of the window, e.g. by varying the longitude coordinate of the window's centre. Taking all the data points obtained in this way and plotting them as a function of their galactic longitude generates an *l-v* diagram of the ^{26}Al emission under the assumption that the assumed sky distribution, in the form of the template sky map

used, matches the actual ^{26}Al distribution.

The first attempt at this procedure was done using a map generated from the NE2001 electron density model without the thick disk and galactic centre components. I divided the template map into two components using a spherical rectangle with a longitude extent of $20°$ and a latitude extent of $10°$. This window was placed in latitude with its centre located on the Galactic plane and successively shifted in longitude to integer multiples of $5°$ within a range of $-50° \leq l \leq 50°$ for the window centre. Fig. 6.5 shows the two input maps for the data point corresponding to the window being centred at $(l, b) = (20°, 0°)$.

Fitting those two sky maps together with a background model to the SPI data with `spimodfit` generates as output the two spectra shown in Fig. 6.6. The left spectrum shows the scaling factors for the map containing data for the outside of the rectangular window, the right spectrum those for the interior. Also shown are the spectral model functions corresponding to the mean of the parameter probability distributions, i.e. the "best fit" parameter values. Fig. 6.7 shows how the line centroid shifts as a function of window position in Galactic longitude. Table 6.1 shows the resulting line energies measured for all the window positions tested. The table also shows the energy relative to the reference energy that is determined by averaging the outermost three samples in both Galactic quadrants, (i.e. -50, -45, -40, 40, 45 and 50 degrees). In addition the radial velocity required to cause a Doppler shift of the respective value is shown.

Fig. 6.8 shows the radial velocities and their uncertainties obtained from fitting the `spimodfit`-produced spectra obtained from this combined "plug"-and-"hole" model fitting. The green data points for the "plug" maps are the same as those shown in table 6.1, the "hole" data points shown in red represent the corresponding part of the model map outside the window. The graphs both show a basically anti-symmetric behaviour with respect to the Galactic centre. The "plug" map, referring to the inside of the selection window, shows a red-shifted line in the first Galactic quadrant and a blue-shift in the fourth quadrant. The behaviour of the "hole" map, i.e. the sky *except* the selection window behaves in an opposite manner. This behaviour can be understood when thinking of the overall line position of the whole inner Galaxy. This position is a constant and therefore also the position of the overall spectrum obtained from model fitting should be constant, provided the model reproduces reality closely enough. The line shifts of both model components weighted by their intensities (not shown here) should sum to zero if the fitting procedure does not assign a significant part of the signal to the background model.

6.4 Results from the sliding-window fit method

To summarise, the sliding-window method of SPI line-position measurement consists of the following steps:

1. Choose a sky intensity map for use as a fitting template.
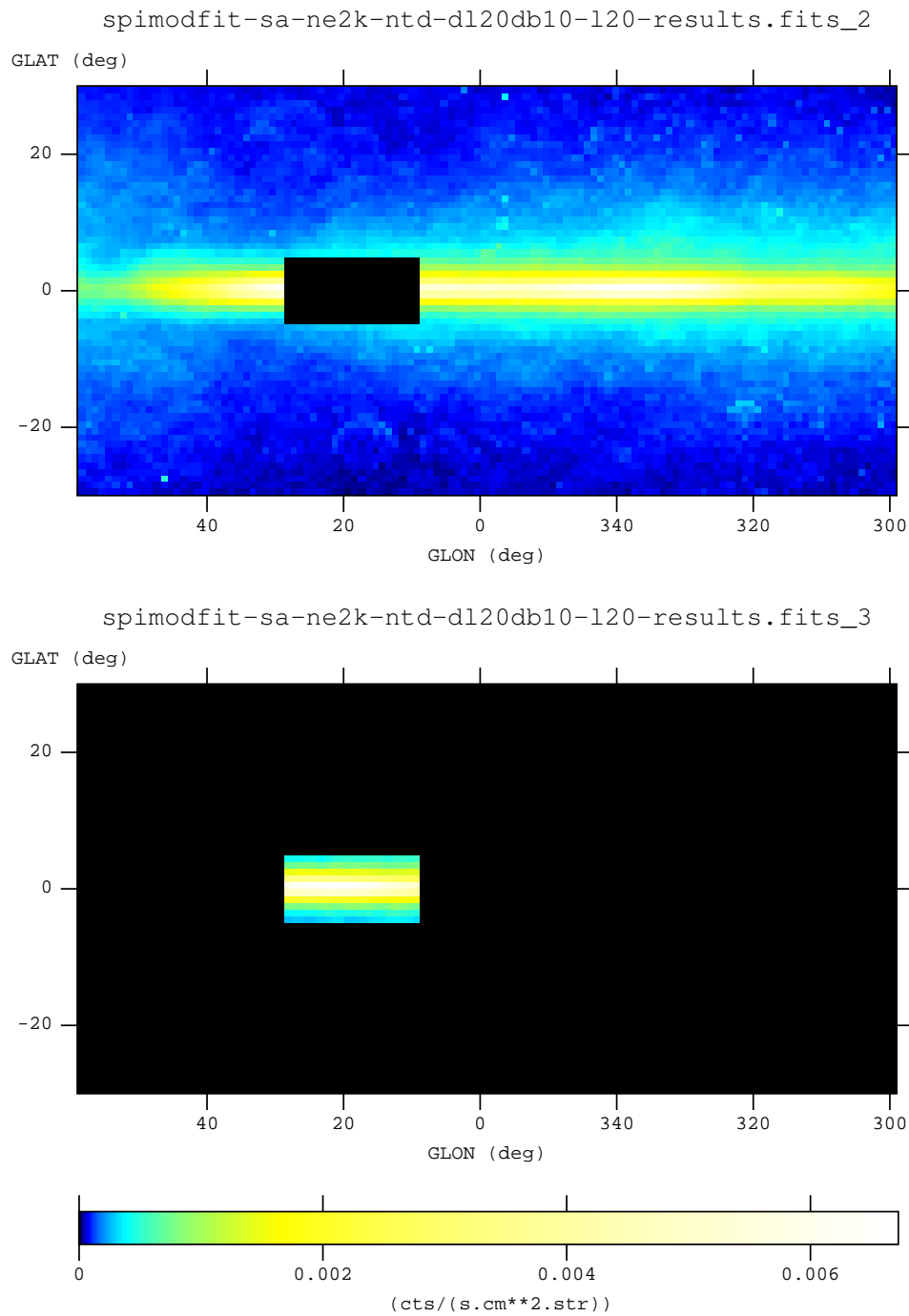
Figure 6.5: A set of complementary sky maps used as templates for model fitting. These maps were made from the NE2001 3D electron density model (Cordes and Lazio 2002) combined with SNR expansion. The extent of the rectangular window is $10° < l < 30°$, $-5° < b < 5°$
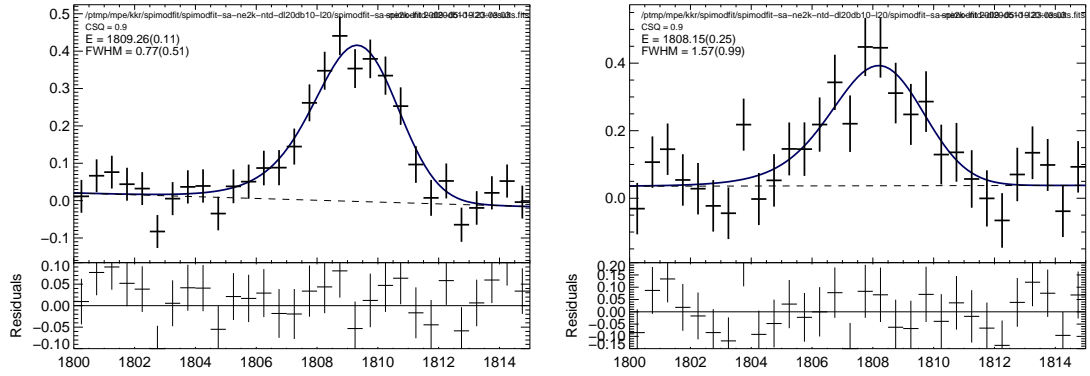
Figure 6.6: The two spectra obtained from model fitting using the sky maps shown in Fig. 6.5 as templates. Left: outside of the window, corresponding to Fig. 6.5 (top), right: window interior, Fig. 6.5 (bottom)
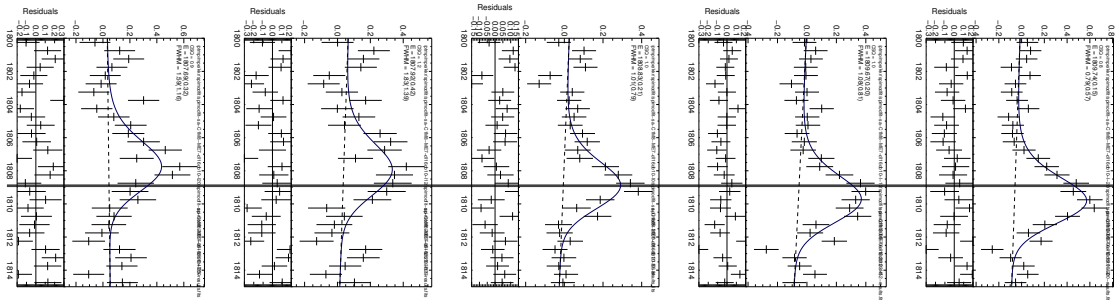
Figure 6.7: Five spectra obtained using the COMPTEL 1.8 MeV map and a $16° \cdot 10°$ window. The central longitudes are (from left to right): 20°, 12°, 0°, −12° and −20°.

Table 6.1: Measured line energies for the sliding-window fit using the NE2001 map and window dimensions of 20° in *l*, 10° in *b*. Relative energy and radial velocity are referred to the average of the outermost three measurements of the map without the window.

| longitude °	energy (abs.) keV	energy (rel.) −1 809.05 keV	rad. vel. km s^{-1}
−50	1 809.50±1.50	0.44±1.50	−73±249
−45	1 810.41±2.99	1.36±2.99	−224±496
−40	1 809.84±0.49	0.79±0.49	−130± 80
−35	1 809.78±0.25	0.73±0.25	−121± 41
−30	1 809.73±0.21	0.68±0.21	−112± 34
−25	1 809.69±0.18	0.64±0.18	−105± 30
−20	1 809.69±0.18	0.64±0.18	−105± 30
−15	1 809.57±0.19	0.51±0.19	−85± 30
−10	1 809.34±0.16	0.29±0.16	−48± 26
−5	1 809.11±0.17	0.06±0.17	−9± 27
0	1 808.90±0.16	−0.15±0.16	24± 25
5	1 808.76±0.16	−0.30±0.16	49± 26
10	1 808.58±0.18	−0.47±0.18	78± 30
15	1 808.36±0.18	−0.69±0.18	114± 30
20	1 808.15±0.23	−0.90±0.23	148± 37
25	1 808.02±0.25	−1.04±0.25	171± 41
30	1 808.04±0.37	−1.01±0.37	167± 60
35	1 807.93±0.49	−1.12±0.49	185± 81
40	1 807.86±0.80	−1.19±0.80	197±131
45	1 808.70±1.49	−0.35±1.49	58±247
50	1 809.45±2.86	0.40±2.86	−66±474
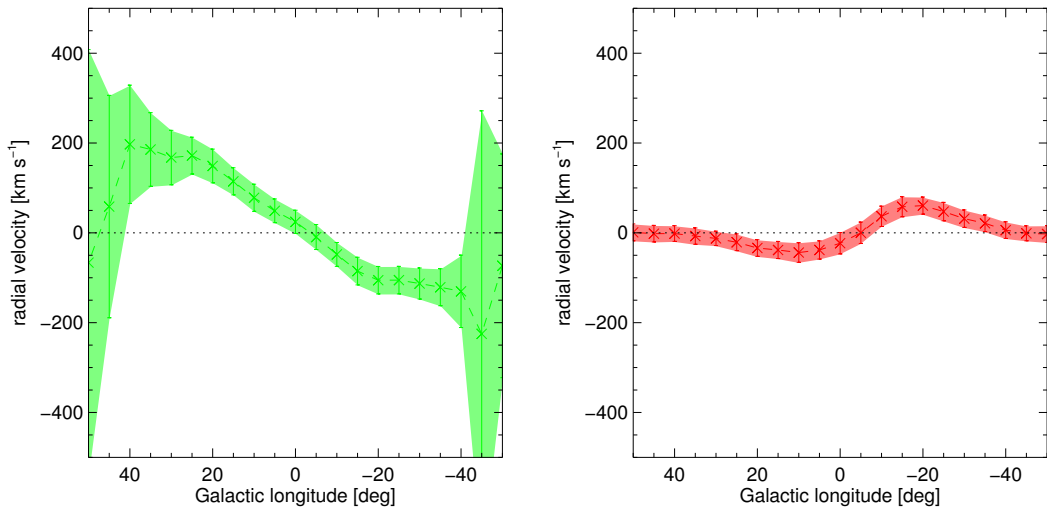
Figure 6.8: A collection of line position measurements with their uncertainties and the used longitude intervals shown for both the target map ("plug", left) and the rest of the sky model ("hole", right)

2. Choose a set of areas on the sky which shall be compared, e.g. spherical rectangles of equal size shifted by a fixed offset angle.

3. Divide the map into two complementary parts which sum to the original distribution for every one of these shapes.

4. For every map pair, fit the SPI data with the map pair and background model(s) for every energy bin in the spectral range of interest.

5. Combine the fitted intensity values into spectra for all the template maps.

6. Fit the spectral line shape in every spectrum, providing a line position measurement for every template map.

7. Plot the measured line position as a function of the corresponding sky area, e.g. the longitude of the rectangle centre.

There is a level of freedom in the choices to be done for the input to the sliding-window method. The first is the choice of the sky map to use as a template from which the two complementary maps used in the model fitting procedure are taken. The second choice is in the division of this very template map into two (or more) parts and how this division is changed from one data point in the *l-v* diagram to the next. These choices are a potential source of systematic uncertainties in the results. Because of this, I am investigating their influence in this section.
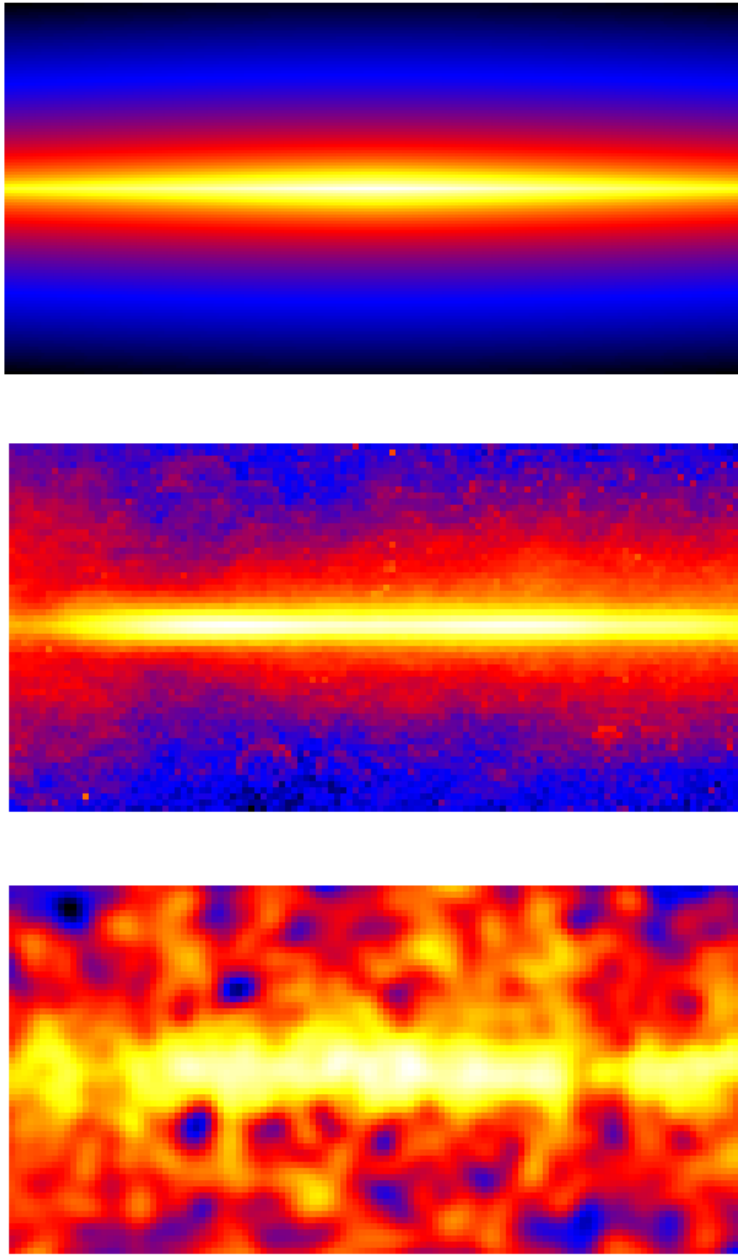
Figure 6.9: Three sky maps used for creating the template maps in the sliding-window method. The areas shown comprise $l \in [60°, -60°], b \in [-30°, 30°]$. From top to bottom: exponential disk $R_0 = 4$ kpc, $z_0 = 180$ pc, NE2001 without thick disk, COMPTEL 1.8 MeV maximum-entropy reconstructed image.

Table 6.2: Tested combinations of sky map template and longitude extent of the sliding window

map template	window size Δl [°]						figure
	6	8	10	12	16	20	
exponential disk	●	●	●	●	●	●	6.10
NE2001				●	●	●	6.11
COMPTEL 1.8 MeV	●	●		●	●	●	6.12

I tested three models of the angular distribution covering the solid angle of the inner Galaxy: an exponential disk, a map derived from the NE2001 electron density model and the COMPTEL 1.8 MeV sky map, reconstructed by the maximum-entropy method. The exponential disk with the parameters $R_0 = 4$ kpc, $z_0 = 180$ pc serves as a baseline because of its radial symmetry and the implied absence of small-scale features. These values have been found to fit the COMPTEL ^{26}Al measurements well. Compared to this are the electron density model with its spiral features as a tracer of massive star environments and the sky map based on the reconstructed angular distribution of the 1.8 MeV emission instead of a spatial model. Fig. 6.9 compares an inner Galaxy subset taken from each of these three sky maps. Table 6.2 shows the matrix of tested combinations of source map and sliding window size, together with the number of the figure showing the corresponding *l-v* diagrams.

The *l-v* diagrams (Figs 6.10-6.12) show that the common aspects in all of them are a basically anti-symmetric shape. It should be kept in mind that all radial velocities in these diagrams have been referenced to the average of the "hole" templates with large absolute longitudes of the hole, i.e. the large-scale average line position assuming the underlying template sky map to approximate the real angular distribution. The radial velocities relative to this average are in all cases compatible with zero in the direction of 0°, i.e. toward the Galactic centre.

With increasing separation from the Galactic centre, the absolute radial velocity increases monotonically for $|l| \leq 15°$. For even larger absolute longitude values, the behaviour differs depending on the template map used. For the NE2001 and COMPTEL maps, radial velocities remain at roughly the peak values, whereas for the exponential disk model, radial velocity values decrease for $|l| \geq 20°$. With further increasing $|l|$, the statistical uncertainties increase due to the decreasing ^{26}Al intensity at larger separations from the Galactic centre. This is shown in Fig. 6.13, where the lower signal-to-noise ratio of the spectrum taken with the window at $l = 40°$ is apparent.

In the inner part of the *l-v* diagrams, $|l| \leq 15°$, the relation between longitude and radial velocity is approximately linear with a small offset at $l = 0$, approaching a proportionality relation. The slope of this relation depends on the width (longitude extent) of the sliding window used. There is a monotonic relation with the slope
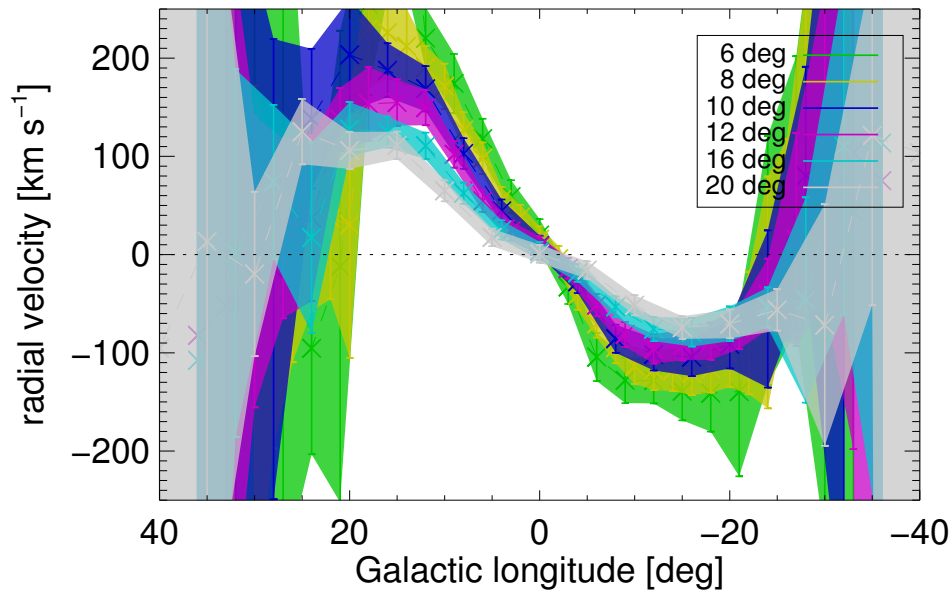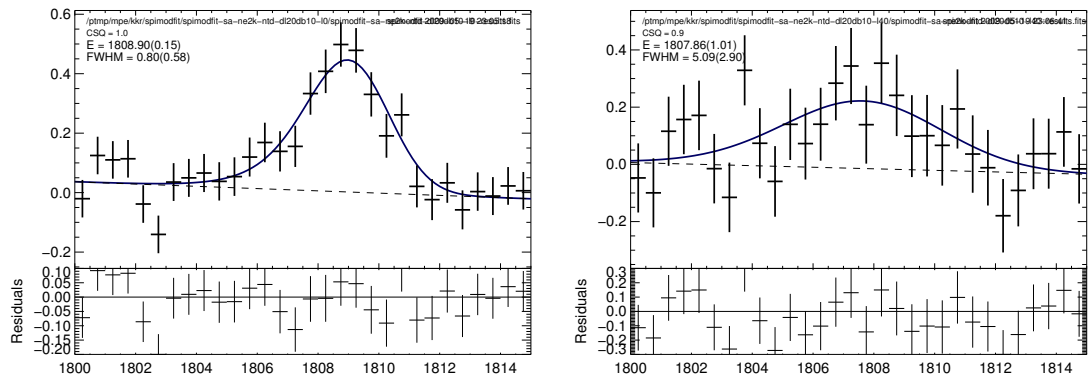
Figure 6.10: Sliding-window *l-v* diagram derived from the exponential disk model ($R_0 = 4$ kpc, $z_0 = 0.18$ kpc)
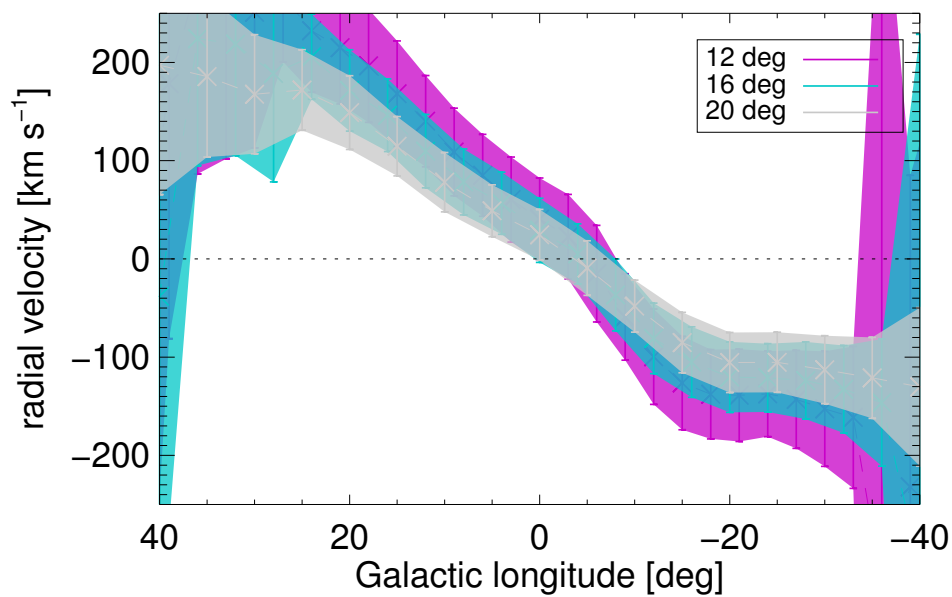
Figure 6.11: Sliding-window *l-v* diagram derived from the NE2001 electron density model (excluding the thick disk component)

91
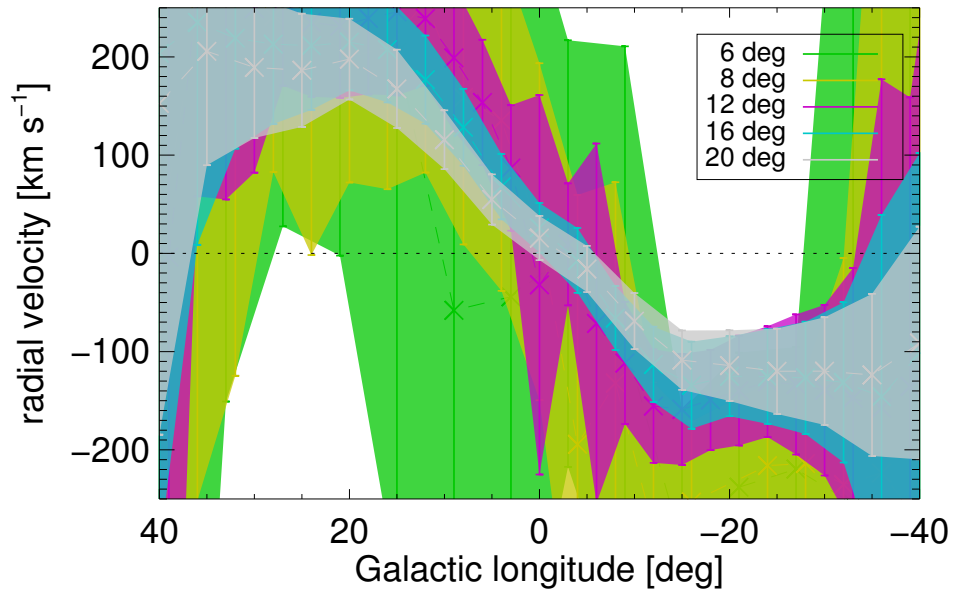
Figure 6.12: Sliding-window *l-v* diagram derived from the COMPTEL 1.8 MeV all-sky survey, using maximum-entropy reconstruction

Figure 6.13: Two spectra obtained from model fitting using the NE2001-derived "plug" sky maps, i.e. the inside of the sliding window. Left: window centred on $l = 0°$, right: window centred on $l = 40°$, window dimensions $l \cdot b$: $20° \cdot 10°$.

dv_R/dl decreasing with increasing longitude extent Δl.

For further discussion, I choose the *l-v* diagram based on the COMPTEL 1.8 MeV sky map with a window size of $16° \cdot 10°$ (Fig. 6.14) as a representative reference of the SPI results. The rationale for this choice is that the COMPTEL map is the closest approximation to the ^{26}Al 1.8 MeV distribution available. The 16° window size strikes a good balance between the loss of spatial resolution incurred for larger window sizes and the loss of velocity resolution that smaller window size choices are subject to.
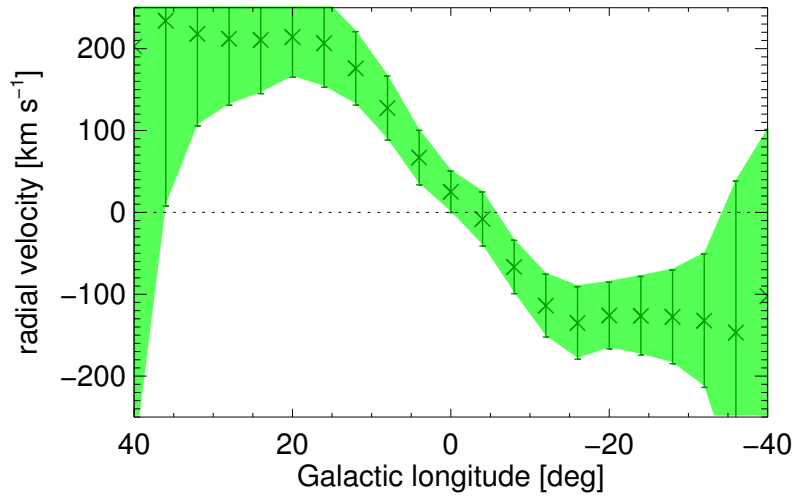
Figure 6.14: SPI *l-v* diagram using the COMPTEL 1.8 MeV sky map with a $16° \cdot 10°$ sliding window size.

7 Inference on Galactic structure with SPI data

7.1 Interpreting the ^{26}Al results

What inferences can be drawn from the sliding-window fit results? Any new information to be gained from an ^{26}Al *l-v* diagram will show up in the differences between it and *l-v* diagrams obtained by other observational methods, such as CO mapping or from simulations of Galactic dynamics. Fig. 7.1 is an example of the first case: comparing two different observed *l-v* diagrams. It compares the SPI *l-v* diagram with that seen in CO by Dame et al. (2001). The overall range of radial velocity values is comparable for both data sets. The CO data reaches larger absolute velocities (ranging from -260 km s^{-1} to $+280$ km s^{-1}, however it does so only in a narrow longitude range very close to the $l = 0$ line. This range where high CO radial velocities are observed is narrower than the window size used for the SPI data. Also, CO velocities within this range are resolved, while SPI velocities are averaged. Within $|l| \leq 20°$, the SPI curve lies within the area covered by the CO velocities, for larger separation from the Galactic centre, the absolute radial velocities of the SPI curve continue to increase, whereas the velocity range of the CO data decreases. The CO data shows a prominent feature reaching from $(30°, +100$ km s$^{-1})$ to $(-30°, -50$ km s$^{-1})$. The overall slope dv_r/dl of this feature, identified with the Molecular Ring by Dame et al. (2001), is clearly flatter than the slope seen in the SPI data. Fig. 7.2 compares the SPI *l-v* diagram with the distribution of SiO maser sources measured by Deguchi et al. (2004, see also Ch. 1.2.4). The maser distribution is similar to the molecular gas distribution as seen in CO. It does show a larger scatter, meaning that it intersects with the SPI *l-v* diagram over a larger interval of Galactic longitudes. The increased scatter compared to the CO data is due to the time scale of stellar evolution to the AGB stage where excitation of maser emission takes place. The longer time allows for greater divergence of the stars' orbits.

The other class of *l-v* diagrams that serve as references relative to the SPI observations are those derived from simulations incorporating the 3D structure and kinematics of the Galaxy. The simulations I have developed are well suited to build *l-v* diagrams corresponding to the SPI sliding-window data. Since the simulation results are in the form of spectra for pixels in galactic coordinates, it is easy to average the spectra of all pixels in a sky rectangle to obtain a spectrum for this area. The mean energy of this averaged spectrum then gives a data point for the radial velocity corresponding to the
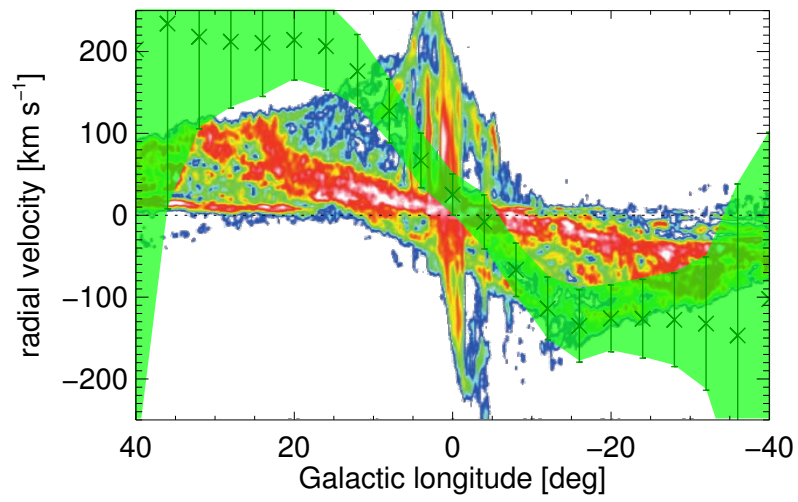
Figure 7.1: SPI *l-v* diagram (green) overlaid on top of the Dame et al. (2001) CO data.
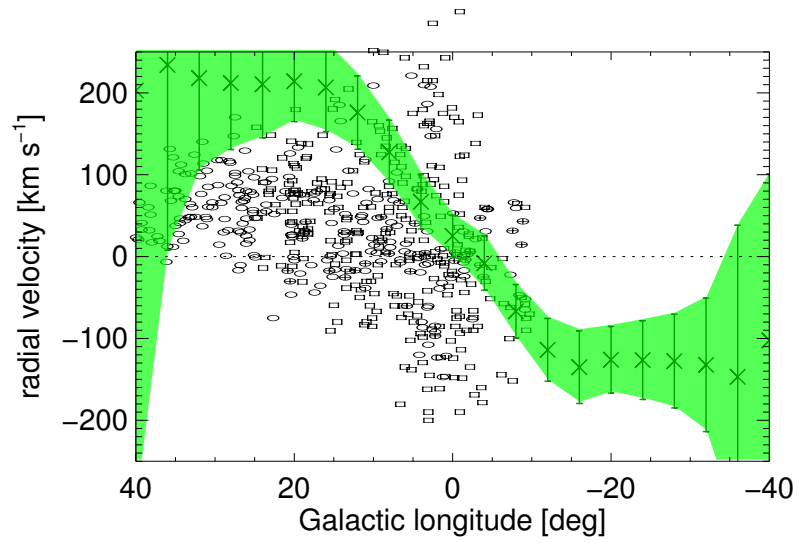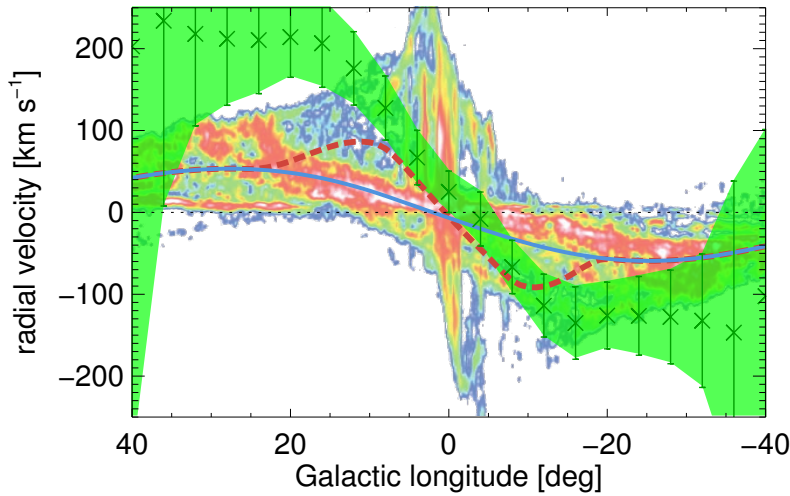

Figure 7.2: SPI *l-v* diagram (green) overlaid on top of the Deguchi et al. (2004) SiO maser data.

rectangular window.



Figure 7.3: SPI *l-v* diagram (green) and CO map with over-plotted simulation results:
blue solid line: thin disk and spiral arms from the NE2001 density distribution,
red dashed line: the same with an additional ellipsoidal bar model contributing
1/4 of the total luminosity.

Fig. 7.3 shows the result of such a comparison in the form of the blue solid line,
which is derived from the NE2001 spatial distribution combined with the Sofue et al.
(2009) rotation curve, assuming circular orbits. The simulated spectral data have been
averaged over the same $16° \cdot 10°$ sliding window as was used for the SPI model fits. In
the simulation, it is practical to use a smaller spacing for the grid of longitude values
sampled than in the model fitting procedure. The simulation *l-v* diagram uses a fine $1°$
spacing of window positions yielding a smooth curve. There is a close correspondence
between the average radial velocity of the NE2001 simulation and the prominent feature
in the CO data (identified with the molecular ring in Dame et al. (2001) for $|l| \leq 20°$.
The SPI *l-v* diagram has much larger radial velocities.

What could be the reason for these differences in v_r versus l? The motivation for
using the NE2001 model for the ^{26}Al spatial distribution is the physical link between
massive stars as sources of free electrons and ^{26}Al as well as the good correlation of the
NE2001 predecessor Taylor and Cordes (1993) with the COMPTEL 1.8 MeV morphology.
However, this model is biased by the lack of source data (pulsar sight lines) in the inner
part of the Galaxy. This means that the electron density there is poorly constrained
by pulsar measurements, and the authors therefore chose a model with a quite low
electron density inwards of $R = 2$ kpc.

The influence of this uncertainty with respect to the ^{26}Al density in the inner part

of the Galaxy can be tested by comparing models that vary the density in this region. One of simplest models that still fits the COMPTEL observations acceptably well is the exponential disk, which has its density maximum in the Galactic centre, decreasing exponentially with increasing galactocentric radius and distance from the Galactic plane. The comparison of a model fit using an exponential disk model with a radial scale length of $R_0 = 4$ kpc and a scale height of $z_0 = 180$ pc – the best fit values obtained with COMPTEL data – is shown in Fig. 7.4. Compared to the NE2001 model, the slope of both the simulated as well as the measured *l-v* diagram is compatible at the central $l = 0$ point and in a longitude interval of about $l \in [-10°, 10°]$. At larger absolute longitude values, the measurements show greater absolute radial velocities than the simulation. When the NE2001 density distribution is combined with an additional component that has a higher density in the inner Galaxy, the resulting slope of the *l-v* increases. This is illustrated by the red dashed curve in Fig. 7.3, where the average radial velocity of a model Galaxy consisting of a NE2001 component with circular orbits is combined with an ellipsoidal bar with a semi-major axis of 4 kpc length seen at an angle of 30°. The resulting curve shows a much steeper slope that matches that of the SPI *l-v* diagram in the $l \in [-10°; 10°]$ interval.
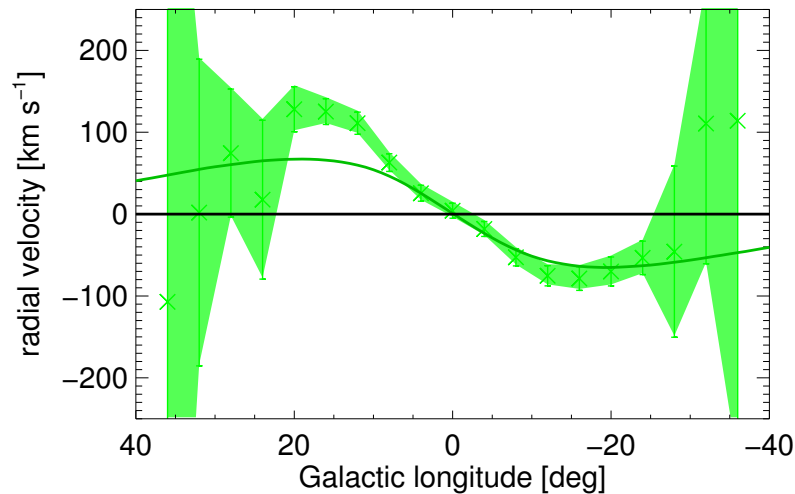

Figure 7.4: SPI *l-v* diagram (shaded) compared with simulation (solid line). Model fit input map: exponential disk simulation; Simulation: exponential disk combined with Sofue et al. (2009) rotation curve. Disk parameters: $R_0 = 4$ kpc, $z_0 = 180$ pc
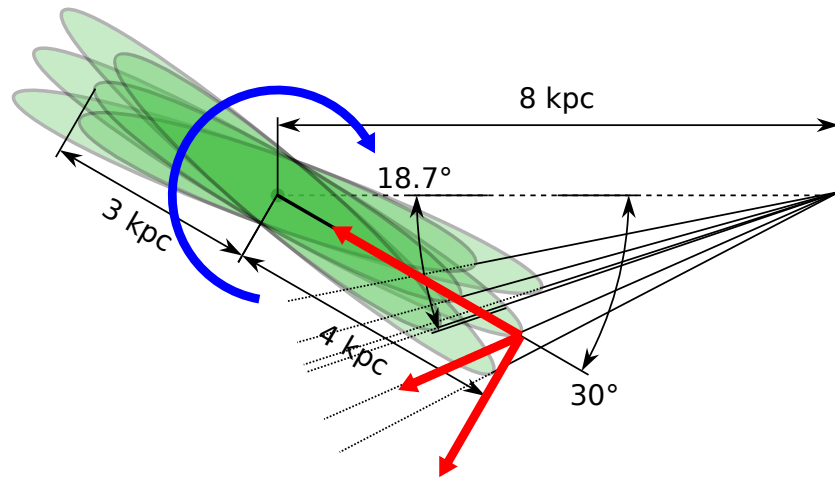
7.2 The Galactic bar

7.3 Evidence for a bar compared to a symmetric disk

In recent years, observational evidence has mounted indicating that the Milky Way is a barred spiral galaxy instead of a "normal" spiral galaxy as had previously been assumed. This evidence comes from radio observations and measurements of the near-infrared light distribution taken with the *COBE* microwave background mission. The case for a barred Milky Way has become much better supported with the results from the *Spitzer* infrared telescope (e.g. Benjamin et al. 2005), which allowed resolving individual stars instead of integrating their combined brightness over a 0.5 square degree solid angle as in the case of *COBE*/DIRBE. A bar is a structure that deviates strongly from cylindrical symmetry and has twofold rotational symmetry about the galaxy's axis instead. Since the Galaxy's rotation velocity does not rise proportionally to radius, such a structure cannot rotate as a solid body, where the constituent parts move along circular orbits with constant angular velocity. It is rather a pattern in the density distribution of stars, gas and dust that forms in a self-synchronisation process, with stars accumulating on a family of orbits which are closed in the rotating frame of reference synchronised with the rotation of the Galactic bar (Binney et al. 1991). These families of orbits, the x_1 and x_2 orbits in some cases show strong converging behaviour or even intersect themselves. When stars are on these orbits, this causes little effect because of the small probability of star-star interaction. It does have a large effect on gas, however. Gas clouds on strongly converging orbits interact by exhibiting shocks that dissipate kinetic energy, causing the gas to be funnelled toward the Galactic centre (Englmaier and Gerhard 2006).
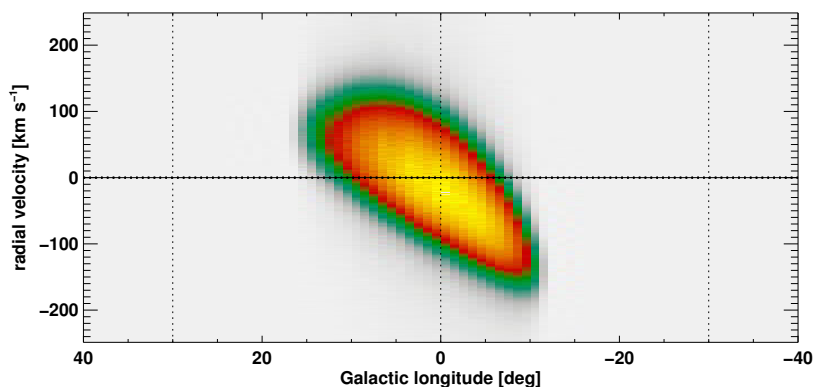
Figure 7.5: *l-v* intensity map of a model representing a bar structure with a triaxial ellipsoid shape with the semi-axis dimensions of $[3.0, 1.25, 0.25]$ kpc, oriented $30°$ off the direction to the Galactic centre

When examining the ^{26}Al line l-v diagram in Fig. 6.8, an east-west asymmetry in the longitudinal extent of the data points with non-zero Doppler shift can be seen. Compared with the behaviour that is expected from an ellipsoidal bar (Fig. 7.5) some similarities are apparent. Towards positive longitudes in the first Galactic quadrant, the region where the line position deviates noticeably from the rest value extends to greater absolute longitude than toward the fourth quadrant. At the same time, the maximum amplitude of the Doppler shift is smaller in the first than in the fourth quadrant. This feature is more readily visible in the plot for the complement to the target area, the "hole" curve in the right panel of Fig. 6.8, where the line position uncertainty at larger absolute longitudes $|l| > 30°$ is not as pronounced as in the plot for the target area itself.


Figure 7.6: Viewing geometry towards the Galactic bar for a variety of different bar parameters. The parameters shown cover semi-major axis dimensions of 3 kpc and 4 kpc, as well as angles toward the sun-Galactic centre line of 20°, 30° and 40°.

Fig. 7.6 shows the viewing geometry for different bar geometries, covering the range found in the literature. The radial velocities seen in ^{26}Al favour emission from sources beyond the ends of the 3 kpc bar in order to explain the observations at $|l| > 20°$ with the bar at angles of 40° or less.

Numerical simulations (Baba et al. 2010, Rodriguez-Fernandez and Combes 2008) as shown in Fig. 7.7 that implement both gravitational as well as hydrodynamic interaction of stars and gas show features that are very similar to the CO measurements, but differ from the ^{26}Al measurements at intermediate galactic longitudes ($|l| \in [15°; 30°]$). Since this type of simulation models cold molecular gas successfully, the ^{26}Al data supports the hypothesis that the time interval of approximately ten million years between the phase where the interstellar medium is visible as molecular clouds and the phase where ^{26}Al is visible is responsible for a substantial difference in kinematics. While the

simulation shows some gas clumps at radial velocities greater than 100 km s^{-1}, these only occur at low galactic longitude values. When the average radial velocity over a longitude interval of $\approx 10°$ is considered, corresponding to the window size used in the SPI analysis, the anti-symmetry about the galactic centre, combined with the sparseness of the high-velocity gas clumps compared to the bulk of them at $|v_r| < 100$ km s^{-1} results in moderate average values compatible with the ^{26}Al results for $|l| < 15°$.
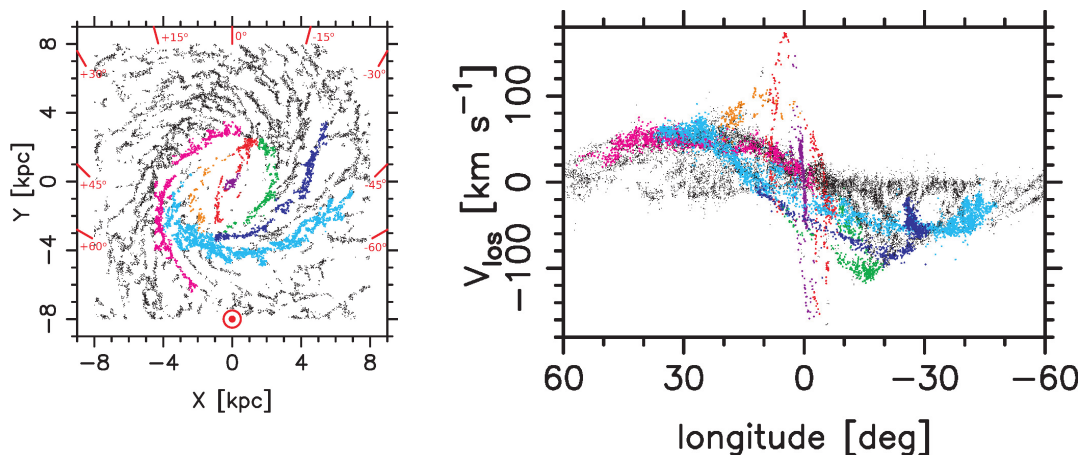
Figure 7.7: Simulated spatial (left) and kinematic (right) distribution of gas in a barred model of the inner Galaxy performed by Baba et al. (2010).

For larger absolute longitudes however, the discrepancy indicates an additional velocity component in the direction of Galactic rotation. Fig. 7.6 shows three arrows representing possible velocity vectors resulting in equal radial velocity components. The range of longitude values for the near end of the Galactic bar in this figure spans from 11.2° for 3 kpc semi-major axis and 20° bar angle to 27.5° for 4 kpc semi-major axis and 40° angle. Conversely, the longitude range $|l| \in [15°; 30°]$ corresponds to a range of galactocentric radii from 1.4 kpc (10° at 8 kpc distance) to 5.2 kpc (near end of a bar seen at $l = 20°$). This annulus about the Galactic centre contains the co-rotation radius, the radius where the angular velocity of the bar pattern equals the circular orbit angular velocity (Bissantz et al. 2003). Simulations of galaxy dynamics including star formation such as Wozniak (2007) indicate that this is the region with the youngest stellar populations, concentrated near the ends of the bar and in a ring structure just outside it.

The SPI ^{26}Al observations suggest that there is a process that favours ejection of matter released by massive stars in these regions in a preferential direction. A hypothesis explaining this phenomenon may be that the newly-formed stars move spin-ward in the rotating reference frame of the bar, compared to the region with high molecular cloud density where they formed. The time interval from star formation to the stage

where the largest amount of ^{26}Al is present in the interstellar material is of the order of $3\,\mathrm{Ma} \leq t \leq 10\,\mathrm{Ma}$ (Voss et al. 2009). With a rotation frequency of the bar of $60\,\mathrm{Ga}^{-1}$ (Bissantz et al. 2003) this is a time comparable to the rotation period of the bar, allowing for the stars to move away from the region of their formation as seen in the rotating reference frame of the bar. When these stars eject matter, they do so in a density gradient where there is a higher probability of material ejected away from the dense cloud complexes to stay at high velocity. Matter ejected towards the high-density region would have a lower mean free path before being decelerated, resulting in an asymmetric velocity distribution with respect to both the emitting stars as well as the parent molecular cloud structures.

8 Summary

Our knowledge about the detailed structure and history of our Galaxy is still patchy and unsatisfactory, major questions are still open and debated, such as number and location of spiral arms and the configuration of a central bar. The reason for this situation is that our position inside the Galaxy introduces biases in all observations. The most prominent example for this type of bias is the absorption of light by dust in the interstellar medium, others are the self-absorption occurring in the radio wavelength line emission or the circular dependency often occurring when in kinematics studies the distance of an object could be determined if its velocity was known, but the determination of the velocity vector depends on distance information.

It has become clear that the combination of observations in different wavelength regimes helps, because they are affected by different constraints. Limitations of a particular observable can be compensated by another. This is one motivation to study the inner Galaxy in the light of radioactive ^{26}Al, through its 1 808.63 keV gamma-ray emission. Because of the low interaction probability of MeV photons, absorption in the interstellar medium is not an issue. ^{26}Al originates in massive stars and becomes observable when it is ejected into the hot interstellar gas that surrounds the stars and is heated by their large energy output. Therefore, ^{26}Al gamma-rays carry information about a different stage in the cycle of matter operating in the Galaxy between star formation and the redistribution of stellar matter into the larger scale interstellar medium after the active phase of stellar evolution. Compared to the radio emission from molecular clouds coming from interstellar gas that is located chronologically before stars form, ^{26}Al probes mainly the phase $3 \text{ Ma} \leq t \leq 10 \text{ Ma}$ after a group of stars has formed. A difference between the behaviour of molecular gas and that of ^{26}Al-containing hot interstellar medium indicates what influence the star formation process has on the gas kinematics ("feedback").

The observations of Galactic ^{26}Al emission are performed with the germanium spectrometer SPI on ESA's INTEGRAL space observatory. Its detector technology provides an energy resolution of about 3 keV at the energy of the ^{26}Al emission line, so that the instrument is also capable of measuring the observed line energy to a precision of better 0.2 keV, corresponding to a Doppler velocity of 30 km s^{-1} when integrating the signal over a target solid angle of the order of 100 square degrees in size in the inner Galactic plane. This opens the possibility to measure the average radial velocity of the ^{26}Al in the hot interstellar medium as a function of Galactic longitude by scanning the plane with a set of these regions of interest. The result is a longitude-velocity diagram similar to those obtained from e.g. radio surveys such as the CO surveys in

the millimetre wavelength regime that trace molecular gas.

Our analysis method yields a range of *l-v* diagrams, each corresponding to a particular choice of sky model used as an input in the SPI data analysis procedure based on spatial model-fitting. We discuss the variety of prior assumptions and methodological parameter options to address the systematic uncertainties.

The resulting *l-v* diagrams are then compared with those of other inner-Galaxy observables in order to determine in which respects ^{26}Al behaves similarly to other observables and where there are differences. This shows that ^{26}Al absolute-average radial velocities increase steeper with increasing absolute Galactic longitude than is seen in other *l-v* diagrams, both from observations as well as from simulations. ^{26}Al reaches $|v_\mathrm{r}| > 100$ km s^{-1} for $|l| \in [10°; 30°]$. Comparison with the radial velocity map (Fig. 3.5) shows that velocities of this magnitude are incompatible with circular orbits. This suggests that the Galactic bar has a major influence on the kinematics of ^{26}Al in the inner Galaxy. Numerical simulations of gas flow in the potential of a barred spiral model for the Milky Way indicate that gas moves inward along the major axis of the bar toward the Galactic centre. Our aspect of the bar at an acute angle implies that we observe a large radial velocity component from this motion that increases the absolute observed radial velocity compared to the case of circular orbits. Simulations show that molecular gas accumulates in narrow spiral arms on the leading edge of the bar's stellar component. A scenario would be that the ejecta from the young stellar populations formed in these molecular cloud complexes are preferentially emitted in the direction of Galactic rotation, possibly as a result of interaction with nearby molecular clouds, thereby reaching the high radial velocities observed. The SPI ^{26}Al observations support such an understanding of star formation and feedback in the inner regions of our Galaxy.

Bibliography

D. Arnett. *Supernovae and nucleosynthesis. an investigation of the history of matter, from the Big Bang to the present.* Princeton series in astrophysics, Princeton, NJ: Princeton University Press, |c1996, 1996.

J. Baba, T. R. Saitoh, and K. Wada. On the Interpretation of the l-v Features in the Milky Way Galaxy. *PASJ*, 62:1413–, December 2010.

I. Baraffe, G. Chabrier, F. Allard, and P. H. Hauschildt. Evolutionary models for low-mass stars and brown dwarfs: Uncertainties and limits at very young ages. *A&A*, 382:563–572, February 2002. doi: 10.1051/0004-6361:20011638.

C. A. Beichman, P. C. Myers, J. P. Emerson, S. Harris, R. Mathieu, P. J. Benson, and R. E. Jennings. Candidate solar-type protostars in nearby molecular cloud cores. *ApJ*, 307: 337–349, August 1986. doi: 10.1086/164421.

R. A. Benjamin, E. Churchwell, B. L. Babler, R. Indebetouw, M. R. Meade, B. A. Whitney, C. Watson, M. G. Wolfire, M. J. Wolff, R. Ignace, T. M. Bania, S. Bracker, D. P. Clemens, L. Chomiuk, M. Cohen, J. M. Dickey, J. M. Jackson, H. A. Kobulnicky, E. P. Mercer, J. S. Mathis, S. R. Stolovy, and B. Uzpen. First GLIMPSE Results on the Stellar Structure of the Galaxy. *ApJ*, 630:L149–L152, September 2005. doi: 10.1086/491785.

J. Binney, O. E. Gerhard, A. A. Stark, J. Bally, and K. I. Uchida. Understanding the kinematics of Galactic centre gas. *MNRAS*, 252:210–218, September 1991.

N. Bissantz, P. Englmaier, and O. Gerhard. Gas dynamics in the Milky Way: second pattern speed and large-scale morphology. *MNRAS*, 340:949–968, April 2003. doi: 10.1046/j.1365-8711.2003.06358.x.

V. Borrel, B. Kandel, F. Albernhe, P. Frabel, B. Cordier, G. Tauzin, S. Crespin, R. Coszach, J. M. Denis, and P. Leleux. Fast neutron-induced damage in INTEGRAL n-type HPGe detectors. *Nuclear Instruments and Methods in Physics Research A*, 430:348–362, July 1999.

W. W. Campbell. Some stars with large radial velocities. *ApJ*, 13:98–99, January 1901. doi: 10.1086/140793.

D. D. Clayton. *Principles of stellar evolution and nucleosynthesis*. Chicago: University of Chicago Press, 1983, 1983.

J. M. Cordes and T. J. W Lazio. Ne2001. i. a new model for the galactic distribution of free electrons and its fluctuations. *astro-ph/0207156*, 2002.

T. M. Dame, D. Hartmann, and P. Thaddeus. The Milky Way in Molecular Clouds: A New Complete CO Survey. *ApJ*, 547:792–813, February 2001. doi: 10.1086/318388.

F. D'Antona and I. Mazzitelli. Cooling of white dwarfs. *ARA&A*, 28:139–181, 1990. doi: 10.1146/annurev.aa.28.090190.001035.

Klaus Debertin and Richard G. Helmer. *Gamma- and X-ray spectrometry with semiconductor detectors*. Elsevier, 1988.

S. Deguchi, T. Fujii, I. S. Glass, H. Imai, Y. Ita, H. Izumiura, O. Kameya, A. Miyazaki, Y. Nakada, and J.-I. Nakashima. SiO Maser Survey of IRAS Sources in the Inner Galactic Disk. *PASJ*, 56:765–802, October 2004.

R. Diehl, H. Halloin, K. Kretschmer, A. W. Strong, W. Wang, P. Jean, G. G. Lichti, J. Knödlseder, J.-P. Roques, S. Schanne, V. Schönfelder, A. von Kienlin, G. Weidenspointner, C. Winkler, and C. Wunderer. ^{26}Al in the inner Galaxy. Large-scale spectral characteristics derived with SPI/INTEGRAL. *A&A*, 449:1025–1031, April 2006. doi: 10.1051/0004-6361:20054301.

M. Elitzur. Astronomical masers. *ARA&A*, 30:75–112, 1992. doi: 10.1146/annurev.aa.30.090192.000451.

P. Englmaier and O. Gerhard. Milky Way Gas Dynamics. *Celestial Mechanics and Dynamical Astronomy*, 94:369–379, April 2006. doi: 10.1007/s10569-006-9003-3.

Richard B. Firestone et al. *Table of Isotopes*. Wiley-Interscience, 8th edition, 1996.

C. Francis and E. Anderson. Calculation of the local standard of rest from 20 574 local stars in the New Hipparcos Reduction with known radial velocities. *New Astronomy*, 14:615–629, October 2009. doi: 10.1016/j.newast.2009.03.004.

O. Gerhard. The Galactic Bar. In *ASP Conf. Ser. 273: The Dynamics, Structure & History of Galaxies: A Workshop in Honour of Professor Ken Freeman*, pages 73–+, 2002.

A. Heger. Stellar evolution. WWW, 2006. URL `http://www.ucolick.org/~alex/stellarevolution`.

J. Holmberg, B. Nordström, and J. Andersen. The Geneva-Copenhagen survey of the Solar neighbourhood II. New uvby calibrations and rediscussion of stellar ages, the G dwarf problem, age-metallicity diagram, and heating mechanisms of the disk. *A&A*, 475:519–537, November 2007. doi: 10.1051/0004-6361:20077221.

Edwin T. Jaynes. *Probability Theory, The Logic of Science*. Cambridge University Press, 2003. ISBN 0-521-59271-2.

P. M. W. Kalberla, W. B. Burton, D. Hartmann, E. M. Arnal, E. Bajaja, R. Morras, and W. G. L. Pöppel. The Leiden/Argentine/Bonn (LAB) Survey of Galactic HI. Final data release of the combined LDS and IAR surveys with improved stray-radiation corrections. *A&A*, 440:775–782, September 2005. doi: 10.1051/0004-6361:20041864.

B. Kandel, V. Borrel, F. Albernhe, P. Frabel, B. Cordier, G. Tauzin, S. Crespin, R. Coszach, J. M. Denis, and P. Leleux. Influence of temperature on the behaviour of INTEGRAL n-type HPGe detectors irradiated with fast neutrons. *Nuclear Instruments and Methods in Physics Research A*, 430:363–372, July 1999.

H. J. G. L. M. Lamers. Stellar Wind Theories. *Ap&SS*, 260:81–100, October 1998. doi: 10.1023/A:1001879004789.

M. S. Longair. *High Energy Astrophysics*. Cambridge University Press, 1994. ISBN 0521435846.

M.-M. Mac Low and R. S. Klessen. Control of star formation by supersonic turbulence. *Reviews of Modern Physics*, 76:125–194, January 2004.

W. A. Mahoney, J. C. Ling, A. S. Jacobson, and R. E. Lingenfelter. Diffuse galactic gamma-ray line emission from nucleosynthetic Fe-60, Al-26, and Na-22 - preliminary limits from HEAO 3. *ApJ*, 262:742+, November 1982.

G. Meynet, A. Maeder, G. Schaller, D. Schaerer, and C. Charbonnel. Grids of massive stars with high mass loss rates. V. from 12 to 120 $M_\odot$ at Z = 0.001, 0.004, 0.008, 0.020 and 0.040. *A&AS*, 103:97–105, January 1994.

J. H. Oort. Dynamics of the galactic system in the vicinity of the Sun. *Bull. Astron. Inst. Netherlands*, 4:269–+, November 1928.

N. J. Rodriguez-Fernandez and F. Combes. Gas flow models in the Milky Way embedded bars. *A&A*, 489:115–133, October 2008. doi: 10.1051/0004-6361:200809644.

D. Russeil, Y. M. Georgelin, P. Amram, Y. P. Georgelin, A. Laval, and M. Marcelin. Marseille Observatory H-alpha survey of the southern Galactic Plane and Magellanic Clouds. *Publications of the Astronomical Society of Australia*, 15:9–13, April 1998.

G. Schaller, D. Schaerer, G. Meynet, and A. Meader. New grids of stellar models from 0.8 to 120 $m_\odot$ at z = 0.020 and z = 0.001. *A&AS*, 96:269–331, dec 1992.

G. K. Skinner, P. von Ballmoos, N. Gehrels, and J. Krizmanic. Fresnel lenses for x-ray and gamma-ray astronomy. In *Optics for EUV, X-Ray, and Gamma-Ray Astronomy. Edited by Citterio, Oberto; O'Dell, Stephen L. Proceedings of the SPIE, Volume 5168, pp. 459-470 (2004).*, pages 459–470, February 2004. doi: 10.1117/12.508316.

Y. Sofue, M. Honma, and T. Omodaka. Unified Rotation Curve of the Galaxy – Decomposition into de Vaucouleurs Bulge, Disk, Dark Halo, and the 9-kpc Rotation Dip –. *PASJ*, 61:227–, April 2009.

L. Spitzer. *Physical processes in the interstellar medium*. New York Wiley-Interscience, 1978. 333 p., 1978.

J. H. Taylor and J. M. Cordes. Pulsar distances and the galactic distribution of free electrons. *ApJ*, 411:674–684, July 1993. doi: 10.1086/172870.

P. von Ballmoos, H. Halloin, J. Evrard, G. Skinner, N. Abrosimov, J. Alvarez, P. Bastie, B. Hamelin, M. Hernanz, P. Jean, J. Knödlseder, V. Lonjou, B. Smither, and G. Vedrenne. CLAIRE's first light. *New Astronomy Review*, 48:243–249, February 2004. doi: 10.1016/j.newar.2003.11.032.

R. Voss, R. Diehl, D. H. Hartmann, M. Cerviño, J. S. Vink, G. Meynet, M. Limongi, and A. Chieffi. Using population synthesis of massive stars to study the interstellar medium near OB associations. *A&A*, 504:531–542, September 2009. doi: 10.1051/0004-6361/200912260.

Alfred Weigert and Heinrich J. Wendker. *Astronomie und Astrophysik*. VCH, third edition, 1996.

S. E. Woosley, A. Heger, and T. A. Weaver. The evolution and explosion of massive stars. *Reviews of Modern Physics*, 74:1015–1071, November 2002.

H. Wozniak. The distribution of stellar population age in galactic bars. *A&A*, 465:L1–L4, April 2007. doi: 10.1051/0004-6361:20067020.

B. Zuckerman and P. Palmer. Radio radiation from interstellar molecules. *ARA&A*, 12:279–313, 1974. doi: 10.1146/annurev.aa.12.090174.001431.