Solar Flare Physics

During the past year we have been working with the HESSI (High Energy Solar Spectroscopic Imager team in preparation for launch in early 2001. HESSI has as its primary scientific goal photometric imaging and spectroscopy of solar flares in hard X-rays and gamma-rays with \( \sim 2'' \) angular resolution, \( \sim \text{keV} \) energy resolution and \( \sim 2\)-s time resolution over the 6-keV to 15-MeV energy range.

We have performed tests of the imager using a specially designed experiment which exploits the second-harmonic response of HESSI’s subcollimators to an artificial X-ray source at a distance of 1550 cm from its front grids. The figures of section I show the response to X-rays at energies in the range where HESSI is expected to image solar flares.

To prepare the team and the solar user community for imaging flares with HESSI, we have written a description of the major imaging concepts. This paper (Section II) will be submitted for publication in a refereed journal.
SECTION I
EXPERIMENTAL DETERMINATION OF THE HESSI MODULATION
THE HESSI GRIDLET TEST

In May 2000, at the Paul Sheerer Institute, where the HESSI telescope elements were installed and aligned, a $^{109}Cd$ source was used to test the modulation capability of the six coarsest subcollimators. The following briefly describes the test for each subcollimator. (A more complete report will be submitted later for publication.)

The $^{109}Cd$ source was supported on an XY table in front of the subcollimator at a distance of 1550 mm from the front grids. This distance is equal to the subcollimator length, and the equality is crucial to the experiment, since in this position, motion of a point source will produce 2nd-harmonic modulation of detected X-rays, equivalent to scanning a source at an infinite distance. Two X-ray detectors were mounted to the rear of the subcollimator, behind two masks which optimized the 2nd-harmonic performance.

The source had dimensions larger than the pitches of the grids, and so by itself no modulation could be expected, so a tertiary grid (called the "gridlet") equal in pitch to the two grids of the subcollimator, was used to collimate the X-rays into multiple beams incident on the telescope.

The figures on the following page show the response of the subcollimators to the motion of the source. In the plots for each subcollimator (1-6), there are two curves, one for each detector. Since the rays incident on each detector pass through different halves of the front grids, their output is a strong function of any relative twist of the front and rear grids. Small differences in phase between the profiles of the two detectors indicates small relative twists. The smallness of the twists (well below the specifications for good imaging performance) indicated that the HESSI imager was properly aligned and was ready for integration into the rest of the instrument package.
SECTION II
HESSI IMAGING PRINCIPLES
HESSI Imaging Principles

Edward J. Schmahl\textsuperscript{1}, Gordon J. Hurford\textsuperscript{2}, Andrew J. Conway\textsuperscript{3,4}, Brian R. Dennis\textsuperscript{2}, Richard A. Schwartz\textsuperscript{3}, Robert P. Lin\textsuperscript{2}, Jim McTiernan\textsuperscript{2}, David Smith\textsuperscript{2}, Chris Johns-Krull\textsuperscript{2}, and Andre Csillaghy\textsuperscript{2}, Markus J. Aschwanden\textsuperscript{4} and Thomas R. Metcalf\textsuperscript{4}

1. University of Maryland, Astronomy Department, College Park, MD 20742
2. Space Science Lab, University of California, Berkeley, CA 94720
3. Lab for Astronomy and Solar Physics, NASA Goddard Space Flight Center, Greenbelt, MD 20771
4. Department of Physics and Astronomy, The Open University, Milton Keynes MK7 6AA, UK.
5. Lockheed Martin Advanced Technology Center, Solar & Astrophysics Laboratory, Dept. L9-41, Palo Alto, CA 94304

ABSTRACT

The High Energy Solar Spectroscopic Imager (HESSI) will use rotational modulation synthesis for imaging hard X-ray and gamma-ray flares with spatial resolution of 2.3" and spectral resolution of ~ 1 keV. Like the Yohkoh/HXT hard X-ray telescope, HESSI relies on Fourier methods to produce images, but HESSI has many more sampled points in the Fourier plane, and is expected to produce higher-resolution maps with greater dynamic range.

With the intention of providing a heuristic or intuitive feel for how HESSI imaging works, we summarize the basic imaging hardware and the methods to be used for HESSI imaging. Using the bare minimum of mathematics, we explain crucial objects such as modulation profiles, modulation patterns, and visibilities. We describe the basic Back projection method and its relation to Fourier Transforms, and we outline four basic tools for image reconstruction: CLEAN, MEM, PIXON, and Forward Fitting, with simulated examples of each.
## Contents

1 Introduction

2 The HESSI Imager—A Brief Hardware Description 3
   2.1 The Subcollimators .......................... 3
   2.2 Decimation and Attenuation ..................... 4
   2.3 The Aspect System .......................... 4

3 Modulation Principles 5
   3.1 One-dimensional Modulation .................... 5
   3.2 The Modulation Pattern ....................... 6
   3.3 Rotational Modulation ........................ 7
   3.4 The Response Geometry ....................... 8

4 Visibility 9
   4.1 The (u,v) Plane ............................ 9
   4.2 From Count Rates to Visibility ................. 10

5 Image Reconstruction 12
   5.1 Back Projection and Fourier Transforms .......... 12
   5.2 CLEAN ....................................... 13
   5.3 Maximum Entropy Methods ...................... 14
   5.4 Forward Fitting ............................ 16
   5.5 The PIXON Method ............................ 16

6 Expected Imaging Performance 17
   6.1 Imaging Strengths and Limitations .......... 17
   6.2 Self Calibration ............................ 18
   6.3 Expected Evolution of Imaging ............... 18

7 References 18
1 Introduction

The High Energy Solar Spectroscopic Imager (HESSI) has as its primary scientific goal photometric imaging and spectroscopy of solar flares in hard X-rays and gamma-rays with ~ 2" angular resolution, ~keV energy resolution and ~2-s time resolution over the 6-keV to 15-MeV energy range. (Lin et al. 1993, 1994, 1998, Holman et al. 1997).

The only practical method of satisfying these requirements within the cost, mass, and launch constraints of a small satellite is to use Fourier-transform imaging. (See Prince et al. 1988 for a review of imaging techniques.) One of the most powerful of the Fourier family techniques is rotational modulation synthesis, first proposed by Mertz (1967) and implemented by Schnopper et al. (1968).

Rotating Modulation Collimators (RMC) were first constructed and used for solar flare X-ray imaging by the Japanese Hinotori telescope (Makashima et al. 1977, Ohki et al. 1982, Enome 1982), with angular resolution of 28" in the 20-40 keV energy range. The next hard X-ray solar imager was the Hard X-ray Imaging Spectrometer (HXIS), not a Fourier-transform imager, but important for developments in the ~ 3 - 30 keV range (Van Beek et al. 1980). Later, the Hard X-ray Telescope (HXT) on the Yohkoh satellite, used non-rotating Fourier synthesis with angular resolution of ~ 8" in the 20-100 keV energy range (Kosugi et al. 1991). A prototype of a rotation modulation telescope—the High Energy Imaging Device (HEIDI)—was flown in 1993 on a balloon, without successful flare imaging, but with a definitive test of a novel solar aspect system (SAS), proving 0.5" performance at balloon altitudes. (Crannell et al. 1994)

Among the new features of HESSI imaging are its high (2.3") angular resolution, its relaxed requirements for pointing (aspect knowledge vs. precision pointing), its wide field of view (full Sun), the wide range of energies that maps can be made in, and its ability to self-calibrate the amplitudes and phases of its imaging hardware.

The purpose of this paper is to summarize briefly the mostly new techniques that HESSI uses to exploit Fourier synthesis imaging. Our goal is to help the solar community get an intuitive feel for the concepts of HESSI imaging, in order to enhance the likelihood of its broad use for analyzing solar flares.

2 The HESSI Imager—A Brief Hardware Description

2.1 The Subcollimators

HESSI uses nine bi-grid collimators, each consisting of a pair of widely separated grids in front of an X-ray/gamma-ray detector. Each grid consists of a planar array of equally-spaced, X-ray-opaque slats separated by transparent slits. The slits of each pair of grids are parallel to each other and their pitches (p) are
identical, so the transmission through the grid pair depends on the direction of the incident X-rays. (See Figure 1) For slits and slats of equal width, the transmission is modulated from zero to 50% and back to zero for a change in source angle to collimator axis (orthogonal to the slits) of $p/L$ where $L$ is the separation between grids (1550 mm). The angular resolution is then defined as $p/(2L)$. For HESSI, the transmission of the source photons through the grids is modulated by the rotation of the spacecraft at about 15 revolutions per minute. The detector records the arrival time and energy of individual photons from anywhere on the Sun, allowing the modulated counting rate to be determined as a function of rotation angle.

Note that the detectors have no spatial resolution and hence have been optimized for high sensitivity and energy resolution. The nine segmented Ge detectors (GeDs), one behind each RMC, detect photons from ~3 keV to 15 MeV. The GeDs are cooled to ~75 K by a space-qualified long-life mechanical cryocooler. As the spacecraft rotates, the RMCs convert the spatial information from the source into temporal modulation of the photon counting rates of the GeDs. The instrument electronics amplify, shape, and digitize the GeD signals, provide power, format the data, and interface to the spacecraft electronics.

2.2 Decimation and Attenuation

Except in extreme circumstances, information on every photon is stored and then sent down. If an extended, extremely active period occurs, a decimation scheme is enabled which digitally decimates the incoming photon event stream, so that only a fraction of the events below an energy threshold get sent to telemetry. Both the fraction and the energy threshold are programmable. There are decimators for both the front and the rear segments of the detectors. The front decimator thresholds are controlled by software to avoid overflowing the spacecraft recorders, and the rear decimators can be controlled from the ground to limit the background rate.

In addition to decimation, there are mechanical attenuators, which if required, reduce the flux below the saturation level of ~50,000 events/s. Together, decimation and attenuation provide minimum loss of imaging and no loss in gamma-ray spectroscopy.

2.3 The Aspect System

HESSI exploits one of the fundamental efficiencies of Fourier imaging: precision spacecraft pointing can be traded for exact knowledge of pointing. With a Fourier imager, it is not necessary to stabilize the spin axis to any better than a few arc minutes, as long as the system gives sufficiently precise and high-bandwidth (i.e. well-sampled in time) pointing information. For HESSI, this information is provided by the solar aspect system (SAS) — a heritage of the HEIDI SAS — and the roll angle system (RAS). The SAS consists of 3 linear
diode arrays on which the full solar image is projected. The solar limb is determined at 6 positions every 10 ms, giving pitch and yaw to $\sim 1.5''$. The RAS consists of a star-imaging linear photodiode, providing roll angle to $2.7''$ every minute.

3 Modulation Principles

HESSI, like all RMCs, relies on temporal modulation by its grids to provide spatial information about incoming photons. The imaging process may be thought of as a superposition of elemental processes of photon passage through a subcollimator. Since the time of arrival of each photon is determined to 1 $\mu$s, and the aspect system determines the pointing of the telescope with arcsecond accuracy, the known geometry of the subcollimators provides well-defined information about the photon's origin.

3.1 One-dimensional Modulation

We illustrate (Fig. 1) how this works by considering an ideal subcollimator with two identical grids, each parallel to the $x$-$y$ plane, whose slats are, at a given instant, parallel to the $y$ axis, and whose spin lies parallel to the $z$ axis. The pitch (slit plus slat width) is denoted by $p$. Let a photon from an infinitely distant point source in the $x$-$z$ plane strike the front grid's plane at a point with coordinate $x$. The spatial phase of the photon relative to the slats of the front grid is $2\pi x/p$. The phase may be considered to be random since photons are arriving on parallel lines from the distant point source.

The probability of passage of the photon through the front grid is proportional to the projected area of the slits of that grid, as viewed from the direction of the photon. For an ideal grid of finite thickness and infinite extent, this value is independent of the $y$ direction, but dependent on the photon direction in a fashion characteristic of the grid thickness and the slit-slat ratio.

In the case of a single thin grid with equal slits and slats, the projected open area, plotted as a function of incidence phase, is a periodic, triangular waveform with equally shaped maxima and minima. (Fig. 2) For a thick grid, when sufficiently off axis, there is internal shadowing by the sides of the slats, and the projected open areas are reduced from the thin case, so the area waveform is still periodic, but with flat minima. (See third plot in Fig. 2.) In the case of HESSI grids, the projected slit-slat ratios are optimized to have equal projected slit and slat widths at incidence angles of about 412''. For HESSI subcollimators 1-6, the aspect ratio of the slats ($x$-width/$z$-thickness) is 1:50. For photons with nearly normal incidence, the effective slit/pitch ratio is 0.6, and for photons incident at angles of 412'' in the $x$-$z$ plane, the effective slit/pitch ratio after shadowing is close to 1:2.
For slit/pitch ratios equal to 1/2, and for energies low enough that the slats are opaque ($E < 100$ keV), and for energies high enough that diffraction effects are negligible ($E > 6$ keV), the effective subcollimator area is approximately a triangle waveform function of phase. Being periodic, it is useful and efficient to characterize a fractional area function ($S$) in terms of the first few harmonics of its expansion as a Fourier cosine series:

$$S_{\text{triangle}}(\Phi) = \frac{1}{4} \cdot \left(1 + \frac{8}{\pi^2} \cos(\Phi) + \frac{8}{9\pi^2} \cos(3\Phi) + \frac{8}{25\pi^2} \cos(5\Phi) + \cdots \right)$$  (1)

where $\Phi = 2\pi x_r/p$ is the photon's phase with respect to the rear grid. $x_r$ is the $x$-coordinate of where the photon struck the rear grid plane, i.e. $x_r = L \tan \theta$, where $\theta$ is the angle of incidence in the $x$-$z$ plane, and $L$ is the distance between front and rear grids.

This is precisely the form used by Schnopper et al. (1970) for their observations of galactic-center with a rocket-borne RMC. But for HESSI, the even harmonic terms are usually not negligible, and the Fourier coefficients are not those of a pure triangle waveform, so we must use the more general form:

$$S(\Phi) = T \cdot \left(1 + a_1 \cos(\Phi - \Psi_1) + a_2 \cos(2\Phi - \Psi_2) + a_3 \cos(3\Phi - \Psi_3) + \cdots \right)$$  (2)

where $T$ is the fractional collimator transmission, which for ideal grids would equal 1/4, but in general for HESSI, which has thick grids of differing X-ray opacity, $T$ is a slowly-varying function of map position and energy. In practice, we keep terms only up to the 5th harmonic. In general, for real conditions, the relative amplitudes $a_1, a_2, \cdots$ and the relative phases $\Psi_1, \Psi_2, \cdots$ are dependent on the grid characteristics, the roll angle, and the angle of incidence. The latter dependence is sufficiently small and slow that for sources on the Sun, it can be considered uniform within a map in a given time bin. These parameters have been computed from the grid characterizations, and are included in the HESSI reconstruction algorithms.

### 3.2 The Modulation Pattern

If we consider a photon not restricted to traveling in the $x$-$z$ plane, we find that the above theory is still applicable if $\theta$ is the angle between the $z$ axis and the photon's trajectory projected onto the $x$-$z$ plane. Let us now imagine a source ($x$-$y$) plane at a large distance $D$ from a non-rotating grid ($D \gg L$).

If a point source in the source plane is moved in the $x$ direction, the detector will show a count time series that has the triangular waveform. In fact simple geometry shows that the phase $\Phi$ will be proportional to $x$, ($\Phi = 2\pi L/pD$) so the observed count profile will be identical in form to the area function $P(\Phi)$. If the source is moved in the $y$ direction, then the counts will remain constant. From this it is clear that HESSI's spatial response to a brightness distribution
in the source plane would resemble a series of stripes in the $y$ direction. A cross-section perpendicular to stripes in the $x$ direction, for any $y$, would be identical in form to the fractional area function $S(\Phi)$. This response is called the modulation pattern.

If we describe the brightness distribution in the source plane as a pixel image $F_m$, where $F_m$ is the photon flux (photons cm$^{-2}$ s$^{-1}$) from pixel $m$ incident on the HESSI's front grids, then the following equation shows how to use the modulation pattern to obtain the expected count rate $C_i$:

\[
   C_i = A \sum_m P_{im} F_m \Delta t_i
\]  

(3)

In this notation, $P_{im}$ is the probability that a photon originating in a pixel $m$ and incident on the front grid will be counted in the $i^{th}$ time bin during interval $\Delta t$. Except for a possible normalizing constant, $P_{im}$ is the modulation pattern, and is related to the area fraction function $S(\Phi)$ by the equation $P_{im} = S(\Phi_m)$ where $x_m$ is the $x$ coordinate of the center of pixel $m$ in the source plane. For notational convenience we introduce $K = 2\pi L/pD$ so that the phase $\Phi_m = K x_m$.

### 3.3 Rotational Modulation

So far we have described the area fraction only for one roll angle. As the spacecraft spins, the projected modulation pattern spins with it. To the lowest order of approximation, if the grids are infinitely wide and infinitely thin, and the spin axis does not move, the modulation pattern would rotate unchanged. Since, however, the grids are thick, of finite extent, and the spin axis is not necessarily constant during a rotation, the projected modulation pattern changes with roll angle in a predictable way. The result is a slow, smooth change in the parameters defining the modulation pattern: the amplitude coefficients $a_n$, and phases $\Psi_n$.

It is useful to take the telescope as our reference frame, and let a point source move on a circle (roll angle $\alpha = 0 - 2\pi$) in the source plane. For an ideal RMC (idealized grids, steady spin), the argument $\Phi$ in equation (2) is

\[
   \Phi = K x = K r \cos \alpha
\]  

(4)

where the source is at radial position $r$, and the wavenumber of the detector is $K = 2\pi/p$.

Note that $\cos \alpha$ is present because the modulation described by equation (2) only depends on the $x$ position of the source. The effect of this factor can be regarded as changing effective pitch $p \sec(\alpha)$. This in turn will cause the the $a_n$ and $\Psi_n$ parameters to vary, with the fundamental coefficients $(a_1, \Psi_1)$ changing the least.
The number of photons \( N \) falling on a HESSI detector from a point source a distance \( r \) from the axis in the source plane is proportional to incident flux \( F_0 \) and the effective area:

\[
N(\alpha) = F_0 AT \cdot (1 + a_1 \cos(Kr \cos(\alpha) - \Psi_1) + \ldots)
\]  

(5)

As HESSI rotates, the roll angle \( \alpha \) will vary smoothly with time, and so a plot of \( N(\alpha) \) with time will have the profile shown in Figure 3. This is the modulation profile for a point source if the spin axis is fixed.

For a given map center, the quantities \((a_n, T, \Psi_n)\) are computed from the aspect system and the grid response matrices. The above function is computable for any time interval and any point source on the Sun. Therefore, since the signal expected for a superposition of sources is the superposition of point source signals, any model source can be used to predict a modulation profile. This is the basis for all HESSI reconstruction algorithms and underlies equation (3).

In general, if detector \( k \) is at roll angle \( \alpha_i \), and the point source has position vector \( (x_0, y_0) \) in the source plane, with polar equivalent \( (r_0, \phi_0) \), then the generalization of the phase \( \Phi \) in equation (4) is:

\[
\Phi_{ik} = K_i \cdot r = (K_k \cos \alpha_i) x_0 + (K_k \sin \alpha_i) y_0 = K_k r_0 \cos(\alpha_i - \phi_0)
\]  

(6)

The set of photon arrival times taken from the HESSI telemetry can be histogrammed to yield a raw count rate profile. The HESSI user can select an almost arbitrary set of time bin sizes \( \{ \Delta t_i \} \) for histogramming. \( \Delta t_i \) is necessarily \( \geq 1 \mu s \), and may be as large as \( \sim 1 \) s, but is more typically \( \sim 0.5 \) ms.

After the time bin selection, the count/bin profile becomes a function of roll angle \( \alpha \), and one can proceed to calibrate the count rates. The size of the time bins depends on the detector, with the finest detectors (smallest \( p \)) requiring the smallest time bins.

### 3.4 The Response Geometry

For phase and amplitude calibration, that is determining \((a_n, \Psi_n)\), one must first select a center for the mapping. Figure 4 shows the contours of the subcollimator phases and map center at one instant of rotation. The lines of constant phase for a subcollimator of angular pitch \( p \) are separated by \( 2\pi R/p \) radians, the map being centered at a distance \( R \) from the spin axis. The phase at map center rises and falls in a quasi-sinusoidal way (precisely sinusoidal if there is no drift or wobble in the pointing) during the spacecraft rotation.

The selected map center should be close to the centroid of source of emission. This point can be found in several ways (coarse mapping, parametric fitting, context observations or using previous flare position).

In general, one must incorporate the live time \( \tau_{ij} \), dependent on roll angle and energy \( E_j \), the detector number \( k \), and the subcollimator transmission \( T_{ijk} \), as well as the modulation amplitudes \( a_{ijk} \) and modulation phases \( \Psi_{ijk} \) to determine the number of photons falling on the detector \( N_{ijk} \):
\[ N_{ijk} = F_0 \cdot \{ \tau_{ijk} T_{ijk} [1 + \sum_n a_n^{ijk} \cos(n(\Phi_{ik} - \Psi_{ik}^{ijk}))] \} \]  

where \( F_0 \) is the incident photon flux on the front grid. When the HESSI analysis software begins image reconstruction it first computes the quantities \( \tau_{ijk}, T_{ijk}, a_n^{ijk} \) and \( \Psi_{ik}^{ijk} \) for the appropriate map position, energies, and bin sizes.

4 Visibilities

The concept of visibilities is borrowed from radio astronomy. In the context of HESSI 'counts' and 'visibilities' can be regarded as two time series that present the same information: the response of HESSI to a brightness distribution.

In radio astronomy the response of an interferometer to point sources in the sky, known as the fringe pattern, can be quantified in terms of complex visibilities, which are the components of a discrete Fourier transform of the sky. Thus far we have considered the response of HESSI to a point source as a modulation profile (Figure 3) expressed as a time series of counts. The modulation profile can be obtained from the brightness distribution using the modulation patterns, which we have represented in terms of a cosine and a series of harmonics. Clearly HESSI imaging has much in common with radio astronomy interferometry and could benefit from the enormous efforts invested in that subject. Also, we will see that the use of visibilities with HESSI has several intrinsic advantages. To exploit this resource it is necessary to consider what a visibility means in relation to counts in the context of HESSI.

4.1 The \((u,v)\) Plane

In radio astronomy the visibility of a point source of flux \( F_0 \) at coordinate \((x_0, y_0)\) is defined to be the Fourier transform:

\[ V(u, v) = F_0 e^{2\pi i (ux_0 + vy_0)} \]  

The variables \((u, v)\) are the coordinates in the Fourier plane and the 2D array \( V(u, v) \) is the Fourier transform of the flux distribution in the \(x-y\) source plane, in this case containing a single point source at \((x_0, y_0)\). It turns out that detector \( k \), while HESSI is at roll angle \( \alpha_i \), will sample a quantity closely related to the visibility transform \( V_{ik} = V(u_{ik}, v_{ik}) \) at \( u_{ik} = \cos \alpha_i / p_k \) and \( v_{ik} = \sin \alpha_i / p_k \).

So, each observation (count integrated over a time-bin) that HESSI makes is a sample upon one of 9 possible circles, one circle for each detector. Samples on the largest circles correspond to the highest spatial frequencies, and come from detectors with the small pitches. In one rotation, for a particular detector, HESSI's observations form a time series of count values, where each time corresponds to a particular roll angle. This time series is a set of sample points...
spread across the entire circumference of one of the 9 circles. In this way the count time series can be related to the 2D Fourier transform of the source flux distribution. Conversely, the visibilities which are the samples of the 2D Fourier transform, can be regarded as a time series in the context of HESSI.

If the spin axis is fixed so that \((a, \Psi_n)\) are constant, if background is negligible, and if the transmission \(T\) and livetimes \(\tau\) are constant, then the relation between the fundamental term in the counts and visibilities can be deduced from equations (6), (7) and (8):

\[
X_{ik} \equiv \text{Real} \left[ V_{ik} \right] = \frac{C_{ik} - F_0 \tau}{T \tau a_1} = F_0 \cos \Psi_{ik}
\]

In this idealized case, the counts are the cosine transform of the image. The imaginary part of the visibility is the corresponding sine transform. Fig. 3 has already shown the real part of the visibility time series, i.e. the visibility modulation profile, for a point source. The dashed lines in that figure show the imaginary part of the visibility (offset by 0.5, since their mean must be zero). The case of a moving spin axis, important for constructing visibilities from observed counts in practice, is dealt with in Section 4.2.

The importance of visibilities to HESSI is that they are an intermediate between the observed counts and images. The visibilities have been corrected for all of the spin-axis wobble, slit-shadowing, transmission and deadtime effects. Consequently, visibilities, unlike count-rate profiles, can be added from one rotation to another.

From many simulations the advantages and some disadvantages can be summarized as follows: (1) The aspect correction can easily be applied while constructing visibilities from the observed counts. (2) Due to 1, visibility modulation patterns do not need to be corrected for the aspect solution for each time bin. (3) Due to 2, operations involving modulation patterns can be performed with a convolution based on Fast Fourier Transform methods, rather than a matrix multiplication. This allows a speed up of order \(M^2\) as compared to \(M \log M\), where \(M\) is the number of time bins. (4) Due to 2, integration of visibility values over several rotations simply involves adding together visibility series from those rotations. (5) In constructing visibilities from observed counts, there is effectively a smoothing, which relaxes problems arising from the presence of zero counts. The price is the introduction of a small systematic error and a correlation of noise on visibilities in neighboring time bins.

4.2 From Count Rates to Visibilities

We now discuss the method by which the visibilities can be constructed from the counts. For simplicity, we discuss only the fundamental term in the Fourier expansion of the count rate. Constructing the visibilities needs to be done in the context of making a map, i.e. reconstructing an image, in a limited spatial region of interest, e.g. a solar flare on the Sun's disk. Central to this
is the concept of the phase at map center \( \Phi_0 \), touched on in section 3.4. In general, \( \Phi_0 \) depends on the results of the aspect solution, the time binning, the subcollimator, and the position of the map center. The phase at map center \((x_0, y_0)\) can be calculated from equation (6):

\[
\Phi_0 = \Phi(x_0, y_0) - \Phi(x_s, y_s).
\] (9)

\( \Phi(x_s, y_s) \) is the phase correction arising from the fact that the map center \((x_0, y_0)\) and the projected position of the subcollimator on the source plane \((x_s, y_s)\) are at different locations. Using the aspect system data, the distance of the line of maximum transmission from map center is computed for each roll angle \( \alpha_i \) and each subcollimator pitch \( p_k \). This gives the phase at map center for each timebin \( t_i \), detector \( k \). (See figure 4.)

Given an incident flux of photons on a HESSI subcollimator, after the selection of a map center, the predicted count rate \( C \) of a point source located at map center is given by equation (7), dropping the subscripts \( j, k \) for clarity:

\[
C_i = F_0 T_i \tau_i \cdot (1 + a_1 \cos[\Phi_0])
\] (10)

The visibility of a point source at the map center is:

\[
\mathcal{V}_i = F_0 e^{i\Phi(x_0, y_0)} = F_0 e^{i(\Phi_0 + \Phi'_1)}
\] (11)

Note that the visibility does not depend on the subcollimator coordinates \((x_s, y_s)\), but the count rate is subject to phase shifts caused by telescope motion. Comparing equations (10) and (11) leads to the following relation between counts and visibilities:

\[
X_i \equiv \Re \left[ \mathcal{V}_i e^{-i\Phi'_1} \right] = \frac{C_i - F_0 T_i \tau_i}{T_i \tau_i a_1}
\] (12)

where \( T_i, \tau_i \) and \( a_1 \) are known quantities, but \( F_0 \) is unknown. There are various ways to estimate \( F_0 \) and the optimum method in the presence of background has not yet been decided on.

Once \( F_0 \) has been determined, the conversion from counts to visibilities depends on oversampling of the circles in the UV plane. In each modulation cycle, there must be at least four time bins. During each cycle, the rotation of the modulation pattern within the map can be thought of as a purely linear shift. So for each successive 4 (or more) count-rates, one may represent the position in the \((u, v)\) plane by its average value \((\bar{u}_i, \bar{v}_i)\) and an offset \( \delta \). Then the corrected countrate (equation 12) can be represented as the cosine transform:

\[
X_i = \int \int \cos(2\pi[\bar{u}_i(x - x_0) + \bar{v}_i(y - y_0)] + \delta) f(x, y) dz dy
\]

The angle \( \delta \) changes by 360° during one cycle, so we can easily have 4 or more independent samples; a regression fit to each modulation cycle then gives both the sine and the cosine components \( X_i^{\sin} \) and \( X_i^{\cos} \).
It is worth mentioning that in early versions of RMC design, there were both sine and cosine subcollimators (e.g., Mertz et al. 1986, Murphy, 1990), but, as we see here, this turns out not to be necessary in hardware, since with sufficient sampling of the modulation profile, relative amplitudes of the sine and cosine components can be determined from the modulation profile itself.

It is important to note that, according to equation (12), the peak-to-peak variation of the visibility is several times larger than the peak-to-peak variation of $C$. This is because the visibility reflects the transform of the flux distribution at the front grid, while the count rate is the modulated profile of the photons reaching the detector. In fact, when the livetime is unity ($\tau_i = 1$), and when the subcollimator pattern is perfectly triangular ($a_T^i = 2/\pi^2$), the peak-to-peak variation of $\text{Re}[V_i]$ is $\pi^2/2 = 4.9$ times that of the count profile.

For a more general case, where the source is an arbitrary superposition of $N$ point sources of flux $F_j$ at locations $(x_j, y_j)$, the visibility is found by summing equation 12 over the sources. The generalization to higher harmonics is simple, and is important for achieving the highest possible resolution and for self calibration, but the details are beyond the scope of this paper.

5 Image Reconstruction

5.1 Back Projection and Fourier Transforms

Back projection (Mertz, Nakano, and Kilner 1986) is the most straightforward and basic method of image reconstruction. It is equivalent to a 2D Fourier transform (Kilner and Nakano, 1989). A map constructed by this method is called the “dirty map”, being the analogue of the radio astronomer’s initial Fourier transform of the observed visibilities. Back projection is a linear process: maps for arbitrary time intervals may be added together, and maps for different pitches and harmonics may be summed, generally leading to improvement of the image. Further improvements to the image by CLEAN or MEM (for example) do not share this property of linearity.

For a given instant in time, the response of a detector to a source flux distribution is its modulation pattern rotated to the roll angle for that time. Detected photons are most likely to have come from regions where the modulation pattern has its highest values. Back projection creates an image by distributing the received photons according to the modulation pattern. In fact, based on one time bin alone back projection gives an image that is identical to the modulation pattern. For many time bins, back-projection simply involves averaging the modulation patterns, weighting each one according to the the number of photons received in that time bin. Of course, in spreading flux across an image in this way some of it will be mis-placed. This is most obvious for the case of a point source, where photons have come from only one point, but will be traced back to many different regions by back projection. Despite this, given many timebins and several detectors, the greatest concentration of flux will usually be
at the location of the point source. In essence, back projection is analogous to performing an inverse Fourier transform, and the problems caused are a result of having insufficient coverage in Fourier space.

Mathematically, back projection can be described as follows. The modulation patterns $P_{im}$ are computed for all roll angle $\alpha_i$ and map pixels $m$. Then they are normalized according to a recipe first introduced by Durouchoux et al (1983). (a) Subtract out the mean over roll angles at each pixel: $(\bar{P}_{im} = P_{im} - < P_m >)$. (b) Then divide each value by its variance over roll angles: $(\tilde{P}_{im} = \bar{P}_{im} / < \bar{P}_m^2 >)$.

The brightness $D_m$ of each pixel $(m)$ in the back-projection map is defined by the following linear combination of the count rates $C_i$:

$$D_m = \frac{1}{A} \sum_{i=1}^{N} \left[ \frac{C_i}{\Delta t} \right] \cdot \tilde{P}_{im}$$  \hspace{1cm} (13)

where $A$ is the effective area (cm$^2$) of the detector and $\Delta t$ is the timebin duration (s). The initial Durouchoux normalization and the division by $A \Delta t$ ensures that the peak of the dirty map equals the strength of a dominant source. The map then has units of counts cm$^{-2}$ s$^{-1}$.

Figure 5 shows an example of a back-projection image for a simulated double source (Gaussian widths = 1" × 1").

All of the practical image reconstruction algorithms for improving a back-projection ("dirty") map are nonlinear, since linear deconvolution algorithms such as Wiener filtering and inverse filtering are inapplicable to applications with incomplete sampling of the Fourier plane, as is the case for HESSI modulation synthesis. A number of reconstruction methods already exist, or are being improved, or are being considered for addition to the HESSI reconstruction suite. We summarize several of these here.

5.2 CLEAN

CLEAN is an iterative algorithm which deconvolves a Point Spread Function—the imager’s response to a delta function source—from a “dirty map”. This algorithm is of fundamental importance in radio astronomy, where it is used to create images of astronomical sources obtained with interferometers. The basic CLEAN method was developed by Högbom (1974). It was originally designed for point sources, but it has been found to work well for extended sources as well when given a reasonable starting model. The Högbom CLEAN constructs discrete approximations to the CLEAN Map in the plane from the convolution equation

$$P \otimes I_{source} = D$$  \hspace{1cm} (14)

where $P$ is the HESSI PSF for one or many subcollimators and/or harmonics, $I_{source}$ is the source distribution, $D$ is the dirty map, and $\otimes$ denotes a convolution.
The algorithm goes as follows: Start with an initial approximation $I_0$; copy it to a new image, called the residual map. Search for the largest value in the residual map, saving the position in a "CLEAN component" table. Center a delta function at the location of the largest residual flux and give it an amplitude $\mu$ (the so-called "loop gain") times this value. Subtract the scaled Point Spread Function $P$ from $I_{n-1}$ to yield $I_n$. Continue this iteratively until a specified iteration limit is reached, or until the peak residual or RMS residual decreases to some level.

In the resulting final map ($I_{final}$), it is assumed that the residual brightness distribution consists mainly of noise. To damp out high spatial frequency features which may be spuriously created in the iteration, one convolves each CLEAN component with the so-called CLEAN PSF, $P_{clean}$ (the "CLEAN beam" in radioastronomy), which is simply a suitably smoothed version or inner portion of the PSF, $P$.

$$I_{clean} = I_{final} \otimes P_{clean} + \{ D - P \otimes I_{final} \}$$

A CLEAN map is produced when the final residual map is added to the approximate solution, in order to include the noise.

The main disadvantage of CLEAN is that it does not, at least in the Högbom version, compare the observed modulation profile with a model modulation profile, to assess the "goodness of fit" during the iteration. Two variants of CLEAN exist for HESSI. One is for rectangular coordinates, and the other is for polar coordinates. In most circumstances, the latter has a significant advantage of speed over the former.

Figure 6 shows an example "CLEAN map" derived from a simulated event list derived from a double source.

5.3 Maximum Entropy Methods

The principle of Maximum Entropy is to find the map which is maximally noncommittal with regard to the unavailable information, i.e. the regions of the Fourier plane for which HESSI has no data. The image to be found is the smoothest one which fits the data to within the noise level.

HESSI’s observations are provided in the form of nine sets of counts in time-bins, each time-bin corresponding to a particular roll angle. The aim is to find an image that is consistent with these counts to within the photon counting noise. Since there can be many images consistent with the observations, MEM is designed to find the smoothest such image. Traditionally, consistency is measured using a $\chi^2$ measure on the counts:

$$\chi^2 = \sum_i \frac{(C_i - E_i)^2}{\sigma_i^2}$$
where $C_i$ is the count for timebin $i$ and $E_i$ is the "expected" count from the reconstructed image, and $\sigma_i$ is the estimated error of count $C_i$. If only photon counting noise is important then $\sigma_i^2 = E_i$, but in practice systematic errors and uncertainties in the background may also be important and so should be included in $\sigma_i$.

The smoothness of the image is quantified using an entropy measure. In one view, the entropy is something which, when maximized, produces a positive image with a compressed range in pixel values (Cornwell 1984), but according to Gull and Skilling (1984) and Sivia (1996) the only function which guarantees that no unwanted correlation is imposed is:

$$H = -\sum_m F_m \log F_m$$

where $F_m$ is the flux in pixel $m$. A smooth image will have a larger $H$. The goal of MEM is to find the smoothest image that fits the data. In practice, this means maximizing $Q = H - \frac{1}{2} \chi^2$ for the smallest possible $\lambda$. The idea is that $\lambda$ is set to a small value to begin with and an image is sought iteratively that has $\chi^2 = 1$, i.e. it is consistent with observations. If such an image cannot be found, the smoothness constraint is probably too strong, and so $\lambda$ is increased to allow a less smooth image. Additional constraints (e.g. total flux) can be added to $Q$ if desired.

MEM has had a rich history as the main imaging algorithm used for image reconstruction with Yohkoh's HXT. Although the details of implementation are somewhat different for HESSI, the basic theory is still relevant (Sato, 1998, Sato, Kosugi, and Makishima, 1999).

Another HESSI image reconstruction algorithm, MEMVIS, applies MEM to the visibilities rather than the counts. The advantages of using visibilities, were discussed in Section 4. Before iterations begin in MEMVIS, the counts are converted to visibilities, correcting for the wandering of the spin axis. With this correction applied the modulation patterns, and therefore the processing per iteration, are greatly simplified. The basic MEM equations outlined above need some alteration because visibilities are complex quantities. For example, the real and imaginary parts in the $\chi^2$ sum must be separated, and the constraints must be reformed to avoid $Q$ being complex. The MEM algorithm then proceeds according to the steps described above.

The main advantages of MEMVIS is its efficiency (lower memory requirements, use of fast Fourier transforms) and the ease in which the visibilities that it uses can be integrated over time (by literally just adding them together). It is apparent from many simulations that MEMVIS appears to be remarkably robust at low count rates.

Figure 7 shows an example simulated source, the "MEM map", and the residuals.
5.4 Forward Fitting

The forward-fitting method is based on models that represent a spatial map by a superposition of multiple source structures, which are quantified by circular Gaussians (4 parameters per source), elliptical Gaussians (6 parameters per source), or curved ellipticals (7 parameters), designed to characterize real solar flare hard X-ray maps with a minimum number of geometric elements. Examples of fits to hard X-ray sources may be seen in Aschwanden et al., 1999.

In one realization of this scheme, maps of superimposed Gaussians are created iteratively, at each step constructing a model modulation profile from the map. In this “pixelized” form, the map evolves into a best fitting image.

In another realization of Forward Fitting, there are no maps computed at intermediate steps. The algorithm uses the beat patterns seen in the modulation patterns to infer the characteristics of the source. The most efficient way to do this is to compute the visibilities from the count profiles (after selecting map center and calibrating). Then a phase-shift to map center eliminates the rapid modulations, leaving only the low-frequency beat patterns. For two circular Gaussians, the extrema of the beat patterns then provide the relative separation and strengths of the sources. The ratios of the beat strengths for successive subcollimators determine the Gaussian widths. The inferred parameters of the source then provide model modulation profiles for comparison with the observed profiles. Chi-square determinations assess the goodness of fit for iterative estimations of the strength of further sources. Figure 8 shows fits to visibilities deriving from simulated count rate profiles for a double source of 10,000 counts/s/SC. In principle, fits such as this can yield the flux, positions and widths of two Gaussian sources, and possibly more if the fluxes are sufficiently high.

5.5 The PIXON Method

The PIXON method is another technique removes the sidelobe pattern of a telescope while mitigating the problems of correlated residuals and spurious sources which are commonly seen in Fourier deconvolution, chi-square fitting, and Maximum Entropy approaches.

The goal of the Pixon method is to construct the simplest, i.e. smoothest, model for the image that would be consistent with the data, i.e. have an acceptable chi-square fit. Being the simplest model, the derived image would be artifact free, i.e. there would be no spurious sources. In addition, it is claimed by PIXON proponents that the model is necessarily most tightly constrained by the data, and consequently has the most accurately determined parameters.

The PIXON method changes the global smoothness idea of MEM into a local condition in which local maximum smoothness of the image is imposed. From an information science point of view, one selects a model with the minimum information content from the family of multiresolution basis functions (pixons) and which statistically fit the data.
Since the model has minimum complexity, spurious sources are unlikely to arise. Each parameter is determined using a larger fraction of the data, hence it is, presumably, determined more accurately. It is claimed that this usually results in superior photometric and positional accuracy, and that, since the minimum number of parameters are used, the data cannot be over fitted. PIXONS, however, pays a price for its photometry: the method is about two orders of magnitude slower than the other reconstruction methods. Therefore, as with Yohkoh/HXT, it will probably be used only after the faster reconstructions have produced maps, and when checking the strength or shape of weak sources is required.

Figure 9 shows an example "PIXON map", and the model source from which simulated countrate profiles were derived. The reconstructed sources are about 1" FWHM, which is the size of the simulated model sources.

6 Expected Imaging Performance

In 2-3 years of mission life, HESSI is expected to obtain observations of tens of thousands of microflares, thousands of hard x-ray flares, and of order a hundred gamma-ray line flares. Detected photons will be tagged with 1 μs time resolution up to a rate of ~ 10⁵ photons/s. The intense 3-150 keV X-ray fluxes that usually accompany large gamma-ray line flares are absorbed by the front segment of the detectors, so the rear segment will always count at moderate rates. This is essential for gamma-ray line measurements with optimal spectral resolution and high throughput.

It is expected that HESSI will provide detailed imaging spectroscopy in each of ten energy intervals with ~ 2s time resolution for one event every ~ 5 days with ~ 10⁴ counts s⁻¹ above 15 keV. With lower energy resolution (~ 100 keV), crude imaging information could be obtained in tens of milliseconds. HESSI provides spatial resolution of 2.3 arcseconds at X-ray energies below ~ 200 keV, 7 arcseconds to 400 keV, and 36 arcseconds for gamma-ray lines and continuum above 1 MeV.

6.1 Imaging Strengths and Limitations

HESSI's imaging strengths lie in the strengths of Fourier imaging: insensitivity to wandering pointing, independence of relative detector-to-detector efficiencies, many independent Fourier components in the azimuthal coordinate, and additivity of harmonic contributions. But in Fourier lies its weakness: incomplete coverage of the visibility plane due to a finite number of UV circles, which causes the maps to become rougher and noisier for time intervals less than one half rotation.
6.2 Self Calibration

One of the features of HESSI imaging is that there are several ways to self-calibrate amplitudes and phases. Among these are:

(i) The modulation profiles of two halves of a single rotation will reveal a phase difference that is characteristic of the subcollimator's phase offset from the nominal telescope axis.

(ii) If the telescope is performing as designed, the third harmonic of count rate profiles from subcollimator n should be identical to the fundamental of profiles from subcollimator n-2. If not, the ratio of the signals will provide information to correct the instrumental database of amplitudes and phases.

(iii) The modulation amplitudes of two subcollimators responding to unresolved sources should be the same at low energies (except for known internal shadowing effects). At high energies (> 300 keV), the relative amplitudes should depend on the nominal relative grid-slat opacities in a predictable way.

(iv) The subcollimator transmission and amplitude functions should be periodic in roll angle, with a period of 180 degrees. Any asymmetries found in this periodicity can provide information about departures from nominal grid positioning.

6.3 Expected Evolution of Imaging

During the course of the HESSI mission, undoubtedly lessons will be learned about unanticipated features of the hardware and software. As more is discovered about the instrument, methods of analysis can be tailored appropriately. As in other missions, imaging software is expected to improve as bugs are eradicated, and better algorithms are developed. The solar community is urged to test software, give constructive criticism, and provide improvements of their own devising. Better science and better understanding of flares will surely follow as a result of such cooperative efforts.

7 References


Figure 1: Schematic geometry of the HESSI subcollimators, showing representative incidence of photons with respect to the collimator axis.
Figure 2: Profiles of the equivalent-area 1-D function for six energy bands (3-6, 6-12, 12-25, 25-50, 50-100, 100-300, 300-1000 keV) and three map centers for subcollimator 5. Only the highest energy profiles are distinguishable from the rest, due to the lower opacity of the grid for E> 300 keV. As the distance from the spin axis increases (300, 600, and 900") the internal shadowing changes, causing the minima to flatten. All curves are plotted as sums of the first 3 harmonics.
Figure 3: Modulation profiles by an ideal subcollimator. The first plot shows the profile of a source of width $= 1/2$ the angular pitch at an angular distance of 8 pitches from the normal to an ideal subcollimator. The second plot shows the profile for a source of half the flux. The third and fourth plots show the response for shifts of the source by two angular pitches in the radial and azimuthal directions (respectively). Note the translation of the profile in the first case and the increase in the number of extrema for the second. The fifth and sixth plots illustrate the effect of decreasing or increasing (respectively) the source size by a factor of 2. The dashed profile shows the imaginary part of the visibility, discussed in section 4.
Figure 4: Lines of constant phase for one subcollimator at one instant of rotation projected on the solar disk. The spacecraft coordinate system rotates clockwise (increasing $\alpha$), and the lines of constant phase lie at a constant angle $\beta$ relative to the spacecraft coordinates. The phase contours, shown by the parallel dashed lines numbered 1,2,3,..., represent "fringes" and are separated by multiples $2\pi$ radians in phase from the subcollimator line of maximum response.
Figure 5: An example of a back-projection map for a simulated double source of 10,000 counts/s/SC.
Figure 7: An example of a MEM map for a simulated double source of 10,000 counts/s/SC.
Figure 6: An example of a clean map for a simulated double source of 10,000 counts/s/SC.
Figure 8: Fits to visibilities derived from simulated count rate profiles for a double source of 10,000 counts/s/SC. The fits yield the flux, positions and widths of two Gaussian sources.
Figure 9: An example of a PIXON map for a simulated double source of 10,000 counts/s/SC. The logarithmic contours show that the dynamic range is over 50:1 in this map.