

MIT Open Access Articles

Remote Sensing of the Heliosphere with the Murchison Widefield Array

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Oberoi, D., and L. Benkevitch. "Remote Sensing of the Heliosphere with the Murchison Widefield Array." *Solar Physics* 265, Numbers 1-2 (2010): 293-307. Copyright © 2010, Springer Science+Business Media

As Published: <http://dx.doi.org/10.1007/s11207-010-9580-x>

Publisher: Springer

Persistent URL: <http://hdl.handle.net/1721.1/66212>

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Terms of Use: Article is made available in accordance with the publisher's policy and may be subject to US copyright law. Please refer to the publisher's site for terms of use.



Remote Sensing of the Heliosphere with the Murchison Widefield Array

D. Oberoi¹ · L. Benkevitch¹

© Springer ●●●

Abstract The Murchison Widefield Array (MWA) is one of the new technology low frequency radio interferometers currently under construction at an extremely radio quiet location in Western Australia. The MWA design brings to bear the recent availability of powerful high-speed computational and digital signal processing capabilities on the problem of low frequency high-fidelity imaging with a rapid cadence and high spectral resolution. Solar and heliosphere science are among the key science objectives of the MWA, and have guided the array design from its very conception. We present here a brief overview of the design and capabilities of the MWA with emphasis on its suitability for solar physics and remote-sensing of the heliosphere. We discuss the solar imaging and interplanetary scintillation (IPS) science capabilities of the MWA and also describe a new software framework. This software, referred to as Haystack InterPlanetary Software System (HIPSS), aims to provide a common data repository, interface, and analysis tools for IPS data from all observatories across the world.

Keywords: Interplanetary Scintillation, Instruments, Techniques

1. Introduction

A vast array of instrumentation routinely observes the solar disc across an impressively wide band of the electro-magnetic spectrum. A large and capable flotilla of satellites provides *in-situ* measurements of the solar wind plasma, most of them in Earth orbit or close to it and in the ecliptic plane. In the vast inner heliosphere however, there are far fewer means available to investigate and understand the solar wind. Yet, it is the evolution and dynamics of the solar wind in this region which plays a dominant role in space weather and need to be understood to relate the observations close to the Sun to the *in-situ* measurements close to 1 AU. Interplanetary scintillation (IPS) has been amongst the earliest techniques to investigate the solar wind in the inner heliosphere (e.g. Hewish, Scott, and Wills, 1964) and has led to many interesting discoveries about the solar wind. The discovery of high velocity polar wind much before the Ulysses

¹ Massachusetts Institute of Technology, Haystack Observatory, Westford, MA 01886, USA email: doberoi@haystack.mit.edu email: benkev@haystack.mit.edu

polar pass (Coles *et al.*, 1980; Kojima and Kakinuma, 1990) is a good example. With the time accumulated learning, the application of the IPS technique and its ability to extract information about the solar wind from the data has grown considerably. Comparatively recently another powerful and complementary data source for studying the inner heliosphere has become available. The Solar Mass Ejection Imager (SMEI) on board the Coriolis satellite was the pioneer in extending the Thompson scattering measurements out to the entire inner heliosphere (e.g. Eyles *et al.*, 2003). The second generation of Heliospheric Imagers (HIs) on-board the STEREO satellites now provide unprecedented views of solar wind flows in large parts of the heliosphere (e.g. Harrison *et al.*, 2009).

Two recent developments suggest that IPS is now poised for the next leap in its capabilities and on its way to becoming a mainstream remote-sensing heliospheric diagnostic. The first of these developments is the expected availability of the new generation radio instruments like the Murchison Widefield Array (MWA) (Lonsdale *et al.*, 2009) and the Low Frequency Array (LOFAR) (de Vos, Gunst, and Nijboer, 2009), along with new instrumentation at a currently unsampled longitude like the Mexico Array Telescope (MEXART) (Mejia-Ambriz *et al.*, this issue) which has recently seen first light. In addition, very considerable upgrades to some of the existing facilities like the Ooty Radio Telescope (ORT) (Joshi *et al.*, 1988) in India and the Solar Wind Imaging Facility (SWIFT) (Kojima *et al.*, 2003) in Japan are well underway. The second is a community-wide desire and effort to move to a common standard for data format and procedures for archiving and analyzing IPS data.

In Section 2 we introduce the MWA, one of the new instruments expected to become available in the next few years to illustrate the design and significantly superior capabilities of the new generation instruments. Section 3 briefly describes the solar and heliospheric science objectives of the MWA. Section 4 presents our expectations for the future of IPS science in the coming years, and Section 5 describes our first steps towards realizing this goal, a design for a unified IPS software system to archive and analyze IPS data, followed by the Conclusions.

2. Introduction to the Murchison Widefield Array

The low frequency radio spectrum is the last remaining part of the radio spectrum accessible from the earth, which is yet to be explored with good fidelity and sensitivity. High fidelity aperture synthesis at low frequencies has been a difficult challenge due to multiple reasons. The most important of these are: 1. corruptions imposed by the dynamic terrestrial ionosphere on the incident radiation, 2. the wide fields-of-view associated with long wavelengths, 3. the rather intense galactic background which lends a different character to the low frequency radio sky, and 4. the increasing menace of radio frequency interference (RFI), a by-product of modern human lifestyle.

The MWA design (Lonsdale *et al.*, 2009) applies the recent advances in capabilities and affordability of high-speed computational and digital signal processing hardware to interferometer design in order to meet the needs of low

frequency radio imaging. This fresh approach is geared towards maximizing the information collected about the sky and retaining it all the way through the processing to the final imaging. The MWA comprises 8,192 dual-polarization broad-band active dipoles optimized to operate in the frequency range 80-300 MHz. These dipoles are grouped in 512 *Tiles*, each with 16 dipoles arranged in a 4×4 grid. The tiles measure $5\text{m}\times 5\text{m}$ and are distributed in a quasi random manner, most of them inside a region 1.5 km in diameter and a small number of outliers spread out to 3 km. These tiles are steered to the desired direction by combining the signals from each of the 16 dipoles of a given polarization with an appropriate delay in an analog circuit, referred to as the *Beamformer*. The beamformed RF signals from a group of 8 tiles are brought to a *Receiver Node* via co-axial cables. The key responsibilities of the receiver nodes are to do appropriate analog signal conditioning, digitizing, and spectrally filtering the entire RF stream. Signal conditioning refers to limiting the useful part of the RF band to 80-300 MHz, where MWA tiles gathers sky-noise limited signal, using appropriate high and low pass filters. The receiver does a direct sampling of the RF at 655.36 Msamples/sec (corresponding to a Nyquist frequency of 327.68 MHz) and implements a 256 channel polyphase filterbank, yielding a channel width of 1.28 MHz. 24 of these coarse spectral channels, amounting to a bandwidth of 30.72 MHz, are selected from each of the 64 receiver nodes and transmitted to a central processing location over optical fiber. At the other end of the optical fiber they are received by the *Digital Correlator and Array Beamformer*, usually referred to simply as the correlator. Some of the F part of the MWA's FX architecture is implemented in the Receiver Nodes and the remaining part, which does the finer spectral filtering, is implemented in the correlator. The full polarization cross-correlation architecture of MWA requires 524,288 signal pairs to be multiplied together, each at a bandwidth of 30.72 MHz, requiring 1.6×10^{13} complex multiply and accumulate (CMAC) operations per second. The cross-correlation operation follows the computationally-intensive fine filtering and a large data re-organization operation, making the correlator design rather challenging. As the name of this device suggests, in addition to the regular interferometric correlation process, this sub-system also implements a digital array beamformer yielding up to 32 single polarization beams. This functionality has been implemented using Field Programmable Gate Array (FPGA) based boards. The technical specifications of the MWA design have been summarized in Table 1.

The MWA design embodies the most aggressive use of the so called *large-N* architectures so far, where N refers to the number of elements in an interferometer. The large number of elements serve a few different purposes. With no moving parts and comparatively simple electronics they are robust yet inexpensive. They provide an affordable way to build up a large collecting area while maintaining a large field-of-view. The 512 tiles provide comprehensive sampling of the spatial frequencies of the radio sky, far beyond any of the existing instruments, endowing the array with the high fidelity radio imaging capabilities. The large collecting area is necessary to build up the sensitivity needed to image a sufficiently dense grid of calibrator sources with sufficient signal-to-noise on a time scale shorter than that of ionospheric evolution. The calibrators are well known sources with

Table 1. MWA Specifications

Frequency range	80–300 MHz
Number of receptors	8,192 dual-pol dipoles
Number of antenna tiles	512
Number of physical baselines	130,816
Collecting area	$\approx 8000\text{m}^2$ (at 200 MHz)
Field of View	$\approx 15^\circ\text{--}50^\circ$ ($\approx 1000\text{ deg}^2$ at 200 MHz)
Configuration	Core array $\approx 1.5\text{ km}$ diameter ($\approx 97\%$ of collecting area) extended array $\approx 3\text{ km}$ diameter ($\approx 3\%$ of collecting area)
Bandwidth	220 MHz (sampled) 30.72 MHz (processed)
Spectral channels	1024 (30 kHz spectral resolution)
Temporal resolution	0.5 s (uncalibrated) 8 s (calibrated)
Polarization	Full Stokes
Continuum point source sensitivity	20mJy in 1 s (at 200 MHz, full bandwidth) 0.34 mJy in 1 hr (at 200 MHz, full bandwidth)
Array voltage beams	32, single pol

comparatively simple morphologies, the departures in their instantaneous images from the expected images are used to determine and undo the effects of propagation through the ionosphere. Mitchell *et al.* (2008) provide the details of the calibration approach for the MWA. The MWA processing flow has been designed to preserve as much information as possible. In order to achieve this goal, the data are averaged over time and frequency bins small enough to maintain the coherence of the data from horizon to horizon. The MWA is located at an extremely radio-quiet site in the remote and inhospitable Western Australian outback to minimize the complexity arising from strong RFI associated with the proximity of human presence.

3. Solar and Heliospheric Science with the MWA

The solar and heliospheric science objectives of the MWA can be categorized into imaging studies of the Sun and studies of the solar wind via propagation effects like IPS and Faraday rotation (FR). We present brief discussions of solar imaging and IPS studies later, and FR studies are discussed in detail by Liu *et al.* (2007) and Jensen *et al.* (2010, this issue).

3.1. Solar Imaging

The MWA design is optimized for high fidelity snapshot mono-chromatic imaging over its wide field-of-view with high sensitivity to emission at large angular scales. These attributes make the MWA exceptionally well suited for studying extended sources with complicated morphology, the emission from which shows significant evolution in both time and frequency on a wide range of scales, like the Sun. The

MWA is expected to play a very productive role in improving our understanding of both the quiet and the active Sun.

For the quiet Sun, the solar emission in the MWA frequency range is dominated by thermal emission from the hot, optically thin coronal plasma. The large gradients in the electron density and temperature from the chromospheric to the inner coronal regions lead to an even larger gradient, of about 7 orders of magnitude, in the value of the absorption coefficient, κ , in this region. These rather large changes translate into the optical depth, τ , being a strong function of the ray path, which is in turn determined by refraction due to the large-scale density gradients. This presents a unique opportunity to use a set of apparent solar brightness temperature images obtained at different observing frequencies to independently determine both the coronal electron density and temperature distribution in the lower corona. Due to refraction, the ray paths at different frequencies traverse different regions of the corona, which makes it possible to put together a consistent three-dimensional (3-D) picture of lower corona using observations spanning the MWA frequency range. This approach will also be very interesting for comparing the observations with the expectations based on well accepted coronal and chromospheric models, and for exploring more subtle effects like the temperatures and densities in coronal streamers and holes.

While a large number of CMEs are easily evident in white-light coronagraphs, their counterparts in other wavelengths have remained more elusive. The advantages of radio observations over those from optical and soft X-ray have been appreciated for some time now (Bastian and Gary, 1997) and the metric and decimetric band is widely regarded as the most promising part of the spectrum for CME studies. With only a handful of successful attempts reported in the literature (Sheridan *et al.*, 1978; Bastian *et al.*, 2001; White *et al.*, 2006; Maia *et al.*, 2007; Magdalenic *et al.*, 2008), direct radio detection of CME emission has remained largely beyond the reach of existing instrumentation. As mentioned earlier, the MWA design is very well matched to the challenge of imaging a complex source like the active Sun. One of the prime objectives of the MWA is to undertake imaging studies of the Type-II radio emission, most often associated with more energetic CMEs. Considerable insight has been gained about the origin and properties of these bursts, based on studies of their dynamic spectra. The mechanism for this radio emission is believed to be plasma radiation from the locations of the shock, which drifts to lower frequencies with time as the shock propagates radially in the corona (Wild, 1950; Gopalswamy, 2000; Cairns *et al.*, 2003; Schmidt and Gopalswamy, 2008). For lack of better information, the matter build up seen in coronagraph images is often used as a proxy for the shock location. In some of the cases where the radio burst emission has actually been imaged, it is difficult to interpret the observations in the framework of the prevalent models for CMEs and Type-II emission (White *et al.*, 2006; Magdalenic *et al.*, 2008). Routine imaging of location, morphology and 'evolution' of Type-II burst emission with the MWA will be very useful in helping us build a better understanding of the Type-II radio emission and their relationship with CMEs.

3.2. Interplanetary Scintillation (IPS)

As will be argued in the next section, we believe that the returns from existing observations of IPS are limited in a large measure by the insufficient sampling of the heliosphere provided by current instrumentation. This arises due to different reasons for different instruments; these include limited sky coverage in declination and hour-angle, insufficient sensitivity, limitations posed by weather and geographic locations, and slow instrument slew rates. More importantly, the conventional instruments have one beam, i.e. at a given time, they can observe only one line-of-sight. The ORT, which allows up to 12 simultaneous beams, is an exception. However, these beams are fixed with respect to one another which greatly diminishes their usefulness for IPS.

The MWA offers a very significant improvement on this front by providing multiple independent simultaneous beams, each of which can be used to monitor a line of sight. The individual MWA tiles are only a few wavelengths across. The beamformed data stream obtained by combining the signal from the 16 dipoles hence maintains a large field-of-view, as emphasized in the name MWA. A second layer of beamforming is done using the digital data streams from all the 512 tiles in the Correlator and Digital Beamformer sub-system to produce up to 32 single polarization beams. Each of these beams uses the entire collecting area and bandwidth of the array, can be independently pointed anywhere in the tile field-of-view and used to track astronomical sources. In terms of existing instruments, this is analogous to having 32 independent instruments co-located and available for observations.

The MWA design also offers another advantage which is often not as well appreciated. The rapid electronic beam steering of the MWA array beams and a 30° elevation limit imply that first, the array has a large part of the hemisphere available for observations at any given time and second, precious observing time is not lost in slewing from one source to the next. In contrast, transit arrays like MEXART, Puschino, and SWIFT can sample a given part of the sky only once per day, during the short period it passes through the meridian, a rather severe limitation on heliospheric sampling. For steerable systems like the ORT, the beam is steered electronically in declination, but requires physical rotation of the antenna to steer the beam in hour angle. ORT provides hour angle coverage of 9.5 hours with a slew rate of $6.25^\circ/\text{min}$, and the observing strategies tend to minimize the dead time due to slewing from one source to the next. This naturally imposes a heliospheric sampling function where sources at the neighboring hour angles tend to get sampled at nearby times and the part of the heliosphere being sampled sweeps across, usually from West to East before starting over again. While this maximizes the observing efficiency and the overall coverage of the heliosphere, it implies that a given region in the heliosphere is sampled rather coarsely in time. This has the unintended consequence that many of the existing IPS data-sets are systematically missing short-term variations (from a few to 24 hours). With no extraneous considerations influencing the observing strategy, the MWA observing plan will be optimized for sampling the heliosphere uniformly in both spatial and temporal sense. The LOFAR design also shares the advantages from multiple independent simultaneous beams and electronic

beam steering. The northern-biased field of view of LOFAR (Latitude = $52^{\circ}55'$) should provide a good complement to the southern-biased field of view of MWA (Latitude = $-26^{\circ}42'$).

We also note that observations of IPS will be especially useful in the current era when one can expect observations from white-light heliospheric imagers (HIs) to be routinely available. HIs probe the large-scale electron distribution in the heliosphere while IPS probes the micro-turbulence properties. The length and time scales probed by IPS are much finer than permitted by HIs and turbulence properties play an important role in heliospheric dynamics. These two techniques complement each other very well and have led to many interesting and informative studies, e.g. Manoharan (2006); Jackson *et al.* (2007); Bisi *et al.* (2009) to list a few illustrative examples. The expectation that the STEREO observations would continue well beyond a separation of 180° between the two spacecraft, is a welcome recent development. As the separation grows beyond 180° , the HI fields of view would tend to be looking back towards the Earth along some of the same lines of sight as IPS sources. This would, for the first time, make available measurements of IPS and Thompson scattering along exactly the same lines of sight at interesting elongations, providing greater context and hopefully an improved understanding of observations of IPS.

4. Future of IPS

As for all propagation effects, the measurable for IPS is integrated all along the line-of-sight through the interplanetary medium (IPM). In the case of single-site studies of IPS, for instance, the measurable of interest is a statistical estimate of the fluctuations in the observed intensity of a distant compact radio source, or the power spectrum of intensity fluctuations. Under the assumption of weak scattering, the IPM can be regarded as comprised of individual thin screens and the observed power spectrum modeled as a weighted sum of the power spectra of intensity fluctuations from these thin screens. The equivalent observable for multi-site studies of IPS is the set of cross-spectra of intensity fluctuations measured at each of the baselines provided by the multiple sites. A less frequently used observable of IPS is the normalized co-spectrum obtained from simultaneous observations of IPS at frequencies differing by a factors of ~ 2 (e.g. Scott, Rickett, and Armstrong, 1983; Zhang, 2007). It does offer some advantages over single site, single frequency measurements of IPS and one can reasonably expect its to be used more widely with the next generation of instruments which provide a broad spectral coverage. It is the dependence of the IPS measurables on the large-scale structure of the solar wind combined with the ability to model it in terms of physical properties of the IPM which gives IPS its considerable strength for investigating the solar wind (e.g. Manoharan *et al.*, 2000; Breen *et al.*, 2002; Bisi *et al.*, 2007, 2010; Tokumaru, Kojima, and Fujiki, 2010).

Developing this line of thought to its logical conclusion brings us to a scenario where a suitable inversion process is able to extract the information about the large-scale structure of the solar wind in the entire heliosphere using a large carefully assembled set of observations of IPS as inputs. Some of the earliest

work along these lines was done by Jackson *et al.* (1998); Kojima *et al.* (1998); Asai *et al.* (1998) and Oberoi (2000). Considerable progress has been made on the technique and its application in the past decade. The most constraining bottleneck for this approach is currently due to insufficient sampling of the solar wind available from current instrumentation.

As mentioned earlier, the new capable instrumentation shortly coming online is expected to vastly improve our heliospheric sampling, and revitalize this field by providing a far richer dataset. The limitation that individual instruments can monitor only a part of the heliosphere at any given time however still remains. This limitation can only be overcome by having IPS instruments distributed in longitude to provide a continuous coverage of the heliosphere, and synthesizing them into a virtual global IPS array. The realization that the returns from the IPS technique can be maximized by simultaneously using data from all available instruments, rather than restricting oneself to data from the instrument one is most familiar with, has steadily been growing in the IPS community. The impending availability of new instruments has catalyzed the community into working on synthesizing a single global instrument comprising all the individual IPS facilities in the world. A few things must be done in order to accomplish this ambitious goal.

1. Different IPS techniques (single-site vs. multi-site) and analysis procedures followed at different observatories need to be cross calibrated against one another to yield consistent results. Some earlier attempts to do this (Moran *et al.*, 2000) have been reasonably successful.
2. It is essential to make it easier to access data from different instruments, perhaps via a unified interface, and implementation of standardized analysis procedures which can be applied to data from different observatories.

The next section describes the software system being designed to meet the second of the aforementioned needs.

5. The Haystack InterPlanetary Software System

A software suite, referred to as the Haystack InterPlanetary Software System (HIPSS) is being developed for storing and analyzing the IPS data from the MWA. As will be discussed in the next sub-section, the MWA IPS data-stream is significantly more complicated than that from any of the existing instruments, and therefore requires a more flexible framework than has conventionally been needed. We make use of this opportunity to come up with a framework general enough to accommodate data not only from the MWA but from all IPS observatories across the world. In addition, HIPSS will provide a common interface with a flexible and efficient querying mechanism to facilitate science studies, and also implement the functionality to do some routine data processing and generate standard data products to reduce the tedium of usual repetitive data processing.

The HIPSS design has been guided by the following principals:

1. The data repository must support IPS data from all existing and planned IPS observatories across the globe. This requires the design to be flexible and make no assumptions about the data it is meant to host.
2. The database must store the data as close to the raw data as reasonably possible. This is motivated by a desire to avoid irreversible loss of information in the data (for example, by averaging), retain as much flexibility as possible for downstream processing, and benefit from the continuing improvement in analysis algorithms.
3. The data repository must host the data at multiple levels of processing, all the way from almost raw time series data to convenient *science ready* parameters resulting from model fits to the data. The underlying thinking is that the availability of sufficiently fine grained IPS data will permit the IPS experts to perform custom or non-standard analysis as needed, while its availability in a form which is easier to understand and make use of by non-experts will further encourage the use of IPS data in the wider community.
4. The database must provide efficient and flexible means for querying. This is driven by a desire to provide an efficient data-mining tool. Making it possible to conveniently run complicated queries for data selection will encourage the users to think fresh, and ask and pursue questions which might be too tedious or difficult to pursue in absence of such a tool.
5. The design is based on a scenario where each contributing observatory has a local installation of HIPSS, along with perhaps a few data centers. HIPSS therefore needs to be designed with the ease of installation and management as a significant consideration.

After examining the various options available for the choice of implementing a data repository, we have chosen to implement it in the Version 5 of Hierarchical Data Format (HDF5) with Python, and Python add-ons, providing the primary tools interacting with the repository. HDF5 is best described as a data model, library and file format for storing and managing data. It is designed for flexible and efficient I/O and for high volume and complex data¹. HDF5 is rapidly gaining popularity in the scientific community. Python is an interpreted, interactive, object-oriented, extensible programming language and is widely used by the scientific community².

5.1. The Data Model

Traditionally, IPS instruments observe only a single source at a given time, usually in a single fixed and contiguous observing band. These data are typically reduced to a single time series of real numbers, representing the intensity scintillation of the radio flux of the IPS source with a time resolution of a few tens of ms. A typical observation lasts a few to many minutes and a 15 min observation with 10ms sampling recorded as floats leads to 36 kBytes of data, a rather trivial

¹More information about HDF5 is available at <http://www.hdfgroup.org/HDF5>

²More information about Python is available at <http://www.python.org>

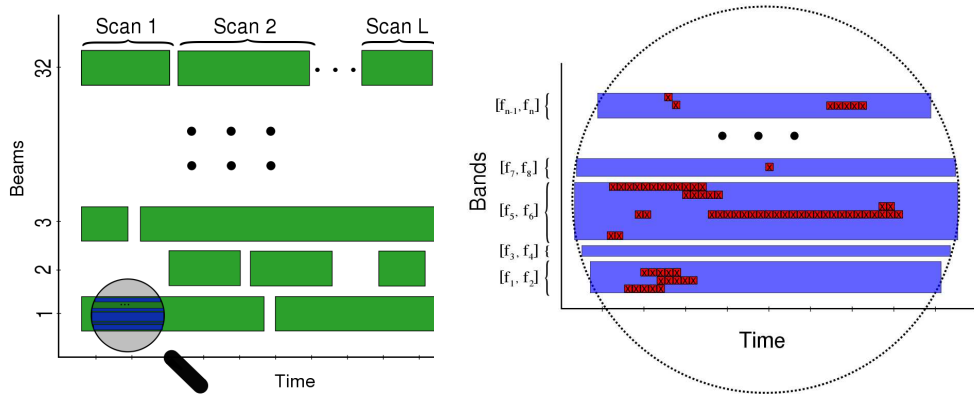


Figure 1. The left panel shows data from a large number of independent beams, represented by the green strips, being gathered simultaneously from an instrument. The right panel shows a zoomed in view of the region under the lens in the left panel to illustrate the frequency structure of the data from one of the beams. The Y axis represents frequency. The data for each stream can consist of many bands represented by the blue strips. The bands themselves can be of different widths and with arbitrary gaps between them. The red boxes marked with crosses represent known blocks of bad data which need to be excluded from all subsequent analyses.

amount of data by most current standards. Multi-station observations comprise one such time series from each of the stations involved.

The new generation of instruments, like the MWA and LOFAR, will have multiple beams and can provide non-contiguous bandwidth over spans too large to be averaged into a single time series. We have developed a flexible data model to accommodate such data, a schematic representation of which is shown in Figure 1. Data flows in simultaneously from a large number of independent beams. The observations from different beams start and stop independently of one another and can have arbitrary data gaps between consecutive observations. The data streams from different beams can differ from one another in their bandwidths and organization. A given beam will comprise multiple *bands*, regions of contiguous bandwidth coverage. The bands can vary in widths and there can be arbitrary gaps between adjacent bands. Some of the bands can be so wide that it will be necessary to divide the data in multiple finer *sub-bands* before they can be averaged across frequency and stored as a time series. A sub-band is defined to be the bandwidth over which the data are averaged before being written to the HDF5 data repository and its size can be controlled by the user. Information about known bad data available from the observatory will also be included in the data repository as flagging information. In addition to the data itself, each of the observations have some meta-data associated with it, these include information specific to the instrument, observational setup, observing geometry and a summary of solar wind parameters obtained from the IPS data itself. An exhaustive list of meta-data are provided in Table 2.

Table 2. List of meta-data to be stored in Scans table

	Meta-data field	Variable name	Variable type	Unit
Instrument and signal specification				
1	Telescope/Instrument	telescope	String(9)	
2	Start time	start_time	Double	Seconds (UT)
3	End time	end_time	Double	Seconds (UT)
4	Beam No.	beam	Integer	Number
5	No. of bands	n_bands	Integer	Number
6	Lower end of frequency band	lower_freq	Float	MHz
7	Upper end of frequency band	upper_freq	Float	MHz
8	Total bandwidth of observations	bandwidth	Float	MHz
9	Sampling time	sampling	Double	Seconds
Source specification				
10	IAU Name	iau_name	String(9)	
11	Alternate Name	alt_name	String(9)	
12	RA(J2000)	ra	Double	radian
13	DEC(J2000)	dec	Double	radian
14	RA-Epoch	ra_epoch	Double	radian
15	DEC-Epoch	dec_epoch	Double	radian
16	Catalog flux at mean observing freq.	flux_cat	Float	Jansky
Observing geometry specification				
17	Elongation	elongation	Float	radian
18	Sun-Earth distance	sun_earth_dist	Float	AU
19	P-pt distance	p_pt_dist	Float	AU
20	Carrington Rotation No. of P-pt	carrington_rotn	Float	Number
21	Helio. Lat of P-pt	p_pt_lat	Float	radian
22	Helio. Long of P-pt	p_pt_long	Float	radian
23	Helio. Lat of Earth	earth_lat	Float	radian
24	Helio. Long of Earth	earth_long	Float	radian
Results of analysis and fiducial model fit				
25	Observed flux	flux	Float	Jansky
26	Scintillation index	m	Float	Number
27	g-value	g	Float	Number
28	Velocity	velocity	Float	km s ⁻¹
29	Cn2	cn2	Float	m ^{-2/3}
30	Alpha (turbulence index)	alpha	Float	Number
31	Axial Ratio	axial_ratio	Float	Number
32	Inner Scale	q_in	Float	km ⁻¹
33	Source Size	source_size	Float	milli arcsec
Database specific				
34	Path to data	path	String(TBD)	
35	Software Version/Build ID.	ver_id	Integer	Number

A set of time series of intensity fluctuations, averaged over sufficiently small time (≈ 10 ms) and frequency (≈ 1 MHz) bins to avoid losing any information, will comprise *Level 0* data for both single-station and multi-station observations of IPS. For single-station IPS data, a time series of power spectra derived from each sub-band in the Level 0 data, log averaged in frequency space to bring down the data volume by a factor of ≈ 20 , are defined to be *Level 1* data. A time-averaged spectrum of these individual spectra is also generated and stored as a part of Level 1 data. In order to provide a quick feel for the level of solar wind activity and data quality, the data repository will also include the results of power spectral fitting of a fiducial model to the Level 1 spectrum and these will constitute the *Level 2* data. Appropriate higher level data products will similarly be defined for multi-station data as well.

5.2. HIPSS Overview

A block level view of the HIPSS architecture and data flow is shown in Figure 2. The data and some of the meta-data will be generated by the individual instruments or using simulation software. A data structure has been defined as a part of HIPSS. A Python based tool will receive the data from observatory specific software, verify its validity and populate the appropriate data structure in the repository. This process may take place off-line or in real-time. The current expectation is that each observatory will adapt the provided software to supply data to this Python tool. The appearance of new Level 0 data in the repository will trigger the generation of Level 1 data, followed by the generation of Level 2 data. A mechanism to synchronize the databases across different installations will be provided to allow data from different observatories to be uploaded into a given installation of the database. A flexible Python based query engine will provide the user interface to the data repository. It will provide the functionality to extract all the level 0, 1 and/or 2 data from the repository matching a query condition. The query output can itself be a HDF5 database, thus allowing all the tools available for browsing, visualizing, and analyzing the repository to become available for this sub set. Options like plain text ASCII tables output will also be supported.

In addition to the meta-data associated with the IPS data itself, information about the state of the Sun and space weather obtained from other independent sources provides very useful context for analysis of IPS data. These could range from lists of Type-II bursts observed across the world and CMEs observed by the various coronagraphs, to time series of sunspot numbers and 10.7cm radio flux. These data fall in two distinct categories, *lists* and *time series*, they are both simple to organize but data structures well suited for them are quite different from the ones needed for the IPS data. These data will therefore be organized in a simple database outside the primary HDF5 database, and is represented by the smaller blue box in Figure 2.

5.3. Database Organization

IPS data are inherently hierarchical and HDF5 is optimized to host such data. The UNIX/Linux file system provides a convenient mental picture for visualizing

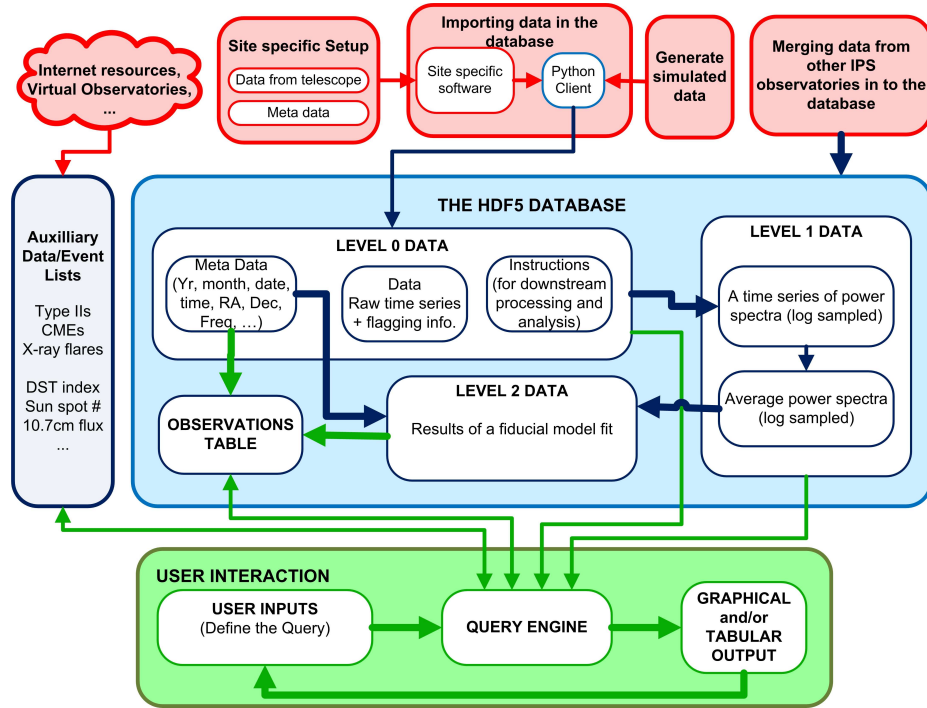


Figure 2. The figure shows the general organization of HIPSS and how it fits into the data flow from the instruments or simulations all the way to the output generated in response to user queries. The boxes in red correspond to aspects related to generating the information to be hosted by the database and importing it into the database, the large blue box shows the database itself, and the green boxes represent the data mining or querying aspects. The arrows are color-coded in a similar manner and their direction represents the flow of data or information. The thin arrows represent data flows which involve little or no processing while the thick arrows represent data flows which require significant processing.

the HIPSS database organization. The data are organized in a strict hierarchy of *nodes*, analogs of UNIX/Linux directories. At the root level are the nodes for each of the different observatories which contribute data to the repository. The hierarchy of the nodes inside the observatory node is *year*, *date*, *beam*, *band* and *sub-band*. The root group can only contain the *observatories* group, the *observatories* group can only contain *year* groups, the *year* group contains only *date* groups and so on. The innermost group, *sub-band*, hosts the time series and can have three groups: *spectra*, which will host the individual and the time-averaged spectra, *bad*, which will host the information of which data needs to be flagged, and *model*, which will store the results of the model fit to the data, along with some information about the model itself.

5.4. Data volumes

Ideally, this data repository will eventually host all existing IPS data, including all available archival data from all the observatories. In fact, in the early years it is the science with archival data which will perhaps be most rewarding. For the

new data coming in, the data volumes will be dominated by the new generation instruments like the MWA. In a given year the MWA expects to make $\approx 1.1 \times 10^6$ individual observations using its 32 beams, which correspond to ≈ 5 TB. Based on MWA data alone, a 50 TB database with 11×10^6 observations is needed to host a decade worth of data. Though large enough to pose an interesting challenge, a redundant database to accommodate this data volume is already well within data storage capabilities available today.

5.5. Optimization for Querying

One of the advantages of keeping all the IPS data in a single repository is the ease of making quick queries to this database to identify the data suitable for a particular study. A hierarchical data architecture implemented as an HDF5 database, however, does not facilitate fast querying. Searches through directory trees are slow and additionally, the data needed to evaluate the query is sprinkled sparsely across large data volumes making rapid data access difficult. In order to optimize for query speed, all the meta-data fields which could potentially figure in a query have been compiled in a single simple table, referred to as the *Observations Table*. This table has one row corresponding to each observation from each of the beams of all the instruments in the database. Each row contains the exhaustive set of meta-data available about that particular observation. A complete list of all the columns in the observations table is provided in Table 2. This concentrates all the information which needs to be accessed for querying at one location in a time-ordered table, making it far more suitable for querying. This is indicated in Figure 2 by the fact that the only double green arrow, which shows the query related data flow, is between the query engine and the observations table. Our experiments to evaluate the query time as function of query complexity and number of rows in the observations table show that using this approach the query times even for the most complex query, involving all the meta-data parameters over a table with 10^7 rows, is ≈ 10 s, small enough for the process to be comfortably interactive³.

6. Conclusions

IPS has a long history of contributing to improving our understanding of the inner heliosphere. The expected availability of substantially more capable new instruments in near future along with significant upgrades of existing instrumentation will lead to a vast increase in the quality and quantity of IPS data captured worldwide. In an era where our means to probe the Sun and heliosphere have grown to include HIs, IPS data will be particularly relevant as it will provide complementary information not available from other sources.

One of the most interesting IPS instruments expected to come online in the near future is the MWA. With its multiple, independent, simultaneous and sensitive digital beams, the MWA will lead to a large increase in the sampling of

³These tests were conducted using one of the cores of a 2.5 GHz Intel Xeon quad-core CPU with 16 GB of RAM

the heliosphere. In addition, with its ability to do high dynamic range and high temporal and spectral resolution imaging of the Sun, and look for FR signatures of heliospheric magnetic fields and CMEs, the MWA will be able to trace the drivers of space weather all the way from the Sun to the terrestrial neighborhood and the ionosphere.

There is a desire in the IPS community to make optimal use of these data from the new and upgraded instruments by synthesizing them into a virtual global IPS array. Using a common data format and analysis procedures, and cross-calibrated observations from different instruments and IPS techniques are among the essential requirements for this process. The design of the HIPSS software suite described here has its origin in the emerging consensus across the world-wide IPS community to develop common software tools and represents a first step in this direction.

Acknowledgements The MWA project is a collaboration of Australian, Indian and US institutions. The Australian partner institutions are the Australian National University, Commonwealth Scientific and Industrial Research Organization, Curtin University of Technology, Swinburne University of Technology, University of Melbourne, University of Sydney, University of Tasmania and University of Western Australia. Raman Research Institute is the partner institution from India, and the US partner institutions are Harvard-Smithsonian Center for Astrophysics, MIT Haystack Observatory and MIT Kavli Institute for Space Research. This work was supported under grants from the National Science Foundation, Division of Astronomical Sciences and Division of Atmospheric and Geospace Sciences. We thank the IPS observatories across the world for their keen and sustained support in working towards the vision of a global IPS instrument. These observatories include European Incoherent SCATter Radar telescopes in northern Scandinavia, LOW Frequency ARray in Europe, Mexico Array Telescope in Mexico, Ooty Radio Telescope in India, Puschino Radio Astronomy Observatory in Russia and, the Solar-Terrestrial Environment Laboratory in Japan. We thank Francesc Alted, author of PyTables, for reviewing and commenting on the HIPSS database design.

References

- Asai, K., Kojima, M., Tokumaru, M., Yokobe, A., Jackson, B.V., Hick, P.L., Manoharan, P.K.: 1998, Heliospheric tomography using interplanetary scintillation observations. III - Correlation between speed and electron density fluctuations in the solar wind. *J. Geophys. Res.* **103**, 1991. doi:10.1029/97JA02750.
- Bastian, T.S., Gary, D.E.: 1997, On the feasibility of imaging coronal mass ejections at radio wavelengths. *J. Geophys. Res.* **102**, 14031–14040. doi:10.1029/97JA00483.
- Bastian, T.S., Pick, M., Kerdraon, A., Maia, D., Vourlidas, A.: 2001, The Coronal Mass Ejection of 1998 April 20: Direct Imaging at Radio Wavelengths. *Astrophys. J. Lett.* **558**, L65–L69. doi:10.1086/323421.

- Bisi, M.M., Fallows, R.A., Breen, A.R., Rifai Habbal, S., Jones, R.A.: 2007, Large-scale structure of the fast solar wind. *Journal of Geophysical Research (Space Physics)* **112**, 6101. doi:10.1029/2006JA012166.
- Bisi, M.M., Jackson, B.V., Buffington, A., Clover, J.M., Hick, P.P., Tokumaru, M.: 2009, Low-Resolution STELab IPS 3D Reconstructions of the Whole Heliosphere Interval and Comparison with in-Ecliptic Solar Wind Measurements from STEREO and Wind Instrumentation. *Solar Phys.* **256**, 201–217. doi:10.1007/s11207-009-9350-9.
- Bisi, M.M., Fallows, R.A., Breen, A.R., O’Neill, I.J.: 2010, Interplanetary Scintillation Observations of Stream Interaction Regions in the Solar Wind. *Solar Phys.* **261**, 149–172. doi:10.1007/s11207-009-9471-1.
- Breen, A.R., Riley, P., Lazarus, A.J., Canals, A., Fallows, R.A., Linker, J., Mikic, Z.: 2002, The solar wind at solar maximum: comparisons of EISCAT IPS and in situ observations. *Annales Geophysicae* **20**, 1291–1309.
- Cairns, I.H., Knock, S.A., Robinson, P.A., Kuncic, Z.: 2003, Type II Solar Radio Bursts: Theory and Space Weather Implications. *Space Science Reviews* **107**, 27–34. doi:10.1023/A:1025503201687.
- Coles, W.A., Rickett, B.J., Rumsey, V.H., Kaufman, J.J., Turley, D.G., Ananthakrishnan, S., Armstrong, J.W., Harmons, J.K., Scott, S.L., Sime, D.G.: 1980, Solar cycle changes in the polar solar wind. *Nature* **286**, 239–241. doi:10.1038/286239a0.
- de Vos, M., Gunst, A.W., Nijboer, R.: 2009, The LOFAR Telescope: System Architecture and Signal Processing. *IEEE Proceedings* **97**, 1431–1437. doi:10.1109/JPROC.2009.2020509.
- Eyles, C.J., Simnett, G.M., Cooke, M.P., Jackson, B.V., Buffington, A., Hick, P.P., Waltham, N.R., King, J.M., Anderson, P.A., Holladay, P.E.: 2003, The Solar Mass Ejection Imager (Smei). *Solar Phys.* **217**, 319–347.
- Gopalswamy, N.: 2000, Type II Solar Radio Bursts. In: R. G. Stone, K. W. Weiler, M. L. Goldstein, & J.-L. Bougeret (ed.) *Radio Astronomy at Long Wavelengths*, 123.
- Harrison, R.A., Davies, J.A., Rouillard, A.P., Davis, C.J., Eyles, C.J., Bewsher, D., Crothers, S.R., Howard, R.A., Sheeley, N.R., Vourlidas, A., Webb, D.F., Brown, D.S., Dorrian, G.D.: 2009, Two Years of the STEREO Heliospheric Imagers. Invited Review. *Solar Phys.* **256**, 219–237. doi:10.1007/s11207-009-9352-7.
- Hewish, A., Scott, P.F., Wills, D.: 1964, Interplanetary Scintillation of Small Diameter Radio Sources. *Nature* **203**, 1214–1217. doi:10.1038/2031214a0.

- Jackson, B.V., Hick, P.L., Kojima, M., Yokobe, A.: 1998, Heliospheric tomography using interplanetary scintillation observations 1. Combined Nagoya and Cambridge data. *J. Geophys. Res.* **103**, 12049–12068. doi:10.1029/97JA02528.
- Jackson, B.V., Hick, P.P., Buffington, A., Bisi, M.M., Kojima, M., Tokumaru, M.: 2007, Comparison of the extent and mass of CME events in the interplanetary medium using IPS and SMEI Thomson scattering observations. *Astronomical and Astrophysical Transactions* **26**, 477–487. doi:10.1080/10556790701612221.
- Jensen, E.A., Hick, P.P., Bisi, M.M., Jackson, B.V., Clover, J., Mulligan, T.: 2010, Faraday Rotation Response to Coronal Mass Ejection Structure. *Solar Phys.*
- Joshi, M.N., Swarup, G., Bagri, D.S., Kher, R.K.: 1988, A new electronically steerable 1056 dipole array at 327 MHz for the Ooty radio telescope. *Bulletin of the Astronomical Society of India* **16**, 111–121.
- Kojima, M., Kakinuma, T.: 1990, Solar cycle dependence of global distribution of solar wind speed. *Space Science Reviews* **53**, 173–222. doi:10.1007/BF00212754.
- Kojima, M., Tokumaru, M., Watanabe, H., Yokobe, A., Asai, K., Jackson, B.V., Hick, P.L.: 1998, Heliospheric tomography using interplanetary scintillation observations 2. Latitude and heliocentric distance dependence of solar wind structure at 0.1-1 AU. *J. Geophys. Res.* **103**, 1981–1990. doi:10.1029/97JA02162.
- Kojima, M., Tokumaru, M., Fujiki, K., Ishida, Y., Ohmi, T., Hayashi, K., Yamashita, M.: 2003, Solar wind imaging facility (SWIFT) for space weather research. In: S. L. Keil & S. V. Avakyan (ed.) *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **4853**, 121–128.
- Liu, Y., Manchester, W.B. IV, Kasper, J.C., Richardson, J.D., Belcher, J.W.: 2007, Determining the Magnetic Field Orientation of Coronal Mass Ejections from Faraday Rotation. *Astrophys. J.* **665**, 1439–1447. doi:10.1086/520038.
- Lonsdale, C.J., Cappallo, R.J., Morales, M.F., Briggs, F.H., Benkevitch, L., Bowman, J.D., Bunton, J.D., Burns, S., Corey, B.E., Desouza, L., Doleman, S.S., Derome, M., Deshpande, A., Gopala, M.R., Greenhill, L.J., Herne, D.E., Hewitt, J.N., Kamini, P.A., Kasper, J.C., Kincaid, B.B., Kocz, J., Kowald, E., Kratzenberg, E., Kumar, D., Lynch, M.J., Madhavi, S., Matejek, M., Mitchell, D.A., Morgan, E., Oberoi, D., Ord, S., Pathikulangara, J., Prabu, T., Rogers, A., Roshi, A., Salah, J.E., Sault, R.J., Shankar, N.U., Srivani, K.S., Stevens, J., Tingay, S., Vaccarella, A., Waterson, M., Wayth, R.B., Webster, R.L., Whitney, A.R., Williams, A., Williams, C.: 2009, The Murchison Widefield Array: Design Overview. *IEEE Proceedings* **97**, 1497–1506. doi:10.1109/JPROC.2009.2017564.

- Magdaleníć, J., Vršnak, B., Pohjolainen, S., Temmer, M., Aurass, H., Lehtinen, N.J.: 2008, A Flare-Generated Shock during a Coronal Mass Ejection on 24 December 1996. *Solar Phys.* **253**, 305–317. doi:10.1007/s11207-008-9220-x.
- Maia, D.J.F., Gama, R., Mercier, C., Pick, M., Kerdraon, A., Karlický, M.: 2007, The Radio-Coronal Mass Ejection Event on 2001 April 15. *Astrophys. J.* **660**, 874–881. doi:10.1086/508011.
- Manoharan, P.K.: 2006, Evolution of Coronal Mass Ejections in the Inner Heliosphere: A Study Using White-Light and Scintillation Images. *Solar Phys.* **235**, 345–368. doi:10.1007/s11207-006-0100-y.
- Manoharan, P.K., Kojima, M., Gopalswamy, N., Kondo, T., Smith, Z.: 2000, Radial Evolution and Turbulence Characteristics of a Coronal Mass Ejection. *Astrophys. J.* **530**, 1061–1070. doi:10.1086/308378.
- Mejia-Ambriz, J., Villanueva-Hernandez, P., Gonzalez-Esparza, J.A., Aguilar-Rodriguez, E., Jeyakumar, S.: IPS observations at the Mexican Array Radio Telescope (MEXART).
- Mitchell, D.A., Greenhill, L.J., Wayth, R.B., Sault, R.J., Lonsdale, C.J., Cappallo, R.J., Morales, M.F., Ord, S.M.: 2008, Real-Time Calibration of the Murchison Widefield Array. *IEEE Journal of Selected Topics in Signal Processing, Vol. 2, Issue 5, p.707-717* **2**, 707–717. doi:DOI: 10.1109/JSTSP.2008.2005327.
- Moran, P.J., Ananthakrishnan, S., Balasubramanian, V., Breen, A.R., Canals, A., Fallows, R.A., Janardhan, P., Tokumaru, M., Williams, P.J.S.: 2000, Observations of interplanetary scintillation during the 1998 Whole Sun Month: a comparison between EISCAT, ORT and Nagoya data. *Annales Geophysicae* **18**, 1003. doi:10.1007/s005850000281.
- Oberoi, D.: 2000, Tomography of the Solar Wind using Interplanetary Scintillations. PhD thesis, Indian Institute of Science.
- Schmidt, J.M., Gopalswamy, N.: 2008, Synthetic radio maps of CME-driven shocks below 4 solar radii heliocentric distance. *Journal of Geophysical Research (Space Physics)* **113**, 8104. doi:10.1029/2007JA013002.
- Scott, S.L., Rickett, B.J., Armstrong, J.W.: 1983, The velocity and the density spectrum of the solar wind from simultaneous three-frequency IPS observations. *Astron. Astrophys.* **123**, 191–206.
- Sheridan, K.V., Jackson, B.V., McLearn, D.J., Dulk, G.A.: 1978, Radio observations of a massive, slow moving ejection of coronal material. *Proceedings of the Astronomical Society of Australia* **3**, 249–250.
- Tokumaru, M., Kojima, M., Fujiki, K.: 2010, Solar cycle evolution of the solar wind speed distribution from 1985 to 2008. *Journal of Geophysical Research (Space Physics)* **115**, 4102. doi:10.1029/2009JA014628.

- White, S.M., Mercier, C., Bradley, R., Bastian, T., Kerdraon, A., Pick, M.: 2006, Imaging Observations of a Very High Frequency Type II Burst. *AGU Spring Meeting Abstracts*, A5.
- Wild, J.P.: 1950, Observations of the Spectrum of High-Intensity Solar Radiation at Metre Wavelengths. III. Isolated Bursts. *Australian Journal of Scientific Research A Physical Sciences* **3**, 541.
- Zhang, X.: 2007, A Study on the Technique of Observing Interplanetary Scintillation with Simultaneous Dual-Frequency Measurements. *Chinese Journal of Astronomy and Astrophysics* **7**, 712–720. doi:10.1088/1009-9271/7/5/12.

