LASER INTERFEROMETER GRAVITATIONAL WAVE OBSERVATORY
- LIGO -
CALIFORNIA INSTITUTE OF TECHNOLOGY
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Document Type	**LIGO-T010095-00-Z**	2001/08/21

Conventions for data and software products of the LIGO and the LSC

Stuart Anderson, Warren Anderson, Kent Blackburn, Patrick Brady Duncan Brown, Philip Charlton, Jolien Creighton, Teviet Creighton, Sam Finn, Joseph Romano, Daniel Sigg, John Whelan, Alan Wiseman, and John Zweizig

Distribution of this draft:

All

California Institute of Technology
LIGO Project - MS 51-33
Pasadena CA 91125
Phone (626) 395-2129
Fax (626) 304-9834
E-mail: info@ligo.caltech.edu

Massachusetts Institute of Technology
LIGO Project - MS 20B-145
Cambridge, MA 01239
Phone (617) 253-4824
Fax (617) 253-7014
E-mail: info@ligo.mit.edu

WWW: http://www.ligo.caltech.edu/

Processed with LaTeX on 2001/08/21

Contents

1 Preamble

This technical note addresses the need for uniform conventions across the multiple software packages being developed by the LSC. The need for concrete conventions is clear if different components of the software are to function together effectively and correctly. The LaTeX source for this document is in under CVS and may be obtained as follows:

```
export CVSROOT=":pserver:anonymous@gravity.phys.uwm.edu:/usr/local/cvs/lscdocs"
cvs login
cvs co T/T010095
```

The password is `lscdocs`.

1.1 Scope of the conventions [REQUIRED]

- The conventions in this document will apply to all data and software products within the LIGO Scientific Collaboration and the Ligo Laboratory.
- This is a living document. A note in each section heading will indicate whether the recommended convention has been adopted by the LSC and the LIGO Lab:

  REQUIRED Indicates that the conventions discussed in this section have been considered by the Software Coordinator, the LSC and the Lab Directorate and have been adopted as the standard convention.

  PENDING Indicates that a recommendation has been forwarded to the software coordinator for consideration.

  RECOMMENDED Indicates that these conventions are being adopted by some subset of the collaboration and would be usefully adopted by all.

- Changes to **REQUIRED** conventions must be forwarded to the Software Coordinator as a formal change request as described in the charter of the Software Change Control Board.

2 Conventions for Discrete Fourier Transforms

2.1 Forward and reverse DFT [REQUIRED]

Consider a time domain quantity $h(t)$ sampled at N discrete points with sampling interval Δt such that $h_j = h(t_j)$. The adopted convention for the DFT is

$$\tilde{h}_k = \sum_{j=0}^{N-1} h_j e^{-i2\pi jk/N} \ , \tag{1}$$

and the inverse DFT is then

$$h_j = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{h}_k e^{i2\pi jk/N} \ , \tag{2}$$

where $i = \sqrt{-1}$.

2.2 Format for DFT's in frequency series [PENDING]

In practice, the time domain data may be windowed before it is transformed. If the window function is described by N real numbers w_j, then the windowed DFT is given by

$$\tilde{H}_k = \sum_{j=0}^{N-1} w_j h_j e^{-i2\pi jk/N} \ , \tag{3}$$

where $i = \sqrt{-1}$. When defining the normalization for frequency series, we make reference to a quantity σ_w which depends on the window function and is unity for a uniform window:

$$\sigma_w^2 = \frac{1}{N} \sum_{j=0}^{N-1} w_j^2 \ . \tag{4}$$

The following conventions should be followed for frequency series:

- Meta-data consistent with the tabular description in Appendix B should be provided.
- The vector $\mathtt{data}[p]$ containing the frequency series data should be packed according to the following rules:
  1. $\mathtt{data}[0]$ should contain the lowest frequency component of the frequency series, i.e. the component at frequency `f0` which can be negative. For a complex time series with an even number N of points, we adopt the convention that $\tilde{h}_{N/2}$ or $\tilde{H}_{N/2}$ is the lowest frequency component of the DFT.
  2. The frequency associated with the pth element of the vector is given by $\mathtt{f0} + \mathtt{p} * \mathtt{df}$. Thus, the frequency is always monotonically increasing with the index p.
- The DFT $\tilde{H}_k$ should be multiplied by the following normalization constant when packed into a frequency series:

$$\mathcal{A} = \frac{\Delta}{\sigma_w} \ . \tag{5}$$

  Thus, elements of the vector $\mathtt{data}[p]$ will have units of seconds $\times$ (Units of time series). Consider a concrete example involving the DFT of real data, then

$$\mathtt{data}[k] = \frac{\Delta}{\sigma_w} \times \tilde{H}_k \tag{6}$$

  for $0 \leq k \leq [N/2]$ where $\tilde{H}_k$ is defined in Eq. (3), Δ is the sampling interval of h_j, and σ_w is defined in Eq. (4). The notation $[x]$ means x rounded down to the nearest integer. The metadata $\mathtt{f0} = 0$ would indicate that $k = 0$ corresponds to DC.

3 Conventions for Power Spectral Densities [REQUIRED]

In this section, we discuss the normalization convention for the power spectral density (PSD) as used within the LSC. Estimation techniques for PSDs are not discussed, *per se*, although some examples are used to demonstrate the normalization explicitly.

For a vector of N time samples h_j with sampling interval Δ, the periodogram provides an estimate of the power distribution in the frequency domain. Assuming $k = 0$ corresponds to DC, the periodogram is defined by

$$P_0 = \frac{\Delta}{N} |\tilde{h}_0|^2 \tag{7}$$

$$P_k = \frac{\Delta}{N} \left\{ |\tilde{h}_k|^2 + |\tilde{h}_{N-k}|^2 \right\} \tag{8}$$

where $\tilde{h}_k$ is given in Eq. (1) and $1 \leq k \leq [(N-1)/2]$. The notation $[x]$ means x rounded down to the nearest integer. When N is even, the Nyquist component is

$$P_{N/2} = \frac{\Delta}{N} |\tilde{h}_{N/2}|^2 \ . \tag{9}$$

This definition respects Parsevals Theorem in the form

$$\Delta \sum_{j=0}^{N-1} |h_j|^2 = \sum_{k=0}^{[N/2]} P_k \ . \tag{10}$$

The definition of the periodogram serves primarily to demonstrate the adopted normalization. It is only an estimate of the underlying power spectrum, and not a very good one at that.

A better estimate of the PSD can be obtained by averaging together the periodogram computed using several windowed DFTs. In this sense, a better estimate of the power spectrum is

$$S_0 = \frac{\Delta}{N\sigma_w^2} \times \langle\ |\tilde{H}_0|^2 \rangle \tag{11}$$

$$S_k = \frac{\Delta}{N\sigma_w^2} \times \left\{ \langle\ |\tilde{H}_k|^2 \rangle + \langle\ |\tilde{H}_{N-k}|^2 \rangle \right\} \tag{12}$$

where $\tilde{H}_k$ is given in Eq. (3) and $1 \leq k \leq [(N-1)/2]$. The notation $[x]$ means x rounded down to the nearest integer. When N is even, the Nyquist component is

$$S_{N/2} = \frac{\Delta}{N\sigma_w^2} \times \langle\ |\tilde{H}_{N/2}|^2 \rangle \tag{13}$$

The notation $\langle \ldots \rangle$ means average. Once again, Eqs. (11)–(13) serve to demonstrate the normalization convention for PSD's and should not be taken to provide the *correct* low-level implementation.

3.1 Format for PSD's in frequency series [REQUIRED]

Only one-sided PSD's should be recorded as frequency series when the following conventions should be followed:

- Meta-data consistent with the tabular description in Appendix B should be provided.
- The vector `PSD`$[p]$ containing the frequency series data should be packed according to the following rules:
    1. `PSD`$[0]$ should contain the lowest frequency component of the frequency series, i.e. the component at frequency `f0`.

2. The frequency associated with the pth element of the vector is given by $\texttt{f0} + \texttt{p} * \texttt{df}$. Thus, the frequency is always monotonically increasing with the index p.

- The vector containing a PSD will have units of seconds $\times$ (Units of time series)^2. Consider a concrete example involving the power spectrum for real data:

$$\texttt{PSD}[0] = \frac{\Delta}{N\sigma_w^2} \times \langle\, |\tilde{H}_0|^2 \rangle \tag{14}$$

$$\texttt{PSD}[k] = \frac{\Delta}{N\sigma_w^2} \times \left\{ \langle\, |\tilde{H}_k|^2 \rangle + \langle\, |\tilde{H}_{N-k}|^2 \rangle \right\} \tag{15}$$

where $\tilde{H}_k$ is defined in Eq. (3), Δ is the sampling interval of h_j, σ_w is defined in Eq. (4) and $1 \leq k \leq [(N-1)/2]$. When N is even, the Nyquist component is

$$\texttt{PSD}[N/2] = \frac{\Delta}{N\sigma_w^2} \times \langle\, |\tilde{H}_{N/2}|^2 \rangle\ . \tag{16}$$

The notation $[x]$ means x rounded down to the nearest integer. The metadata $\texttt{f0} = 0$ would indicate that $k = 0$ corresponds to DC. Equivalently, the two-sided PSD can be obtained from a frequency series containing a DFT as

$$\texttt{PSD}[0] = \texttt{df} \times \langle\, |\texttt{data}[0]|^2 \rangle \tag{17}$$

$$\texttt{PSD}[k] = \texttt{df} \times \left\{ \langle\, |\texttt{data}[k]|^2 \rangle + \langle\, |\texttt{data}[N-k]|^2 \rangle \right\} \tag{18}$$

$$\tag{19}$$

where $\texttt{data}[k]$ is defined in Eq. (6) and $1 \leq k \leq [(N-1)/2]$. When N is even, the Nyquist component is

$$\texttt{PSD}[N/2] = \texttt{df} \times \langle\, |\texttt{data}[N/2]|^2 \rangle\ . \tag{20}$$

- A sample relationship between the `PSD` and the `amplitude` used to plot instrumental sensitivity is

$$\texttt{amplitude}[k] = \sqrt{\texttt{PSD}[k]} \tag{21}$$

where $0 \leq k \leq [N/2]$, $\texttt{PSD}[k]$ is defined in Eqs. (14)–(16). The notation $[x]$ means x rounded down to the nearest integer.

A Fundamental data types [PENDING]

Name	C/C++ data type on ix86	LAL	Short	Type name
8 bit integer	char	CHAR	c	byte
16 bit integer	short	INT2	s	short
32 bit integer	int	INT4	i	int
64 bit integer	longlong	INT8	ll	long
boolean	bool	BOOLEAN	b	boolean
single precision floating point	float	REAL4	f	float
double precision floating point	double	REAL8	d	double
single precision complex number	complex⟨float⟩	COMPLEX8	zf	floatComplex
double precision complex number	complex⟨double⟩	COMPLEX16	zd	doubleComplex
string	char	CHAR	st/ch	string
time	longlong	LIGOTimeGPS	time	GPS
unit	char	LALUnit	unit	Unit

Notes

The *time* type in LAL follows the `frame spec` storing the time in two `INT4` types (one for seconds and one for the remainder in nsec) rather than a single `INT8`.

The *unit* type in LAL is spcialized to allow machine manipulation of units for verification purposes. This is somewhat restrictive in naming conventions for units (e.g. units of sec / sec are not allowed), but is completely general and benefits easy verification of units at the end.

B Frequency Series [PENDING]

Name	Type	Dim	Man.	Description
Subtype	i	1	x	0 - DFT in format (Y) 1 - DFT in format (f,Y) 2 - Power spectrum (Sec. 3.1) in format (Y) 3 - Power spectrum (Sec. 3.1) in format (f,Y) 4 - cross-spectrum in format (Y) 5 - cross-spectrum in format (f,Y) 6 - coherence in format (Y) 7 - coherence in format (f,Y)
f0	d	1	x	Physical start frequency in Hz.
df	d	1	x	Frequency spacing in Hz
fref	d	1		Physical reference frequency in Hz (must be represented in frequency series)
kref	i	1		Index corresponding to reference frequency fref.
t0	time	1	x	Start time in GPS nsec.
dt	d	1		Temporal spacing in sec (only useful for averaged power spectrum).
powerOfTen	s	1		Multiply data by $10^{\mathrm{powerOfTen}}$ to obtain normalization. (Needed to avoid underflow errors when storing strain power spectra.)
BW	d	1		Resolution bandwidth.
Window	i	1		0 - uniform (no window) 1 - Hanning 2 - Flat-top 3 - Welch 4 - Bartlet 5 - BMH 6 - Hamming 7 - Kaiser Default to *uniform* if unspecified.
Average type	i	1		0 - fixed number 1 - running (exponential weight)
Averages	i	1		Number of averages
ChannelA	ch	1		Channel name
ChannelB[M]	ch	1		2nd channel names for cross spectra
N	i	1	x	Number of points
M	i	1		Number of spectra
Unit	unit	1		Physical unit
	f	$M \times N$ $+N$		Power spectral density/coherence in format (Y) Power spectral density/coherence in format (f,Y)
	zf	$M \times N$ $+N$		DFT/cross-spectrum in format (Y) DFT/cross-spectrum in format (f,Y)