
SumPS

Program and notes by James Conway, 26-Mar-2001.

Program Version: 1.2.7s (18-Mar-2001)

Contact me:

Laboratoire de Microscopie Electronique Structurale
 Institut de Biologie Structurale
 41 rue Jules Horowitz
 38027 Grenoble Cedex 1
France
 Telephone: +33-(0)476-884568
 Fax: +33-(0)476-885494
 Email: James.Conway@ibs.fr

Contents:

1. Overview	1
2. Input/Output.....	2
3. Example	3
4. Using the SumPS output for CTF correction.....	5
5. Known problems	5
6. Program history	5
7. Wish list	6
8. References.....	6

1. Overview

SumPS (Sum Power Spectra) is a 'C' program for calculating the power spectra for a series of images, and averaging them (actually, the amplitudes are averaged). This is a good way of determining characteristics of the Contrast Transfer Function (assuming all the images were extracted from the same scanned micrograph). It is an alternative to another program of mine, OptDiff, which calculates the power spectrum of a single large image, such as the entire scanned micrograph. See the references section for more detail about procedures for CTF correction of cryo-electron micrographs – especially [1].

SumPS averages the power spectra from a series of 2D images extracted from electron micrographs. The average spectrum is saved in both a 4-byte-real format and log-values of the averaged spectrum are saved as a 1-byte unsigned integer file. A 1D profile of the whole spectrum is saved as a data file for importing into a graphing program where characteristics of the contrast transfer function may be estimated (see section 4), and this file may also include angular segments of the spectral plane from which astigmatism may be assessed.

Input images may be one or more series of separate files of the form:

rootnnn.ext

where *nnn* are numbers from 001 up. PIC's BP format² and MRC-files³ are supported in this manner. The numeric field can be 3 to 6 digits wide. When a file in the sequence cannot be found, SumPS assumes that the end of data has been reached. Alternatively, the images may be contained in one or more packed files, such as the Purdue PIF file.⁴

Images may be drawn from several populations of particles on the same micrograph. For example, Hepatitis B virus (HBV) capsids usually come in two sizes with surface lattice geometries of $T=3$ or $T=4$, and bacteriophages HK97, T7 and ϕ have procapsid and mature capsid forms which may both be present on a micrograph. Particles of different types will usually be selected and saved in separate PIF files or sets of BP or MRC images, but the SumPS should be calculated over all the kinds of images from a single micrograph so as to get the best signal-to-noise ratio.

The images will be padded at least to the next highest power of two, and it is probably necessary to pad further to get sufficiently fine sampling in Fourier space. For example, a 115x115 size might be padded to 256x256, 512x512 or 1024x1024. Keep in mind that the larger the size, the slower the program. Generally, I use 1024x1024.

Output images are of a similar format to that specified for input. PIC's 4-byte-real format is a BQ file, and MRC and PIF have 4-byte real variants. Note that PIC formats are only used on OpenVMS systems.

The program uses the "pthread" library for parallelization. This library is available on a wide variety of systems, including Compaq's OpenVMS and SGI's Irix. On a single CPU system this allows some overlap of data input with processing. On multiple-CPU machines, several processing threads may execute concurrently with the data input thread. SumPS has been tested on these operating systems:

Compaq Alpha/OpenVMS 7.1
 Compaq Alpha/Linux-RedHat 6.2.
 SGI Irix 6.5
 Apple Macintosh 8.5-9.1

To run, type: `sumps [-t n]`

[...] indicates optional command line arguments. -t specifies the number of processing threads to be n , an integer from 1 up. The default is 1 thread.

2. Input/Output

- 1 Digitization rate in Ångstroms/pixel. For example, a micrograph taken at a magnification of 38000 and scanned at $7\mu\text{m}/\text{pixel}$ will have an at-the-sample digitization rate of $70000\text{Å}/38000 \Rightarrow 1.842\text{Å}$.
- 2 Size to pad images to (power of 2). I suggest 512 or 1024.
- 3 Number of segments to profile. These are additional to the full-plane average which is always calculated. I don't usually do any segments, in which case I enter "0".
- 4 Input dataset format and filename(s). Dataset of images extracted from the scanned micrograph, using for example the X3dPreprocess program (see the separate document "X3dPreprocess" for information).
- 5 Output format and filenames for the PS images and profile. The spectrum images will be the same format as the input files. The profile is saved as a text file.

3. Example

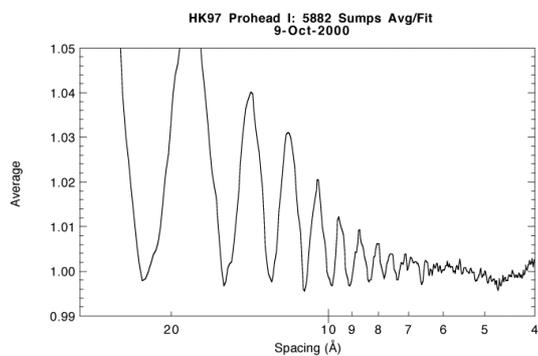
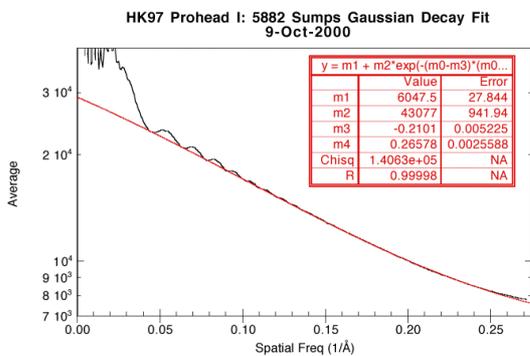
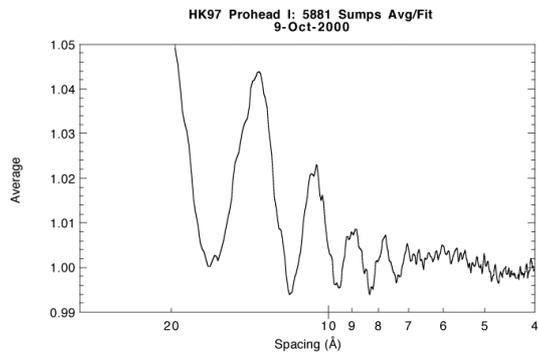
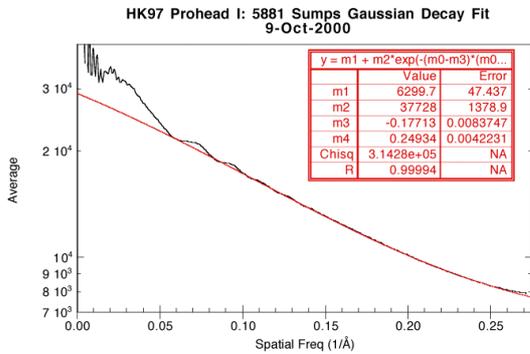
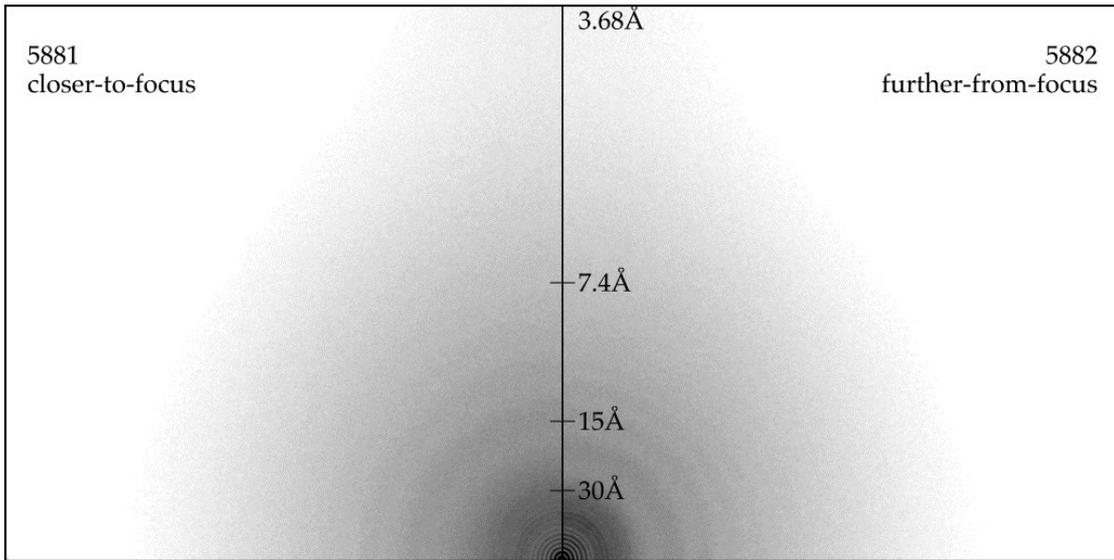
The following is the analysis of images extracted from a defocus pair of micrographs. Two runs of SumPS are made, one for each dataset of images extracted from a micrograph. The averaged PS images are shown, together with plots of the 1D profiles. The sample is HK97 Prohead I from a Phillips CM20 FEG at 38,000, scanned at 7 μ m (1.842 \AA /pixel). Micrographs 5881 and 5882 are a focal pair (closer-to- & further-from-focus, respectively).

Note: user input is in **bold & underline**.

```
SumPS versions 1.2.7s
=====
Compiled: Mar 18 2001, 17:29:55
James F. Conway, LMES/IBS, Grenoble, France.
Problems: email me at James.Conway@ibs.fr
SumPS averages power spectra calculated from a series of 2D images, such as
those extracted from cryo-electron micrographs. The average image is saved
in a 4-byte real format, and the logs are saved in a 1-byte integer file.
In addition, 1D averages of the log values are saved in a text file for the
whole plane, and for segments - these can be used to assess astigmatism.
Input image formats: PIC's BP, Purdue's PIF, MRC modes 0,1,2.
Output PS formats: PIC's BQ&BP, Purdue's PIF, MRC modes 2&0.
This version is not multi-threaded.
Multi-threading disabled
What is the digitization rate (in Angstroms/pixel) ? 1.8421
What size to pad arrays to (power of 2) ? 1024
Calculate 1-d profiles for how many segments (1 up) 10
Enter 2D image file format:
  1 or B: Series of BP files (only on VMS systems)
  2 or P: Packed PIF file
  3 or M: Series of MRC files
Enter choice > p
PIF file chosen: Filename of the form name.pif expected.
>> Enter packed PIF filename >> 5881.pif
Opening PIF file...
- Image data is big-endian (eg, Unix)
- Created from 'JFC-PIF 1.4.8 of Sep 25 2000'
>> Number of images in PIF file: 607.
>> All images in PIF file have same size: 367 x 367.
>> All images in PIF file are 2-dimensional.
Processed: 0
Processed: 7
.....
Processed: 600
Processed: 607
Enter 2D image file format:
  0 or X: Exit
  1 or B: Series of BP files (only on VMS systems)
  2 or P: Packed PIF file
  3 or M: Series of MRC files
Enter choice > x
Completed 1D profiles
Writing results...
Filename for PIF image of PS (real) ? 5881 PS.pif
Writing - 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Filename for PIF image of log(PS) (byte) ? 5881 LOGPS.pif
Writing - 0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
Filename for 1d average ? 5881 PS.dat
Finished.
```

In this example, data from 607 images have been averaged. In addition to the 1D profile over the whole plane, profiles have been calculated within 10 segments, each corresponding to 18° (180°/10). The process was repeated for the further-from-focus

micrograph (5882) and results are shown below. Parts of the two average power spectra images are shown side-by-side for comparison – minima are the light rings – followed by plots of the 1D profiles (see section 4). These plots include a gaussian fit of the minima (left – curves in red) which is subsequently deconvoluted to give the plots on the right where the positions of the minima are more easily determined. The minima correspond to positions of phase reversal in the phase contrast transfer function of the electron microscope. Their positions can be used to determine the defocus value used for each micrograph. Note that only the full-plane profiles have been plotted. In principle, the plots of segments can be used to estimate and correct for astigmatism, but in practice the curves tend to be noisy, and I rarely use them.



4. Using the SumPS output for CTF correction

The 1D profiles shown above were plotted in a Macintosh/Windows program called Kaleidagraph (Synergy Software, www.synergy.com). The Gaussian fit, deconvolution, and estimation of the positions of minima were also done with Kaleidagraph. Instructions for performing these steps are described in another document "SumPS Analysis: CTF Estimation". The minima positions can be used with the program CTFZEROS for estimating the defocus value of the micrograph. These measurements are used by the CTFMIX program to correct images for the contrast transfer function effects so as to allow 3D analysis to 10Å resolution and better.

5. Known problems

- 1 On OpenVMS systems, programs may be terminated with control-C or control-Y. In general, this action should be followed by the EXIT command which cleans up the program's memory and resource footprints. THIS IS EXPECIALLY IMPORTANT WITH A MULTITHREADED PROGRAM SUCH AS SumPS. Failure to do so will likely cause the program to continue execution of an unterminated thread on the input of the next command.
- 2a On a multi-CPU system running OpenVMS 7.0 or earlier, kernal threads were unavailable and a process could only schedule threads on the CPU it was running on. In other words, threads would not be spread over CPU's. The solution - upgrade to 7.1, but see 2b...
- 2b On a multi-CPU system running OpenVMS 7.1, threads do not reliably use all available CPU's, although the execution itself is reliably performed and completed. So performance does not scale with the number of CPU's. The solution - likely fixed in 7.2...

(Note that IRIX 6.5 does not have these problems with pthreads and multiple CPU's).

- 3 A problem with the multi-threaded version on the Alpha/Linux platform has been difficult to track down, hence the current serial-only version.
-

6. Program history

1.2.7s 18-Mar-2001 (s for serial)

* Fix for analysis of more-than-one input file.

1.2.6s 23-Jan-2001 (s for serial)

* Minor fiddles - constants are now typed variables

1.2.5s 10-Jan-2001 (s for serial)

* Compatible with Alpha/Linux (working on it...)

* Disable threading for now, to get it working

1.2.4 2-Mar-2000

* Improved flexibility of file sequence code, in case field width is minimal (eg, 1, 2, 3,...).

1.2.3 16-Apr-1999

* MRC - add an extension argument, and an optional '.' before it.

1.2.2 2-Dec-1998

* Fixed problem reading BP - increased queuing thread's stacksize (OpenVMS only).

1.2.1 26-Nov-1998

* Fix to log image - was not recalculating min and max after taking logs

v1.2 27-Oct-1998

- * Changed from DEC's parallel processing library to pthreads.
- * Added input for number of processing threads, number of segments for 1d profile
- * Added MRC format
- * Options for output image formats
- * Conditional compilation for SGI (BP/BQ)
- * Fixed bug in normalizing 1d-profiles of segments
- * Progress report during processing

v1.1.3 27-Oct-1998 (not released)

- * Add PIF format for input images

v1.1.2 17-Nov-1997

- * Added freq column to output in 1/Å.
- * Normalized output by number of images to aid in comparisons between power spectra calculated from different numbers of images.

v1.1.1 6-May-1997

- * Bug fix - reading input filename root string as %ls instead of %. It worked until we upgraded to OpenVMS 7.1!

v1.1 23-Aug-1996

- * Added code to try opening files with different sized id fields, eg try fred_001.bp, then try fred_0001.bp, etc.
- * Write out BQ of average PS, and BP of logs of average PS.

v1.0 4-May-1996

- * First release, based on FRC3D program structure.

7. Wish list

Let me know!

8. References

1. Conway JF & Steven AC (1999) Methods for reconstructing density maps of "single" particles from cryoelectron micrographs to subnanometer resolution. *J Struct Biol* **128**, 106-118.
<http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10600565&dopt=Abstract>
 2. Trus BL, Kocsis E, Conway JF, Steven AC (1996) Digital image processing of electron micrographs: the PIC system-III. *J Struct Biol* **116**, 61-67.
http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=8742724&dopt=Abstract
 3. Crowther RA, Henderson R, Smith JM (1996) MRC image processing programs. *J Struct Biol* **116**, 9-16.
http://www.ncbi.nlm.nih.gov:80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=8742717&dopt=Abstract
 4. PIF (Purdue Image Format) is described on the following web-page:
http://bilbo.bio.purdue.edu/~baker/programs/doc/sys_info/pif.pdf
-