Chapter 19.8. Use of SPIDER and SPIRE in image reconstruction

A. LEITH, W. BAXTER AND J. FRANK

19.8.1. Introduction

SPIDER (System for Processing Image Data in Electron microscopy and Related fields) was one of the earliest image-processing packages for single-particle reconstruction and electron tomography. It was first released in 1978, and has been continually improved and updated since then (Frank et al., 1981, 1996; Baxter et al., 2007; Shaikh et al., 2008; Yang et al., 2007). It is currently available as free open-source code under Gnu Public License (GPL).

In its most common usage, SPIDER allows a researcher to create a three-dimensional (3D) reconstruction of a macromolecule from a collection of transmission electron micrographs, interpreted as projections showing the molecule in different orientations. Using cryo-electron microscopy these methods, pioneered by Joachim Frank at the Wadsworth Center in Albany (see Frank, 2006), have recently allowed researchers to determine the 3D structures of several macromolecular complexes at near-atomic resolution (~3.8–4.5 Å; Zhou, 2008). These methods provide a way of studying the structure of such complexes in a more natural environment than is possible using X-ray crystallography and of studying conformational changes of molecular machines as they perform their work (Saibil, 2000).

One way in which SPIDER differs from most other software for single-particle reconstruction [Imagic (van Heel et al., 1996; Chapter 19.9), Frealign (Grigorieff, 2007), EMAN2 (Tang et al., 2007; Chapter 19.10), SPARX (Hohn et al., 2007) and Xmipp (Sorzano et al., 2004)] is that it provides alternative methods for many procedures and thus can serve as a developmental platform for testing different approaches to single-particle reconstruction.

For instance, alternative alignment operations and statistical analysis operations can be combined inside scripts to investigate different approaches to overcoming molecular heterogeneity.

The prime reasons for the continued success of SPIDER are its validated reputation for providing the highest-resolution reconstructions, its ease of installation and its comprehensive documentation, which is constantly being updated (Shaikh et al., 2008). SPIDER's procedure language makes it easy to use for controlling the flow of a complex train of operations and addressing problems related to heterogeneity of particles. With the Quick Start Guide and tutorials as an introduction, a new user can quickly acquire competence in the procedure language with only a few hours’ introduction.

The SPIDER system consists of six major components:

(i) SPIDER: the Fortran program with a command interpreter recognizing SPIDER commands and procedure calls.
(ii) SPIDER procedures: text files containing SPIDER commands, parameters, and script-specific operations for conditional execution and looping (see Section 19.8.3).
(iii) Web: a graphical user interface (GUI) for use in Linux and OS X (Frank et al., 1996). Web is available in two different versions, the original X-Window version written in C and a newer Java version.
(iv) SPIDER reconstruction engine (SPIRE), which represents a metastructure that enables procedure files, file numbering and directories required for a single-particle reconstruction project to be managed from a simple GUI (see Section 19.8.4; Baxter et al., 2007).
(v) PubSub: a set of Perl programs that enable SPIDER procedures to be run in parallel on a computer cluster.
(vi) SPIDER documentation: an extensive collection of more than 800 HTML-based documents.

The system is available in versions for Linux, AIX and OS X. Both source code and precompiled binaries for popular platforms are provided. SPIDER is available for free download as a GPL Open Source distribution. Its documentation is available under a Creative Commons Attribution 2.5 Licence.

19.8.2. Basic philosophy of single-particle reconstruction

‘Single-particle reconstruction’ is the term used for the reconstruction of a biological macromolecule from images of a specimen in which the molecule exists in many ‘copies’ in the form of single isolated particles, i.e., with no contact with neighbouring molecules. Since there is no need for crystallization, there is in principle no restriction on the kinds of macromolecules that can be reconstructed, except that they must be above a critical size required for accurate alignment. Combined with cryogenic electron microscopy (cryo-EM), the method is capable of visualizing molecules in their native states. Although the method was originally conceived for a homogeneous population, the introduction of powerful classification techniques is now allowing heterogeneous populations to be disentangled and represented by a series of reconstructions which, suitably ordered, may reflect the development of a system of interacting molecules (a molecular machine) over time.

Each particle image is interpreted as a noisy projection of the three-dimensional Coulomb potential representing the molecule, which is, for practical purposes, identical to the electron-density distribution rendered by X-ray crystallography. The noise (signal-to-noise ratio $\approx 0.1$) is due mainly to the low exposure required to avoid radiation damage. For reconstruction, all projections must be placed in a common coordinate frame. Since, unlike the case in electron tomography, the angles are initially unknown, the most challenging and computationally intensive task in single-particle reconstruction is the determination of particle orientations, usually done in an iterative manner with increasing angular resolution. In the following, it is assumed that a reference density map of a closely related molecular complex is already available. For instance, a map of an empty ribosome may be used as reference for a data set obtained from a ribosome complexed with EF-G, mRNA and tRNA. For ab initio reconstructions of an unknown structure, random-conical and common-lines techniques are available (see Shaikh et al., 2008).

Owing to the oscillatory behaviour of the contrast transfer function (CTF), an entire defocus series must be collected in order to cover the whole range of information in Fourier space. For each micrograph, the defocus is determined by computing the power spectrum and matching it with the CTF. Two strategies are in use in the field: CTF correction is done either at the stage of the raw micrograph, by phase flipping and pooling all data for
reconstruction, or by pooling data into groups with a narrow defocus range, which are then separately reconstructed and combined at the very end using CTF correction by Wiener filtering (Penczek et al., 1997). Although both strategies can be readily realized in SPIDER, the one most thoroughly tested and proven is the latter, sometimes referred to as defocus group reconstruction.

For a step-by-step description of the processing path, we refer to the flow diagram in Fig. 19.8.2.1. After defocus determination, particles are selected and aligned to a set of reference projections using rotational and translational cross-correlation. Eulerian angles yielding the highest cross-correlation are assigned. Next, the data are pooled into defocus groups, defined by the range of allowed defocus variation within each group. The range needs to be small enough to avoid resolution loss. Reconstruction of each defocus group yields a preliminary density map. These density maps are merged by CTF correction via Wiener filtering, yielding an initial CTF-corrected reconstruction. This reconstruction is then used as the reference in the first angular refinement of defocus group reconstructions, which follows the same principle as the initial alignment. Several cycles of angular refinement are used with progressively smaller angular steps, from the initial ~15° to 0.5° or even 0.2°. From a certain point on, comparisons are only done within a narrow cone of the previous Eulerian angle assignment for a given projection. The progress of the refinement is monitored by computing the Fourier shell correlation at each stage and by following the statistics of angle reassignment, watching for evidence of stability.

For supervised classification, not one but two three-dimensional references are used in the projection alignment (Fig. 19.8.2.1), and the assignment of the class and set of Eulerian angles is done according to the parameters of the reference projection producing the highest cross-correlation coefficient. For example, ribosome data may be split on the basis of inter-subunit ratchet rotation, by presenting ‘ratcheted’ and ‘non-ratcheted’ ribosomes as references.

19.8.3. Implementation of single-particle reconstruction in SPIDER

The implementation of the reconstruction strategy presented in the flow diagram (Fig. 19.8.2.1) makes use of the modular features of SPIDER and its hierarchical calling structure (Fig. 19.8.3.1). The computational task of the entire project (level 4, or uppermost in the hierarchy) can be visualized as being composed of building blocks (level 3), each representing a subtask, such as image alignment, defocus determination, angular refinement etc. Each of these building blocks is realized by either a set of elementary commands (level 1) or procedure files containing such commands (level 2). In the reconstruction task, examples of procedures are *deffsc.spi* and *ctf.spi*, which compute the resolution of each defocus group and apply CTF correction, respectively. On the uppermost project level, all data file names, directories and values of parameters must be specified, while the input for lower-level procedures can be variables whose value is not known until the time of execution.

19.8.3.1. Level 1: elementary commands

Commands in SPIDER consist of a short two-to-three letter operation mnemonic and a variable sub-option mnemonic. For example, **BP CG** invokes a program performing back-projection reconstruction using a conjugate gradient algorithm. In the interactive on-line mode, commands are entered at a prompt and the system responds with a series of subsequent prompts for more input specific to the operation the command invokes. In the off-line batch mode, relevant in the context of time-consuming reconstructions, both commands and subsequent inputs are placed in a batch or procedure file in the expected order for successive execution.

---

**Figure 19.8.2.1**
Flow diagram of reference-based single-particle reconstruction by defocus groups. For details, see text.