

EMICOMATBG.DOC

(last update Jan 31, 1997)

CONTENTS

- A. INTRODUCTION
- B. PROGRAM INPUT
- C. PROGRAM EXECUTION
- D. EXAMPLE EMICOMATBG BATCH JOB
- E. FLOW CHART FOR EMICOMATBG PROGRAM

A. INTRODUCTION

EMICOMATBG is the first of two major programs in the icosahedral image reconstruction process that are used to produce a 3D map from image data. EMICOMATBG extracts the data for normal equations from each particle view and combines the normal matrices and solves them to give big G's which are stored in a disk file (DEFAULT = BG.DAT).

This program is now used in place of the programs, EMICOMAT and EMICOBG and is normally run after EMPFT or EMICOGRAD. EMICOLGFB is run immediately after EMICOMATBG.

A big advantage of this routine is that it does NOT produce the EMICOMAT.DAT and *.MAT files previously produced by EMICOMAT. Hence, the only disk space need by EMICOMATBG is that required to store the output of the big G's (BG.DAT). EMICOMAT needed enormous amounts of disk storage especially for jobs with large numbers of particles or requiring large number of annuli for higher resolution analysis. In

THEORY, there EMICOMATBG has no restrictions as to number of particles or resolution limit. Of course, the memory capacity of the computer and disk space available WILL pose ultimate limits if you attempt to step outside these bounds.

B. PROGRAM INPUT

1. ICO_IDIM, NANNULI, FFT_STEPSIZE, ICO_DIAM, IPRINT, INCR, ICO_SPREAD (2I,2F,2I,F)
2. INPUT IMAGE FILENAME (A)
3. ID, THETA, PHI, OMEGA, FFT_ORIGX, FFT_ORIGY, SCALEF (I,6F)

1. ICO_IDIM,NANNULI,FFT_STEPSIZE,ICO_DIAM,IPRINT,INCR,ICO_SPREAD

FORMAT (2I,2F,2I,F)

ICO_IDIM defines the size of the FFT calculated for each image. ICO_IDIM must = 128, 256, 512, or 1024 and it MUST be greater than the largest image dimension (all particle images typically have identical NCOL,NROW dimensions).

NANNULI, the number of annuli in the Fourier transform included in the calculations, defines the highest possible resolution of the final 3D MAP that can be computed. If NANNULI is set too large, then errors are likely to be generated especially for data sets with very few particles. The upper limit to NANNULI depends on the size of the FFT. A warning is given when an attempt is made to exceed this limit.

FFT_STEPSIZE determines how finely or coarsely the FFT is sampled. It is given by the equation:

$$\text{FFT_STEPSIZE} = \frac{(\text{ICO_IDIM} * \text{SCALEF})}{(\text{ICO_DIAM} * \text{ICO_NSAMP})} \quad (\text{transform X-steps})$$

ICO_NSAMP specifies the number of annuli per interval of 1/ICO_DIAM, where ICO_DIAM is the diameter of the 'particle' (original image box dimension) in pixels. ICO_NSAMP = 1 for single or 2 for double sampling of the Fourier transform data. In most instances, ICO_NSAMP is set = 2 for double sampling of the Fourier transform. It is recommended that ICO_NSAMP = 1 only be used during preliminary processing to speed up calculations. Double sampling of the Fourier transforms (ICO_NSAMP=2) should normally be used when the final 3D reconstruction calculations are performed (see pp. 321-323 of Crowther, DeRosier and Klug (1970) Proc. Roy. Soc. Lond. A. Vol. 317 for further discussion).

ICO_DIAM is the diameter of the input particle images.

IPRINT controls the amount of output (mainly for debugging purposes). Extra output is only produced for IPRINT > 0.

NOTE: The following parameter (INCR) is generally no longer used for images of frozen-hydrated specimens from a single micrograph. It is believed that radial scaling is inappropriate for such data since the particle dimensions should be well preserved by the low temperature technique.

INCR was formerly the size of each of NBAND bands (NBAND has been removed from the program) in number of transform annuli. NBAND*INCR must be < ICO_IDIM/2. The width of each annulus is FFT_STEPSIZE.

ICO_SPREAD sets the condition on the mean inverse eigenvalue. This is usually set equal to 1.0 (The DEFAULT). Set to a lower value for a less stringent test.

2. INPUT IMAGE FILENAME (A)

Input to EMICOMATBG is in the form of IMAGE data stored as a group of individual images stored in a single file (PURDUE BYTE-PACKED FORMAT). Read BABE3:[TSB.DOC]EMPROGS.DOC for a description of the IMAGE file format.

3. ID,THETA,PHI,OMEGA,FFT_ORIGX,FFT_ORIGY,SCALEF (I,6F)

ID is the image number--the storage position of the IMAGE data for byte-packed format data.

THETA,PHI,OMEGA are the current values of the three orientation angles (in degrees) for each particle image (obtained from previous runs of EMPFT or EMICOGRAD or EMICOFV).

FFT_ORIGX,FFT_ORIGY are the pixel coordinates of the particle center (the point 0.0,0.0 corresponding to the lower left corner of the boxed particle image). These values are estimated or refined by EMPFT (or with older programs like EMCORORG, EMICOORG, and EMICOORG2).

SCALEF is a radial scale factor (EMPFT or EMICOCMP can be used to determine this). With images of frozen-hydrated particles boxed from a single micrograph, SCALEF = 1.0 (DEFAULT) is normally assumed for all particles.

Examples of different acceptable ways to enter IMAGE data parameters are given in [TSB.DOC]EMICOGRAD.DOC.

C. PROGRAM EXECUTION

The maximum pixel resolution obtained in the three-dimensional reconstruction is estimated in following way:

Assume the particle images were originally boxed in 99 square pixel arrays, with each pixel corresponding to a 0.5 nm sampling interval (e.g. from a 50,000X micrograph digitized at 25 micron

intervals). When 128 by 128 FFT's of such images are computed, and SCALEF = 1.0, and the transform is double-sampled (ICO_NSAMP = 2), then FFT_STEPSIZE = (128*1.0)/(99.0*2) = 0.646 TPU. The resolution obtained then depends on how large NANNULI is set. The maximum radius, IRAD, at which data in the transform are used = NANNULI*FFT_STEPSIZE.

The following table illustrates the relationships between several of these parameters for the above example:

NANNULI	IRAD	RESOLUTION (pixels)	RESOLUTION (nm)
16	11	11.64	5.82
32	21	6.10	3.05
48	32	4.00	2.00
64	42	3.05	1.52
80	52	2.46	1.23

WARNING: it is senseless to set NANNULI large if the data is not reliable to the corresponding resolution. Since the typical rule of thumb is to digitize the data at a pixel resolution that is 3 or 4 times the expected resolution, then IRAD should NORMALLY not exceed ICO_IDIM/3 or ICO_IDIM/4. In the example above, if the IMAGE RESOLUTION is 2.0 nm and it is digitized at 0.5 nm intervals, IRAD should not exceed 128/4 = 32.

NOTE: This program is generally run in BATCH mode. The following is an example batch job command file

D. EXAMPLE EMICOMAT BATCH JOB

```
!      COMMAND FILE FOR RUNNING EMICOMATBG FOR SEVERAL PARTICLES
!
!      THE INPUT ARE AS FOLLOWS:
!
!      1. ICO_IDIM, NANNULI, FFT_STEPSIZE, ICO_DIAM, IPRINT,
!         INCR,ICO_SPREAD
!         (ICO_IDIM must be 128, 256, 512, or 1024)
!      2. INPUT IMAGE FILENAME
!      3. ID,THETA,PHI,OMEGA,FFT_ORIGX,FFT_ORIGY,SCALEF (I,6F)
!
$ RUN SYS$EXELIB:EMICOMATBG.EXE
256, 30, 0.757, 169.0, 0, 0, 1.0
1019-box.PCK
10, 85.000, 20.076, 169.805, 85.658, 84.264, 1.002
13, 79.000, 9.168, 53.789, 90.083, 80.719, 1.002
```

15,	88.000,	-28.017,	330.117,	88.976,	83.627,	1.002
20,	80.000,	9.139,	108.984,	87.322,	81.299,	1.001
35,	87.000,	-15.021,	87.188,	83.139,	79.838,	1.002
38,	82.000,	-16.157,	216.914,	86.053,	81.466,	1.000
45,	88.000,	25.015,	211.641,	88.203,	79.416,	1.002
47,	89.000,	-28.004,	169.805,	86.051,	81.753,	0.998
58,	87.000,	-25.034,	346.992,	85.577,	82.448,	0.998
62,	88.000,	26.016,	203.555,	87.421,	81.606,	0.998
64,	81.000,	12.150,	107.227,	86.847,	79.774,	0.998
65,	84.000,	-19.105,	133.594,	83.676,	78.427,	0.998
66,	75.000,	3.106,	201.094,	86.338,	79.218,	1.000
67,	81.000,	14.175,	18.281,	85.537,	76.735,	0.998
68,	79.000,	-11.206,	20.391,	85.976,	81.223,	1.000

\$ WAIT

\$ EXIT

E. FLOW CHART FOR EMICOMATBG PROGRAM

```

*****
*      MAIN      *
* (EMICOMATBG.FOR) *
*****
*
*      | - PIRADDEG* | - STRING_UPPER*
*- ICO_MATBG_INFO - | - IMG_OPEN -- | - IMG_PACK_BIMGCOM_CLEAR*
*
*      | - FILE_CHECK*
*- ICO_MATBG_GETPARAMS*
*
*      | - PIRADDEG*
*      | - IMG_PACK_FIND - IMG_PACK_NRECS*
*      | - IMG_FILL - IMG$FILL(IMG$PACK_FILL)
*- ICO_MATBG_FFTS - | - IMG_FFT - [see below]
*      | - FFT_TO_ATBT - PIRADDEG*
*      | - ICO_ATBT_STORE*
*
*      | - ICO_EM4IMR*
*      | - ICO_EMMEET*
*      | - ICO_EM4RMI*
*- ICO_MATBG - | - ICO_ZEROZ* | -
ICO_LOCATE*
*      | - ICO_INTRP* | - ICO_TRIDI*
*      | - ICO_NMAT* | - ICO_HOW ----- | -
ICO_EIGVAL*
*      | - ICO_BGSOLVE - | - ICO_BGPROD* | -
ICO_EIGVEC*
*      | - ICO_BGOUT* | -
ICO_RELOCT*
*- TSBEND

*
*- IMG__FFT - | - IMG_FFT_FILL - FFT_CLEAR*
*      | - FFT_2D - FOURT - L6TOL9*

```