1st STEPS IN CCPEM

Chris Wood
STFC Rutherford Appleton Laboratory
Scientific Computing Department
chris.wood@stfc.ac.uk

About me

- No EM background
- BSc Chemistry / PhD Biogeochemical modelling
- Junior computational scientist post for CCPEM:

  "To build up the general infrastructure for cryo-EM, detailed in the Case for Support. Duties will include helpdesk support, porting of programs, maintenance of the web site, development of tutorials, etc. "

Science & Technology Facilities Council

Achieved so far…

- Website (www.ccpem.ac.uk)
- Collaborations started (Birkbeck, Manchester, Bristol, Oxford… + ~15 supporters on the grant yet to visit)
- Community meeting organised - ~35 attendees
- Software repository setup (http://ccpforge.cse.rl.ac.uk/gf/project/ccpem) – also potential place to host wiki

Science & Technology Facilities Council

Current Projects

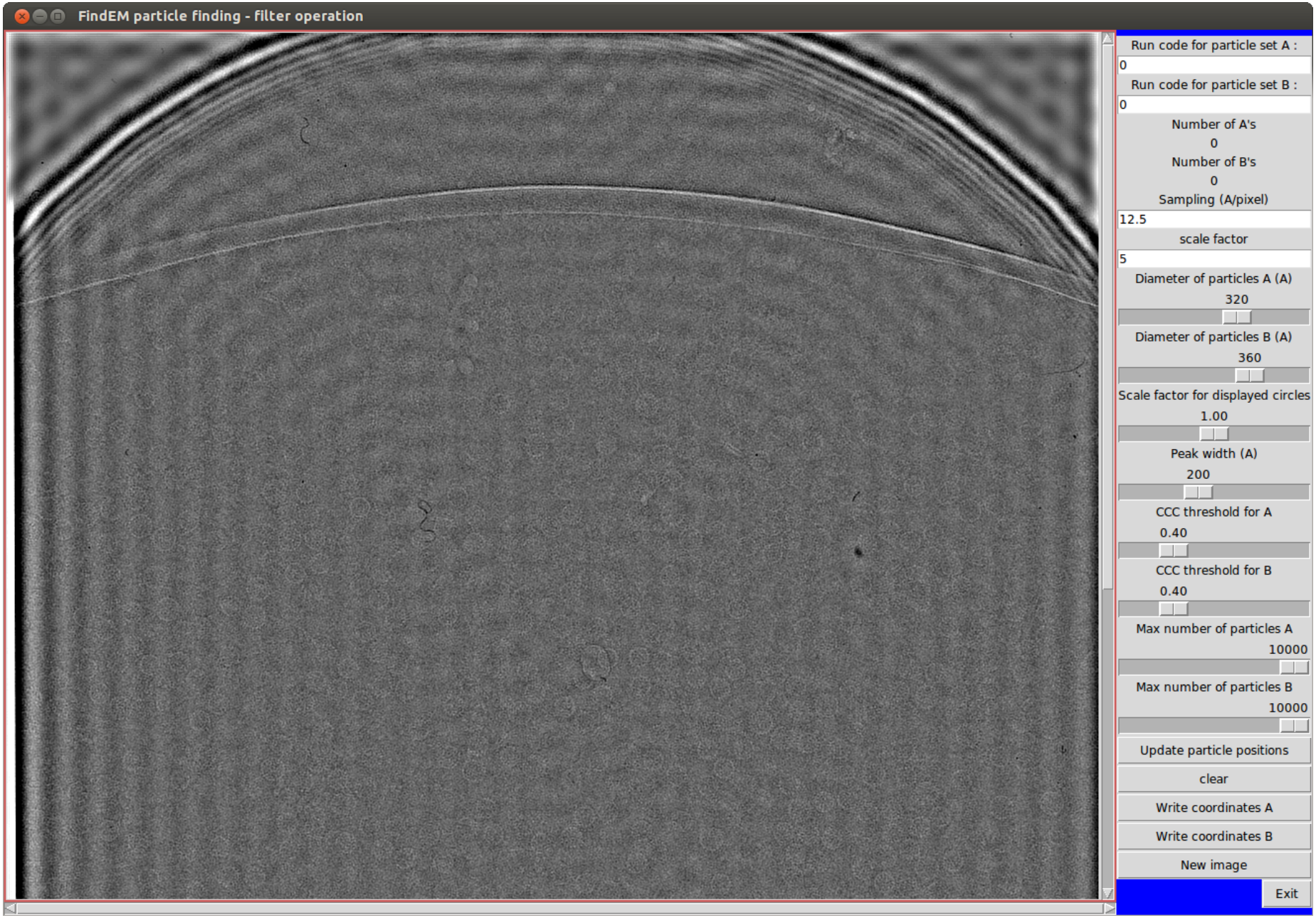
Science & Technology
Facilities Council

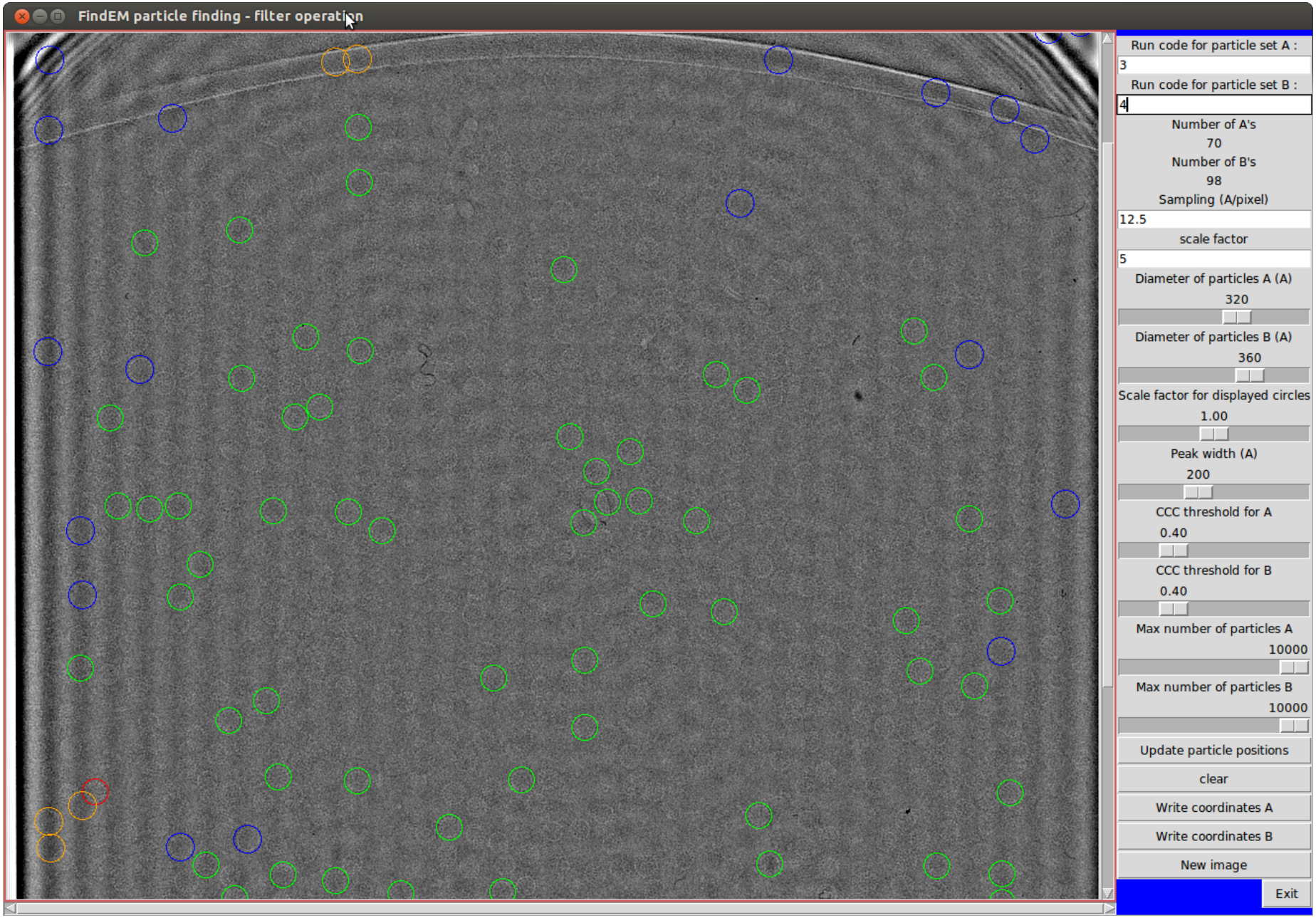
FindEM

- Automated particle-picking from micrographs
- Developed by Alan Roseman
- 5 step process:
  - Bandpass filter image and template, and make them the same scale (2 templates can be used)
  - Perform correlations (FLCF)
  - Process results
  - Convert .mrc file(s) to .gif
  - Use GUI to display and filter the particles found
- Roseman (2003) Ultramicroscopy; **94**; 225
  Roseman (2004) JSB; **145**; 91

FindEM

- So far:
  - Aim to integrate FindEM into a standard software environment that can be used across a future CCPEM software suite
  - Rewritten scripts from C-Shell (relatively old shell-script) to python (more favoured scripting language)
  - Improved GUI / added extra buttons
  - Written a Cmake file
- Still to do
  - Rewrite GUI in python
  - (Complete) cross-platform compatibility


FindEM particle finding - filter operation
Run code for particle set A :
0
Run code for particle set B :
0
Number of A's
0
Number of B's
0
Sampling (A/pixel)
12.5
scale factor
5
Diameter of particles A (A)
320
Diameter of particles B (A)
360
Scale factor for displayed circles
1.00
Peak width (A)
200
CCC threshold for A
0.40
CCC threshold for B
0.40
Max number of particles A
10000
Max number of particles B
10000
Update particle positions
clear
Write coordinates A
Write coordinates B
New image
Exit


FindEM particle finding - filter operation
Run code for particle set A :
3
Run code for particle set B :
4
Number of A's
70
Number of B's
98
Sampling (A/pixel)
12.5
scale factor
5
Diameter of particles A (A)
320
Diameter of particles B (A)
360
Scale factor for displayed circles
1.00
Peak width (A)
200
CCC threshold for A
0.40
CCC threshold for B
0.40
Max number of particles A
10000
Max number of particles B
10000
Update particle positions
clear
Write coordinates A
Write coordinates B
New image
Exit

PIMMS-EM

- New python library being developed by Maya Topf's group
- Functions for (flexible) fitting of atomic structures into EM maps of assemblies
- Uses a variety of scoring functions to determine goodness-of-fit
- Aims to automate fitting, to prevent problems associated with manual fitting bias
- Will provide functionality for regular users, and a toolkit for advanced users to build personalised functions.

PIMMS-EM

- To do:
  - Create Cmake / installation files
  - Create GUI
  - Test cross-platform compatibility
  - Produce example datasets & help files
  - Link with other parts of the EM pipeline within the CCPEM suite

- Will help users determine structures using the modelling approach with a small learning curve

Science & Technology Facilities Council

MRC Image Library

- Existed for ~30 years
- Used to determine structure of macromolecular assemblies
- Approach has been to study specimens with some form of symmetry: crystals / helices / icosahedral particles
- Currently ~80 individual programs of varying complexity
- Standard file formats are used (CCP4 format for data / postscript for images)

MRC Image Library

- CCPEM will maintain a copy of the MRC library and programs, in collaboration with Jude Short and Richard Henderson
- To do:
  - Fix incompatibilities (/bugs) between MRC / CCP4 formats
  - Preserve standards
- Done:
  - Cmake file created for Linux

EMX File Format

- Developed by a Working Group of EM Software developers, led by Jose-Maria Carazo & Steve Ludtke
- “Easy and unambiguous interchange of information”
- XML-based
- Currently only micrographs and 2D objects (i.e. particles) are represented
  - Future versions may have 3D object support
- Still finalising details
- We will help publicise it within the UK and support the community with using it

Science & Technology Facilities Council

Future projects…

- …we need user / community input to help drive our direction!
  - What software do you use?
  - What problems do you have with software?
  - What have you solved by yourself?
- Engage with other UK-based software developments such as RELION, JSubTomo, & ATAP
- Metadata management
- Validation functions (in collaboration with EBI)
- Training on other related software
- Facilitate access to HPCs?
- Webservices?

Science & Technology Facilities Council

Discussion!

Science & Technology
Facilities Council