Merging in 2D

Henning Stahlberg,
Biozentrum, Uni Basel, Switzerland
c-cina.org

2dx Workshop
Basel, August 23–26, 2016
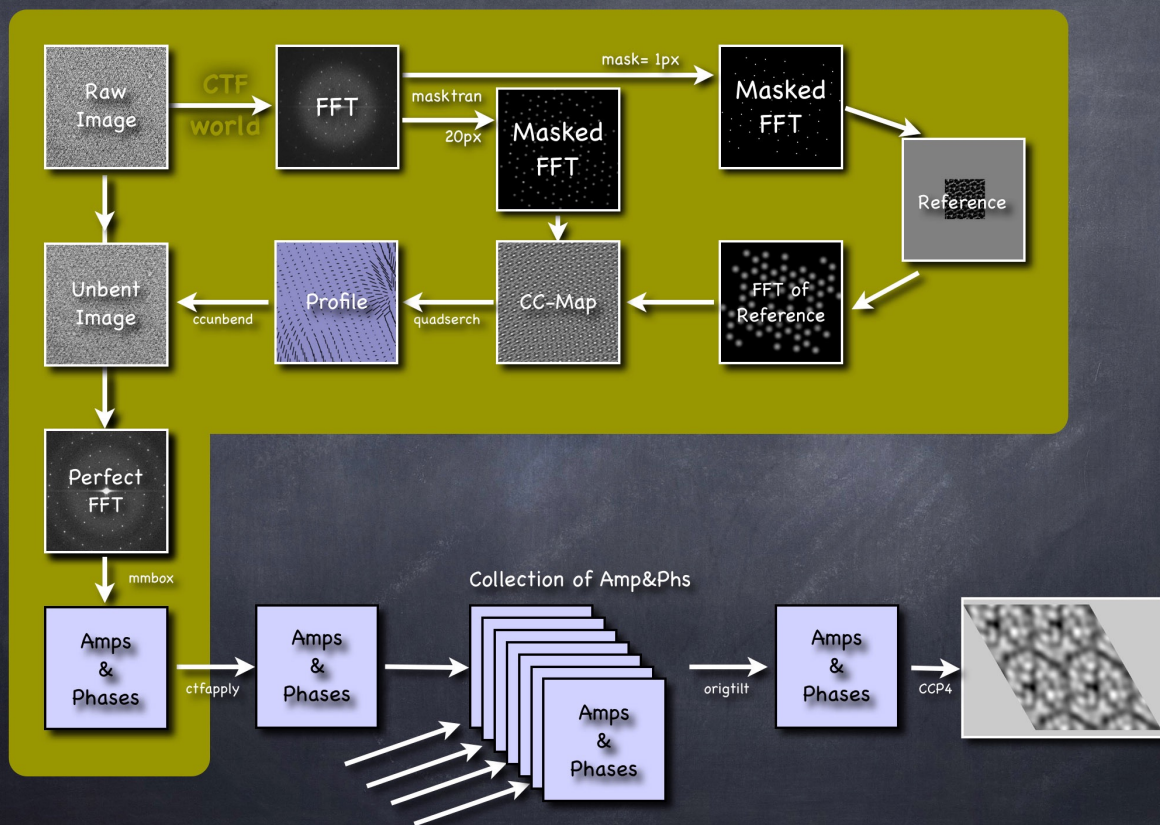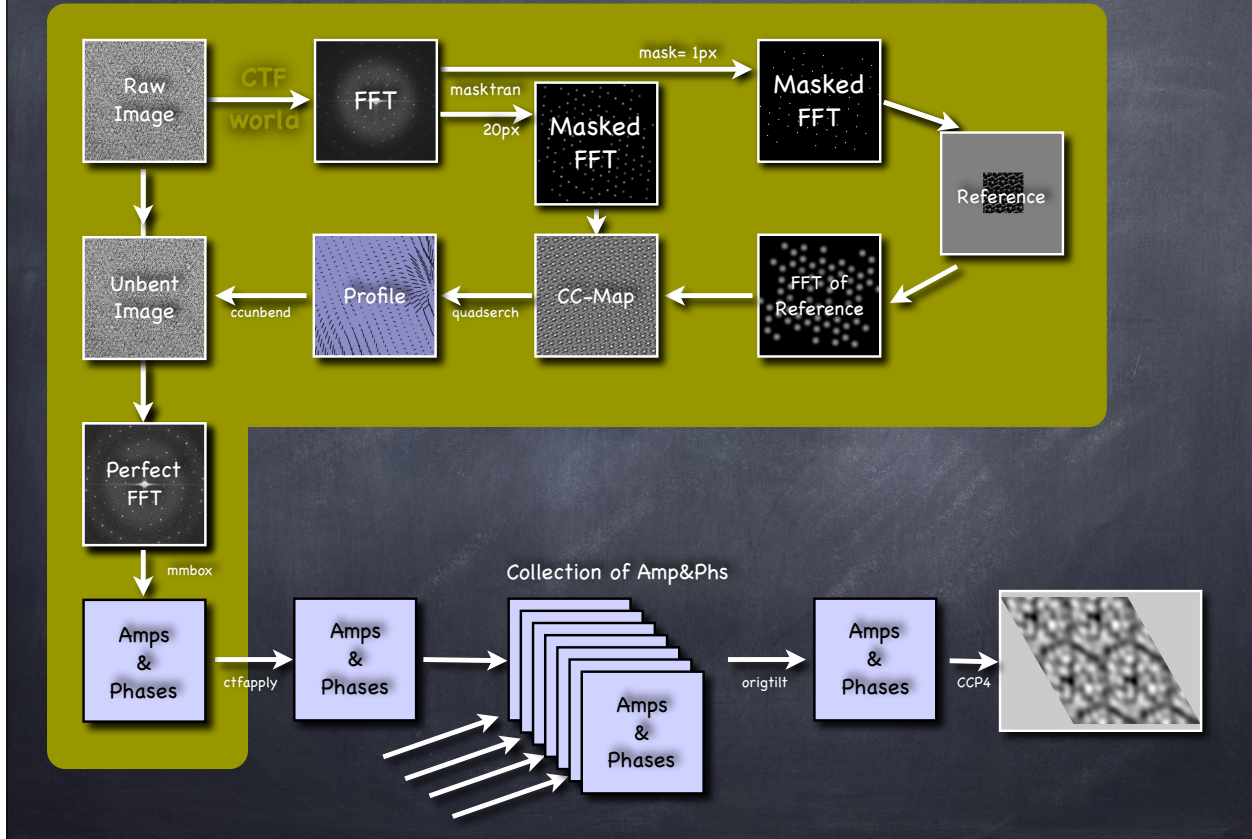
Algorithm Non-tilted

Algorithm Non-tilted

Algorithm for tilted or non-tilted data
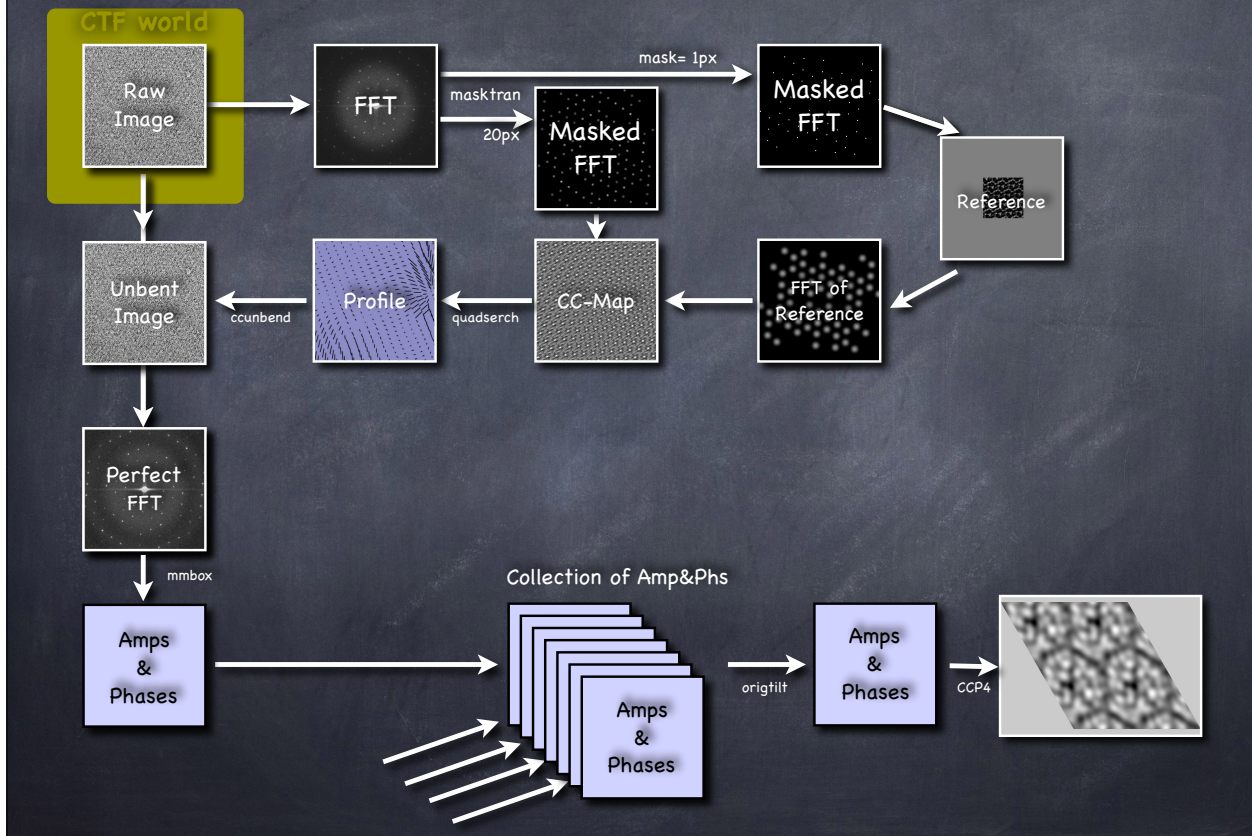
More Complete Unbending Algorithm
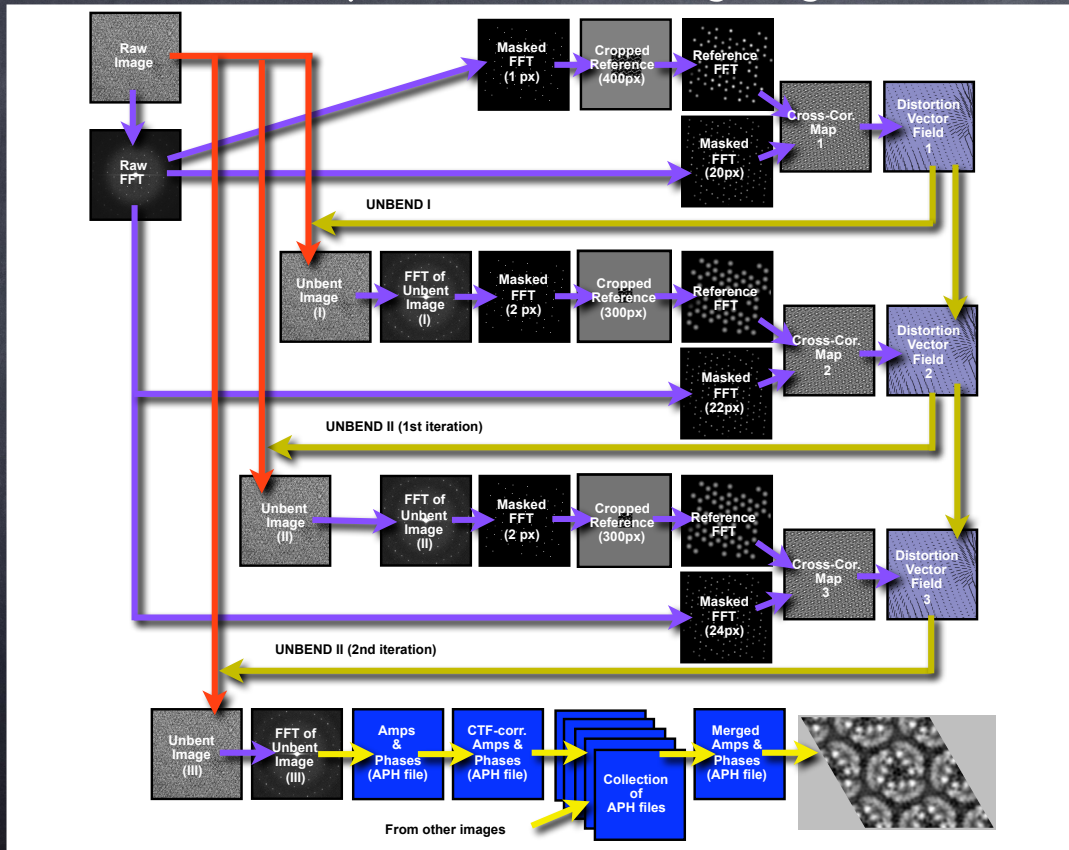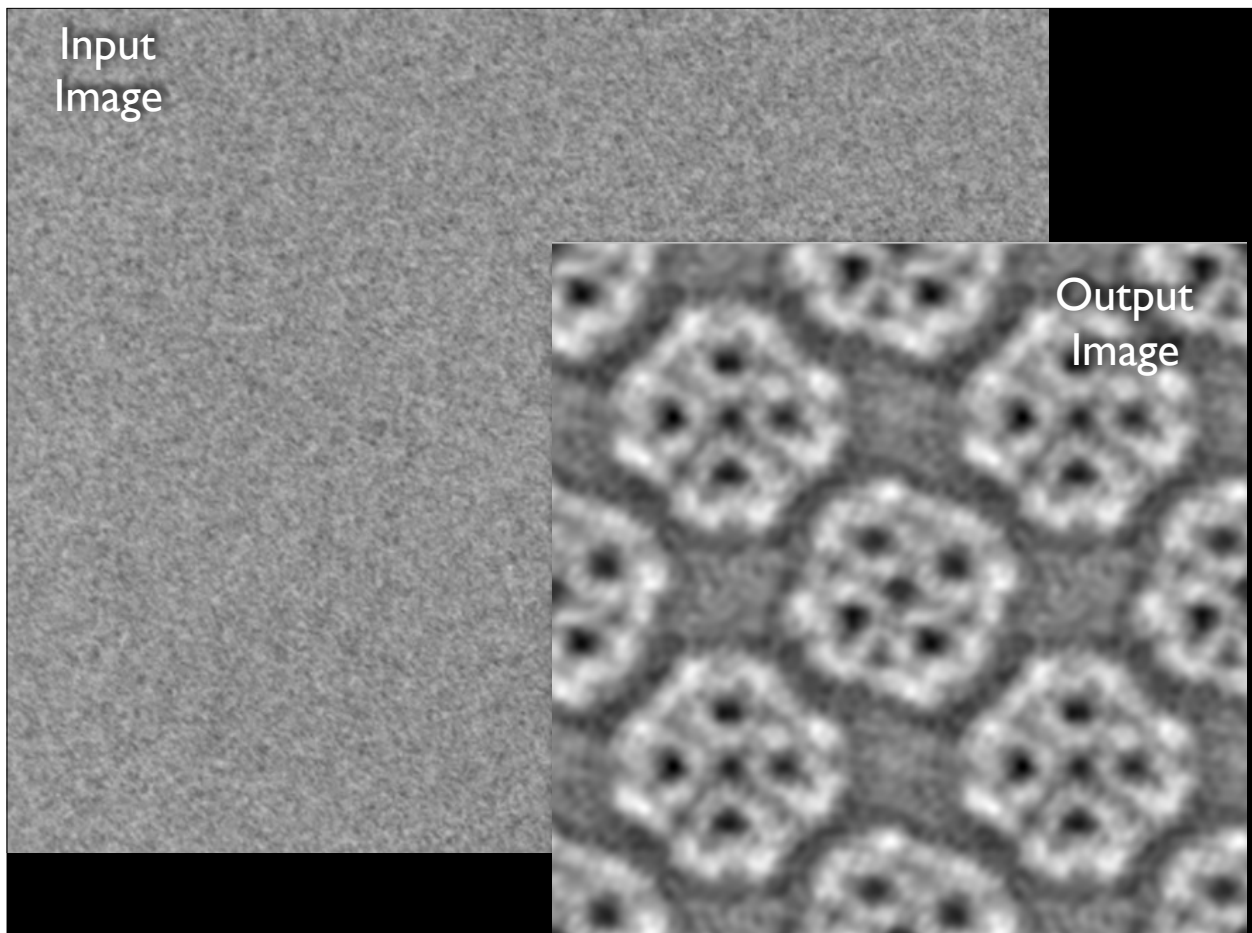
Algorithm Non-tilted
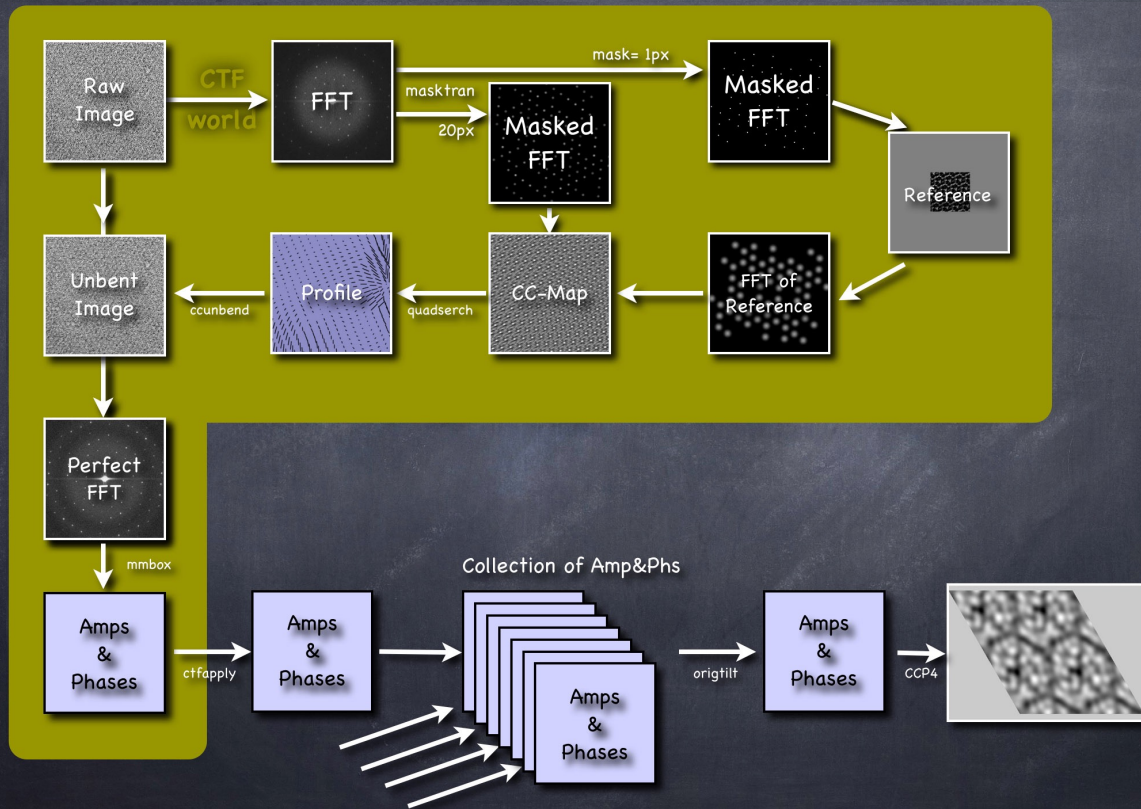
Project Management

- Launch *2dx_gui*

- Import only **ONE** image

- Open and process it,
  save database as project-default

- Import all other images into *2dx_gui*

- Process them all automatically **.....** (This will need finetuning...)

- Merge in 2D

- Merge in 3D

File Formats and Flow

mmbox => ctfapply => origtilt => avramphs => f2mtz

mmbox evaluates the Fourier transform of the unbent crystal image and produces a list of **Amplitudes**, **Phases**, and **Background Amplitudes**.

mmbox:

- takes FFTIR/cor${imagename}.fft.mrc
- generates APH/${imagename}.fou.nolimit.aph

```
Filename: APH/${imagename}.fou.nolimit.aph
1: Header
2:      H           K           AMP       PHS         IQ          BCK         CTF
Example:
    655201 gf06552, Unbend2, Mon Jun 16 00:33:23 CEST 2008
            0           1           0.0       60.9         9          72.2        0.0
            0           2         165.8      330.8         3          49.3        0.0
            0           3          17.1       52.3         8          39.3        0.0
            0           4         151.8      123.1         2          36.5        0.0
            0           5         411.8      127.5         2          50.9        0.0
```
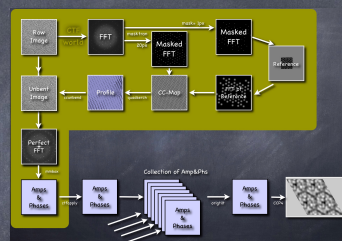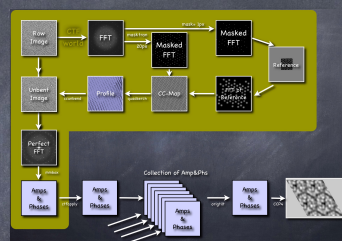

File Formats and Flow

mmbox => **ctfapply** => origtilt => avramphs => f2mtz

ctfapply calculates the CTF correction factor for each reflection, and produces a CTF-corrected APH file, where only the phases are potentially "flipped".
This is one file for each image.

ctfapply:

- takes ${imagename}.fou.nolimit.aph
- generates ${imagename}.fou.ctf.nolimit.aph

```
Filename: ${imagename}.fou.ctf.nolimit.aph
1: Header
2: H   K       AMP         PHS    IQ      BCK         CTF
                AMP is not changed
                PHS is already CTF-applied, possibly +180.
                                            CTF lists the contrast.
                                            Only phase (sign) was already applied.

Example:
    655201 gf06552, Thu Jul 3 22:49:19 PDT 2008
    0   1       0.0         240.9   9       72.2        -0.096
    0   2       165.8       150.8   3       49.3        -0.172
    0   3       17.1        232.3   8       39.3        -0.295
    0   4       151.8       303.1   2       36.5        -0.459
    0   5       411.8       307.5   2       50.9        -0.648
```
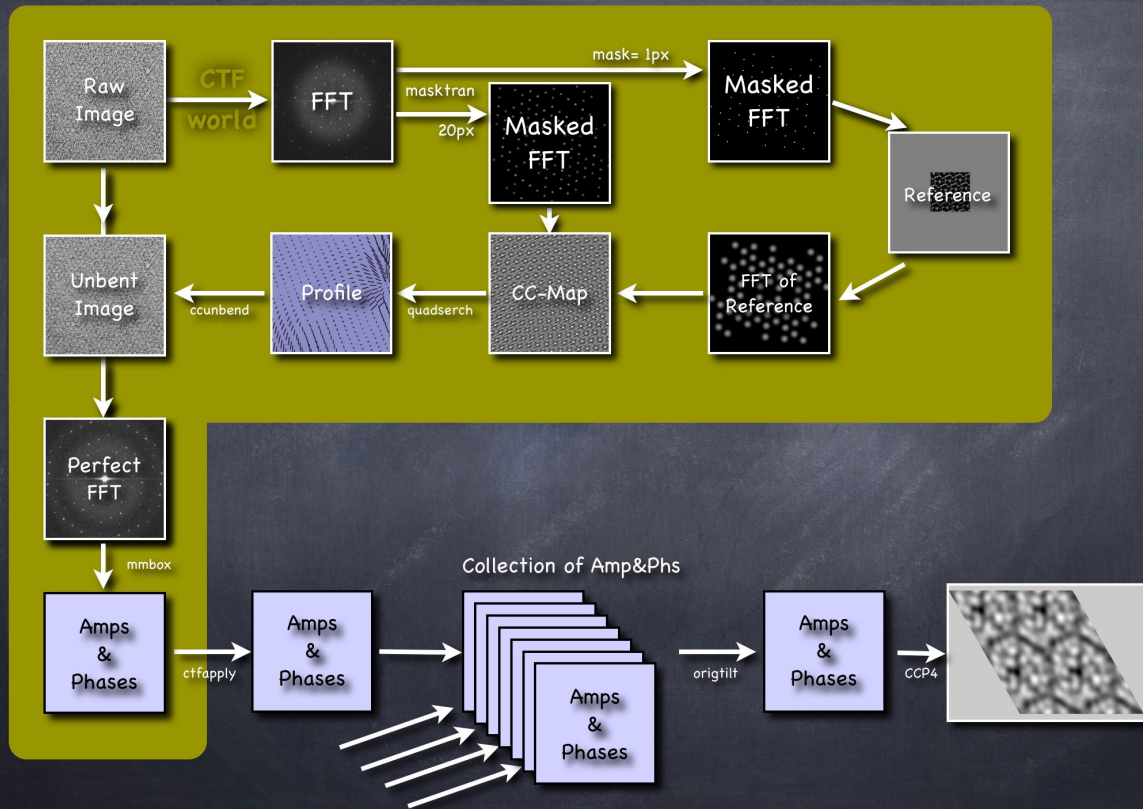
Algorithm Non-tilted

File Formats and Flow

mmbox => ctfapply => **origtilt** => avramphs => f2mtz

origtilt takes the several files from mmbox and combines them into one single large file merge.aph that contains AMP, PHS, original image number, and further data for each reflection.
This is one single file for the entire project.

origtilt:

- takes several APH/${imagename}.fou.cor.aph
- generates APH/merge.aph, which now has for each spot potentially many entries

```
Filename: APH/merge.aph
1: Header
2:   H    K    Z*         AMP          PHS           NO    IQ  WEIGHT        BACK         CTF
                          AMP is the raw measured value, not yet changed
                                      PHS is CTF phase flipped, and has also the PhaseOrigin applied to it
Example:
     1001
     1    0   -0.0000    70.8670      -88.1100       65551 -6  0.078805      47.8734      -0.083
     1    0    0.0000    112.300      -21.6200        6543  6  0.192624      74.5000      -0.093
     1    0    0.0000    25.0594      -131.430       65551  8  0.027867      49.3105      -0.088
...
     2    0   -0.0000    120.801      -160.000        6557  4  0.192729      48.7613      -0.165
     2    0   -0.0000    156.540      169.480       665202 -2  0.276022      34.6150      -0.154
     2    0   -0.0000    202.253      170.080       655201 -3  0.375352      51.5399      -0.167
...
     2    0    0.0013    411.287      -178.300        7032  1  0.440457      50.2215      -0.383
     2    0    0.0013    313.198      -178.300        8032  2  0.307053      40.1020      -0.385
     2    0    0.0016    298.834      -161.900        5685  2  0.360523      61.4043      -0.198
```
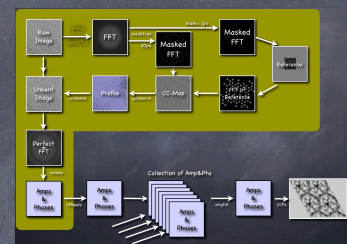
File Formats and Flow

mmbox => ctfapply => origtilt => **avramphs** => f2mtz

avramphs takes the merge.aph file that contains for each reflection several measurements, and averages these into one AMPlitude and PHAse value for each reflection.
This is one single file for the entire project.
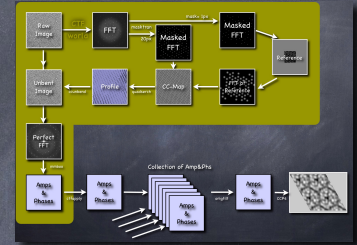
avramphs:

- uses zminmax = –0.025...0.025, to make sure only non–tilted data participate.
- takes APH/merge.aph
- generates APH/avrg2D.hkl

```
Filename: APH/avrg2D.hkl
1: merging number
2:   H    K    L       AMP           PHS             FOM
                 AMP is an average of several spots, and the values are now amplitude-CTF corrected (without inverting the sign).
                             PHS is an average of several spots, which are already CTF and phaseorigin corrected
                                         FOM is between 0 and 100 and quantifies the reliability of the spots.

Example:
    1001
    1    0    0      196.833       -64.0629        96.9797
    1    1    0     1018.19        -6.10076        99.9556
    1    2    0      162.905      -177.142         98.7745
    1    3    0      262.806      -179.891         99.5678
...
    2    0    0      453.627      -177.303         99.8751
    2    1    0      109.832         4.30570       94.4169
    2    2    0     1278.94        177.975         99.9585
```
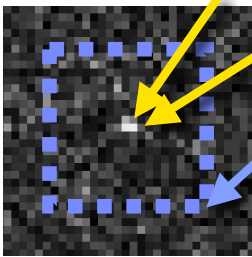

How to average Amplitudes

FT Readout:

H, K, Z*, **Amp**litude, **Pha**se, **Back**groundAmplitude

For each reflection H,K,z* :

$$SNR = Amp \,/\, Back$$

$$FOM = \frac{I_1(SNR)}{I_0(SNR)}$$

Simple Approach:

$$AvgAmp = \frac{\sum Amp}{n}$$

Better:

$$AvgAmp = \frac{\sum Amp * FOM}{\sum FOM}$$

In 2dx now:

$$AvgAmp = \frac{\sum \dfrac{Amp * CTF}{Back^2}}{\sum \left(\dfrac{CTF^2}{Back^2} + \dfrac{1}{Amp^2} \right)}$$

How to average Phases

R. Henderson et al. / Structure of purple membrane from Halobacterium halobium   163

(10, 3) 4.7Å

(6, 8) 4.4Å

(7, 7) 4.5Å

(1,12) 4.3Å
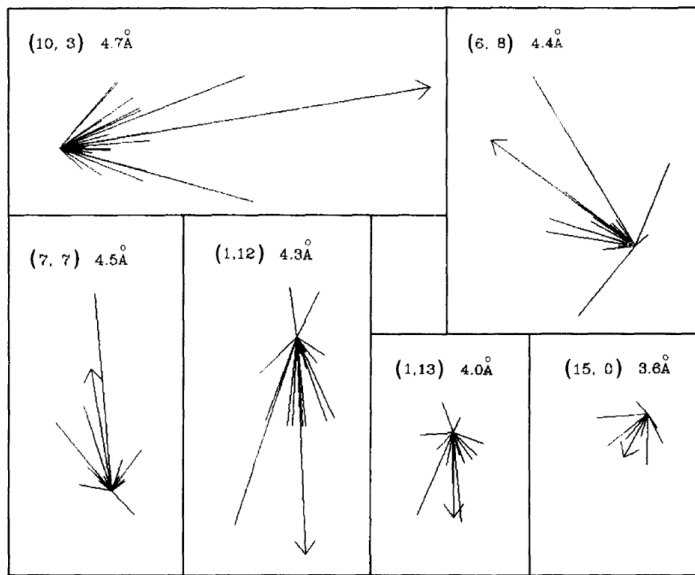
(1,13) 4.0Å

(15, 0) 3.6Å

Fig. 7. Graphical comparison of all phases determined for six spots with resolution beyond 4.7 Å. Phases are plotted as vectors on a polar diagram with the length of each vector being proportional to $1/IQ^2$. Thus the strongest spots show up as longer vectors. The result of the summation of all the vectors is also shown as a vector, but this is plotted at $1/4$ of its true length. The direction of this vector is our best estimate of the phase of the Fourier component in the structure, and its length can be used to provide an estimate of the error.

$$AvgPhs = \tan^{-1}\left(\frac{\sum \cos(Pha) * SNR(FOM)}{\sum \sin(Pha) * SNR(FOM)}\right)$$

Averaging of data from different images in the MRC/2dx software is done in Fourier space by dealing with each reflection individually:

Amplitudes and Phases are averaged, taking $1/IQ^2$ (or now FOM) as weight.
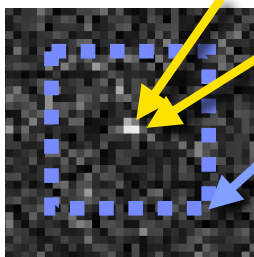
On the left, the Phase averaging is shown graphically.

How to average Figures of Merit (FOM)

FT Readout:

G. Sim, Acta Cryst. 13, 511 (1960)
D. Blow & F. Crick, Acta Cryst. 12, 795 (1959)
G. Sim, Acta Cryst. 12, 813 (1959)

H, K, Z*, *Amp*litude, *Pha*se, *Back*groundAmplitude

For each reflection H,K,z* :

$$SNR = Amp / Back$$

$$FOM = \frac{I_1(SNR)}{I_0(SNR)}$$

$$FOM = lut(SNR)$$

$lut(IQ) = 0.990, 0.982, 0.939, 0.870, 0.763, 0.630, 0.505, 0.124$

$$FOM = \exp\left(-\frac{\sigma_{Phase}^2}{2}\right) \qquad FOM = \cos(\max(\sigma_{Phase}, 90))$$

Use Lookup Table to switch between FOM and *XARG*

$$AvgFOM = \frac{I_1(AvgSNR)}{I_0(AvgSNR)} \qquad I = \text{modified Bessel Function of the first kind}$$

$$AvgSNR = \sqrt{\left(\sum \cos(Pha) * SNR(FOM)\right)^2 + \left(\sum \sin(Pha) * SNR(FOM)\right)^2}$$

```
1: Header
2:  H     K    Z*        AMP         PHS           NO   IQ  WEIGHT       BACK       CTF
                         AMP is the raw measured value, not yet changed
                                PHS is CTF phase flipped, and has also the PhaseOrigin applied to it
Example:
   1001
      1   0  -0.0000    70.8670     -88.1100      65551 -6  0.078805    47.8734    -0.083
      1   0   0.0000    112.300     -21.6200       6543  6  0.192624    74.5000    -0.093
      1   0   0.0000    25.0594    -131.430       65551  8  0.027867    49.3105    -0.088
...
      2   0  -0.0000    120.801    -160.000        6557  4  0.192729    48.7613    -0.165
      2   0  -0.0000    156.540     169.480      665202 -2  0.276022    34.6150    -0.154
      2   0  -0.0000    202.253     170.080      655201 -3  0.375352    51.5399    -0.167
...
      2   0   0.0013    411.287    -178.300        7032  1  0.440457    50.2215    -0.383
      2   0   0.0013    313.198    -178.300        8032  2  0.307053    40.1020    -0.385
      2   0   0.0016    298.834    -161.900        5685  2  0.360523    61.4043    -0.198
```

avramphs:

```
Filename: APH/avrg2D.hkl
1: merging number
2:  H     K     L        AMP         PHS           FOM
                         AMP is an average of several spots, and the values are now amplitude-CTF corrected (without inverting the sign).
                                PHS is an average of several spots, which are already CTF and phaseorigin corrected
                                       FOM is between 0 and 100 and quantifies the reliability of the spots.

Example:
   1001
      1   0   0    196.833     -64.0629       96.9797
      1   1   0    1018.19      -6.10076      99.9556
      1   2   0    162.905    -177.142        98.7745
      1   3   0    262.806    -179.891        99.5678
...
      2   0   0    453.627    -177.303        99.8751
      2   1   0    109.832       4.30570      94.4169
      2   2   0    1278.94     177.975        99.9585
```
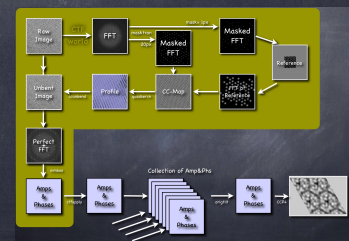

File Formats and Flow

mmbox => ctfapply => origtilt => avramphs => **f2mtz**

centric and **hklsym** apply boundary conditions to reflection values. Centric enforces for certain symmetries that reflections of non-tilted crystals must have phases of either 0º or 180º.
Hklsym avearges symmetry-related reflections.

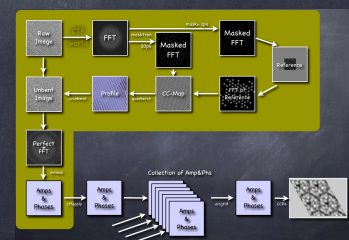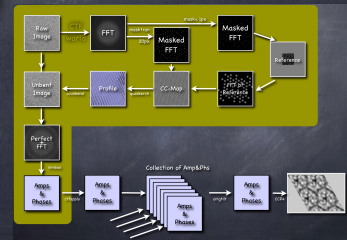**f2mtz** transforms the AMP&PHS file into a binary MTZ file.

f2mtz:

- takes APH/sym2D.hkl
- generates merge2D.mtz

```
Filename: merge2D.mtz
SYMMETRY ${CCP4_SYM}
LABOUT H K L F PHI FOM
CTYPOUT H H H F P W
```

File Formats and Flow

mmbox => ctfapply => origtilt => avramphs => **f2mtz**

centric and **hklsym** apply boundary conditions in reflection values. Centric reflections for crystals symmetries that render non-tilted crystals must have phases of either 0º or 180º. Hklsym avearges symmetry-related reflections.

f2mtz transforms the AMP&PHS file into a binary MTZ file.

f2mtz:
- takes APH/sym2D.hkl
- generates merge2D.mtz

Filename: merge2D.mtz
SYMMETRY ${CCP4_SYM}
LABOUT H K L F PHI FOM
CTYPOUT H H H F P W

File Workflow is described on: 2dx.org

2D Merging

- Select best image only

- Run "Merge Selected Images" to create reference

- Select all images with low sample tilt (< 7 deg)

- Run "Refine PhaseOrigins Once" to align against reference

- Run "Merge & Refine" 5 times
  to refine PhaseOrigins

- Run "Final Merge", look at final map.
- Run "Quality Evaluation"
- Run "Generate Image Maps", check phase origins of images.

Final Merge in 2D

Final Merge in 2D

Slide showing the 2dx (4.0.0) application "MloK1-cAMP" with the Final Merge panel.

Left navigation (2D / 3D):
- Initialization
- Merge Selected Images
- Refine PhaseOrigins Once
- Merge & Refine (Iterative)
- Final Merge
- Quality Evaluation
- Generate Image Maps

Final Merge

Setup — Level: Simplified

Search Parameters

Common Image Processing

- Lower Resolution Limit (RESMIN): 90.0
- Upper Resolution Limit (RESMAX, in Angstroems): 4.5
- Merging resolution limit
- 0=Use global resolution limit
- Resolution of the merged dataset for reference: 4.5

Output (Double click for logbrowser) — Verbosity Level: Low

```
 8              0    0   10   58  139  142  143  744  1236

 9   5.0  4.7   0.0  5.8 65.2 61.2 63.5 69.7 78.5 75.7  73.4
 9              0    3   11   85  103  138  144  953  1437

10   4.7  4.5   0.0  0.0 82.5 61.0 69.0 74.4 76.4 79.8  76.8
10              0    0   10   69   98  124  138  735  1174
-----------------------------------------------------------
              9.2 13.5 20.7 33.4 44.1 51.8 61.9 69.5
             2023 4337 2964 2513 2385 2166 1881 9157
-----------------------------------------------------------
 Overall:     Phaseresidual = 43.489        Number of spots =  27426

The last column shows the average phase residual in that resolution range,
which is the mean weighted squared distance of the phase values from the averaged v
(90 deg = random)

==== fomstats - to calculate FOM statistics ============================

FOMSTATS output:

 N   RESOLUTION   NDAT   PHASERR(0/180, 45 is random)
 1    14.3        42     21.9
 2    10.1        38     13.0
 3     8.2        38     19.0
 4     7.1        36     16.8
 5     6.4        40     22.8
 6     5.8        36     25.9
 7     5.4        38     25.3
 8     5.0        33     34.2
 9     4.7        41     34.6
10     4.5        33     31.3
THE OVERALL ERROR FOR 375 REFL. WAS:  24.3 DEG

The last column shows the average distance of the averaged phase values
from the symmetry-constrained target values of 0 or 180 degrees.
(45 deg = random)

==== 2dx_centric2 - to correct phases to 0 or 180 for 2D run ===============
==== 2dx_hklsym4 - to apply symmetry to APH file for 2D run ===============
==== 2dx_plotreska - to plot the powerspectrum with resolution circles ==========
==== f2mtz - to transform APH file into MTZ file for 2D run ===============
==== 2dx_processor.exe - to create a MAP file from HKL ===============
==== 2dx_merge normal end. ===============
#######################2dx_finalmerge2D finished.#######################
```

DEPENDENT SCRIPTS
- 2dx_origtilt_merge.com
- 2dx_origtilt_aph_to_mtz.com
- 2dx_sym2spcgrp_sub.com
- 2dx_merge_makedirs
- initialize

Results

Parameter	Value
JREFL	54832
JREFL_2D	27426
initialization_executable	y
initialization_reset	y
overall_phase_residual_2D	43.489

Images
- CSH: First (01) merging script
- LOG: First (01) origtilt A output
- CSH: Last (12) merging script
- LOG: Last (12) origtilt A output
- LOG: Console output from merging
- APH: merge.aph
- merge.aph [H,K,Z,A,P,#,IQ,W,Bk,CTF]
- LOG: 2dx_avrgamphs output for 2D run
- TXT: Phase residual table
- APH: FOMSTATS HKL output
- TXT: FOMSTATS statistics
- APH after hklsym4 [H,K,L,F,P,FOM,1.0]
- PS: Resolution Circle Plot of non-tilted data
- MTZ: Merged full reciproc. space 2D data
- MTZ: Merged full reciproc. space 2D data (copy for ref)
- MRC: FINAL p4212 symmetrized 2D map

Quality Evaluation in 2D

Slide showing the 2dx (4.0.0) application "MloK1-cAMP" with the Quality Evaluation panel.

Left navigation (2D / 3D):
- Initialization
- Merge Selected Images
- Refine PhaseOrigins Once
- Merge & Refine (Iterative)
- Final Merge
- Quality Evaluation
- Generate Image Maps

Quality Evaluation

Setup — Level: Simplified

Search Parameters

- Lower Resolution Limit (RESMIN): 90.0
- Upper Resolution Limit (RESMAX, in Angstroems): 4.5
- Resolution of the merged dataset for reference: 4.5
- ALAT (Z-dimension of unit cell to reconstruct): 200

Output (Double click for logbrowser) — Verbosity Level: Low

```
#######################2dx_QualityEvaluation2D#######################
==== Sourcing sym2spsgrp_sub.com =====================================
==== get_max_defocus.py - to extract defocus and TANGL maxima. ==============

=====================================================================
2D Reconstruction Parameters
=====================================================================
Crystal plane group symmetry:                              p4212
Crystal unit cell parameters:                    a=135.0A, b=135.0A
                                                     gamma=90.0deg
---------------------------------------------------------------------
Number of images:                                           35
Range of defocus [micrometers]:                  -0.750 ... 1.937
Number of observed reflections:                          27426
IQ range used:                                         1 ... 8
---------------------------------------------------------------------
Resolution cut-off [Angstroems]:                           4.5
---------------------------------------------------------------------
Overall phase residual [deg]:                             43.5
---------------------------------------------------------------------
 19 images in MloK1-II
  5 images in MloK1-K2-2014-05-26-MloK1 cAMP
  2 images in MloK1-K2-2014-07-29-MloK1 camp
  2 images in MloK1-K2-2014-09-08 MloK1 cAMP
  5 images in MloK1-K2-2014-09-16 MloK1 cAMP
  1 images in MloK1-K2-2014-11-17 MloK1 cAMP
  1 images in automatic
---------------------------------------------------------------------


              PHASE RESIDUALS IN RESOLUTION RANGES

 #  DMIN DMAX  IQ= 1     2     3     4     5     6     7     8 all IQs

 1 1000.0 14.2    9.0  13.6  20.0  29.3  36.8  43.3  52.3  63.7   23.5
 1             1667  2727  1402   647   405   304   829   8224

 2  14.2 10.1   10.1  13.3  18.3  25.5  32.7  38.2  46.4  53.7   26.7
 2              332  1257   806   524   396   295   200   749   4559

 3  10.1  8.2    8.5  14.0  21.4  29.3  39.3  43.1  54.9  59.7   40.1
 3               24   313   532   531   388   318   248  1033   3387

 4   8.2  7.1    0.0  11.4  21.7  30.3  45.1  48.8  57.5  66.6   53.4
 4                0    36   108   263   298   268   195  1053   2221

 5   7.1  6.4    0.0  22.1  34.8  41.5  43.4  56.9  60.1  72.0   62.3
```

DEPENDENT SCRIPTS
- 2dx_sym2spcgrp_sub.com
- 2dx_merge_makedirs
- initialize
- get_max_defocus.py

Results

Parameter	Value
initialization_executable	y
initialization_reset	y

Images
- TXT: Summary_Statistics_2D
- TXT: Phase residual table
- PS: Resolution Circle Plot of non-tilted data