Age Concern

Mining established software for unpublished wisdom

Taken for granted (Users, Developers)

- Program feature that works well and reasonably well documented to users e.g. constrained Hydrogens in ShelXL
- Detailed algorithm (math, pseudo-code)
 - o non-issue for the user
 - useful only to write a similar program: but why doing so?

The Monolith Syndrome (Users)

- © Can I combine different features from different programs?
- At best difficult, at worst impossible
 - free restraints from Crystals and ShelX constraints: NO
 - ShelX refinement with masked solvent: FUDGE

The Monolith Syndrome (Users > Developers)

- Dear developer, can you combine those 2 programs?
 - NO
 - Scripts emulating the chores the user has to go through
- Not good enough

The Monolith Syndrome (Developers)

- @ Could we extend them?
- Could we cleave those monolith into reusable bricks?
- Not realistically: The Great FORTRAN Sins

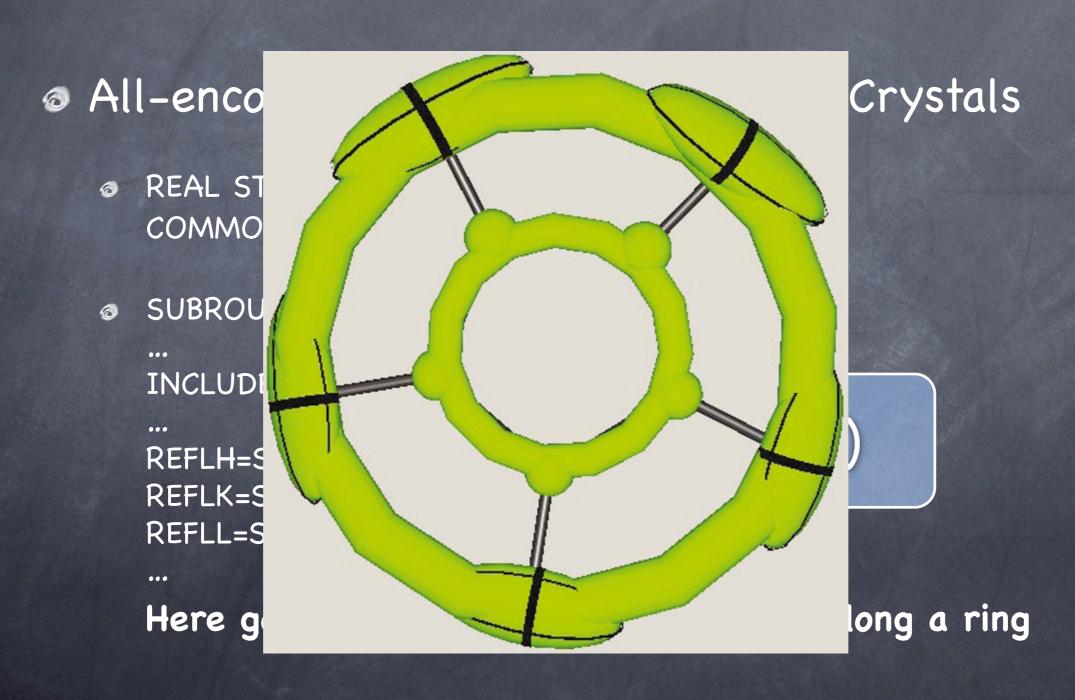
- All-encompassing COMMON block: Crystals
 - REAL STORE COMMON/XDATA/STORE(16777216)
 - SUBROUTINE XRING (..., M6RI, ...)

```
...
INCLUDE 'STORE.INC'
```

REFLH=STORE(M6RI)
REFLK=STORE(M6RI+1)
REFLL=STORE(M6RI+2)

Fcalc(hkl)

Here goes the computation of disorder along a ring



- All-encompassing COMMON block: Crystals
 - REAL STORE COMMON/XDATA/STORE(16777216)
 - SUBROUTINE XRING (..., M6RI, ...)

```
...
INCLUDE 'STORE.INC'
```

REFLH=STORE(M6RI)
REFLK=STORE(M6RI+1)
REFLL=STORE(M6RI+2)

Fcalc(hkl)

Here goes the computation of disorder along a ring

- All-encompassing COMMON block: Crystals
 - REAL STORE COMMON/XDATA/STORE(16777216)
 - SUBROUTINE XRING (..., M6RI,).)

```
INCLUDE 'STORE.INC'
```

```
REFLH=STOR E(M6RI)
REFLK=STOR E(M6RI+1)
REFLL=STOR E(M6RI+2)
```

Fcalc(hkl)

Here goes the computation of disorder along a ring

- Global state shared by every function
 - ShelX: two arrays A and B store unit cell, sites, ADP's, etc
 - every routine: CALL SX3G(..., A, B)
 - ø direct access to A and B elements

M=INT(ABS(C(NG)))
IF(M.EQ.N)GOTO 60
51 M=M+MD
... 50 lines later ...
60 KU=KX
IF(NS.NE.0)GOTO 47

several 100 lines

- Arrays = most advanced data structure
- Compare:
 - X=STORE(M5A)
 Y=STORE(M5A+1)
 Z=STORE(M5A+2)
 X1=X*STORE(M2LI) + Y*STORE(M2LI+1) + Z*STORE(M2LI+2)
 - sc = scatterers[m5a]
 r = symmetries[m2li]
 site_1 = r*sc.site

Refinement FORTRAN Direct badness Method Merging

The Redemption

- Toolbox approach
- C++ / Python
- Mine algorithm from old software
 - then recode into C++ / Python
 - help from fable

Fable (FORTRAN to C++)

- Produce C++ code one can further develop on:
 - programmers with both FORTRAN and C++ backgrounds
- COMMON blocks encapsulated in classes:
 - several instances of the converted FORTRAN inside the same C++ program
 - static memory allocation > dynamic

Mining Bag of tricks

- Talk to the author
- Static analysis of code structure
 - Look for comments in the code
 - © Crystals: 25% vs ShelXL: 3.1%
 - LAPACK: 31.5%
 - Search for WRITE

Mining Bag of tricks

Talk to the author

Static analysis of code structure

© Crystals: 25%

LAPACK: 31.5%

Search for WRIT

Look for commen 1 FORMAT(' some header:') 2 FORMAT(2F8)

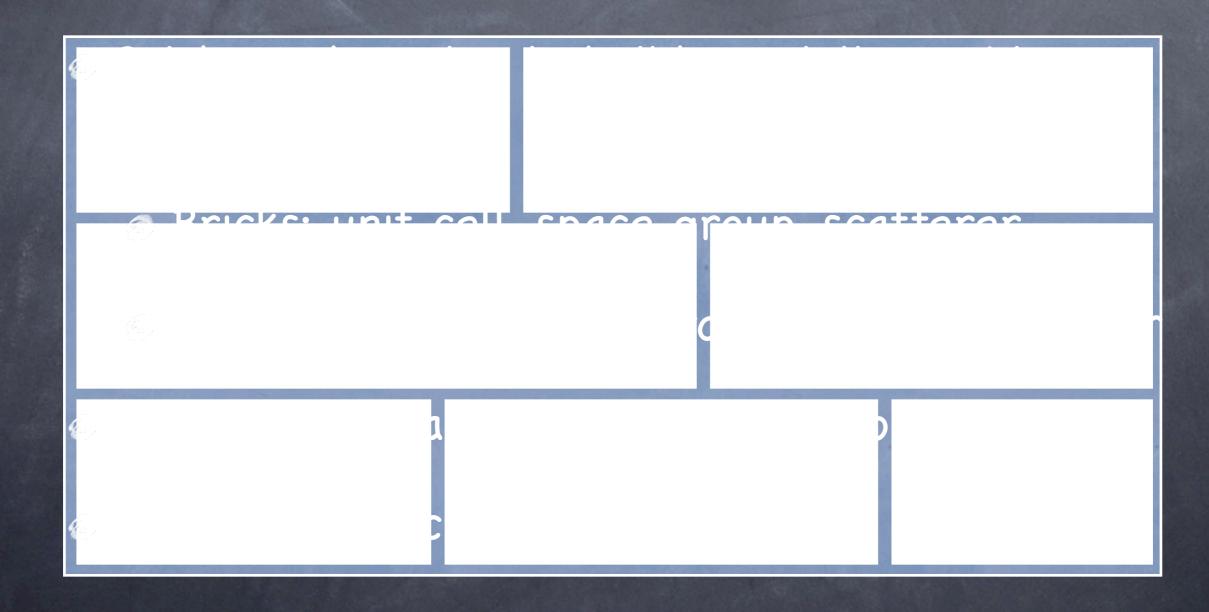
... 50 lines later ...

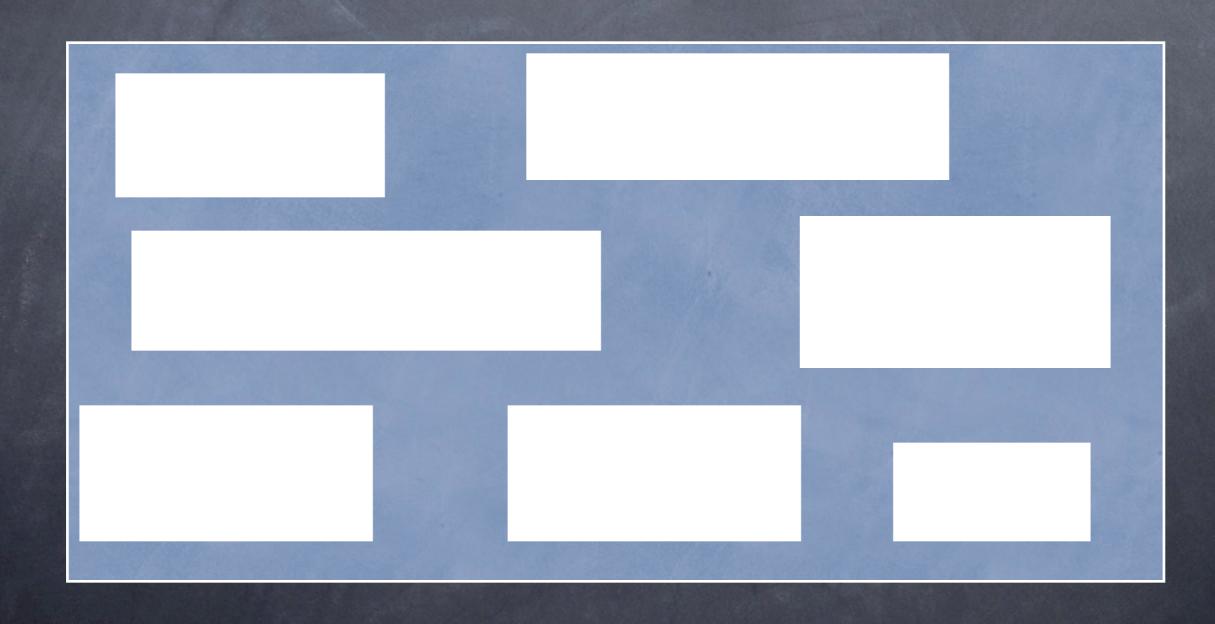
WRITE(OUT, 1) DO I=1,10 WRITE(OUT, 2) A, B **ENDDO**

Mining Bag of tricks

- Dynamic analysis of code structure
 - Run the program step by step through a debugger
 - and/or add WRITE statements
- Go back to previous slide

- Bricks and mortar to build crystallographic software, small or big
 - Bricks: unit cell, space group, scatterer, ...
 - Mortar: infrastructure to bind them together
- Each brick: fearsome quality control (tests)
- Thin mortar: scripting language





- Bricks and mortar to build crystallographic software, small or big
 - Bricks: unit cell, space group, scatterer, ...
 - Mortar: infrastructure to bind them together
- Each brick: fearsome quality control (tests)
- Thin mortar: scripting language

The Dream Comes True: CCTBX

- Initially not very well suited for small molecule work => SMTBX
- full matrix L.S. => esd's and correlations
- weighting schemes
- restraints (bonds, angles, dihedral angles)

The Dream Comes True: SMTBX

- constraints:
 - @ all ShelX
 - reasonably easy to add new ones, e.g.
 - terminal sp3 —NH2
 - water molecules (rigid)

The Dream Comes True: SMTBX

- Squeeze done right: refinement of
 - | F_{calc} + F_{masked solvent}|² vs observed F²
- Like in Crystals
- Work done by our PhD student Richard Gildea

Acknowledgement

- Oleg Dolomanov, Richard Gildea Horst Puschmann, Judith Howard (Durham University, UK)
- David Watkin, Georges Sheldrick
- Ralf Grosse-Kunstleve and PHENIX people
- EPSRC and Bruker AXS