INTRODUCTORY TALK

Ton Spek
National Single Crystal Service Facility
Utrecht University

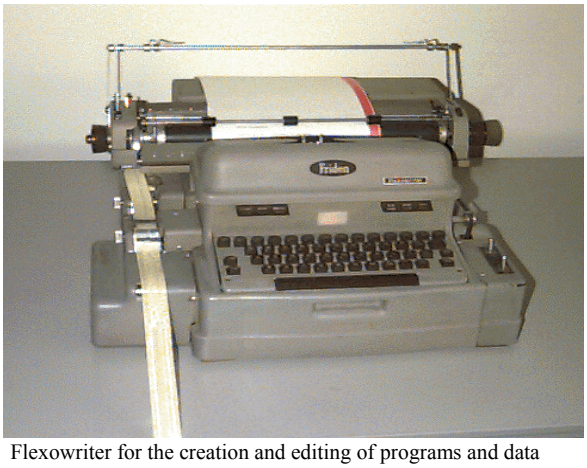
Some History

- On crystallographic computing from the perspective of a small-molecule crystallographer who started to work in crystallography in the mid 60's at Utrecht University, The Netherlands.
- Many of the older software developers, like me, have a background in Direct Methods. Mine started as follows:
- As a student, I was given a colorless crystal of unknown composition with the assignment to determine its structure using X-ray techniques only. It took me more than ½ a year to determine that it was *methoxyglutaconic acid.*
- Today, 40 years later, a problem like this is solved in a matter of seconds on my notebook, but not in those days.

Not so 40 Years ago

- The crystallography group in Utrecht already had a tradition in Direct Methods (Paul Beurskens, one of the first authors implementing the *Symbolic Addition Method*). However, none of the locally available programs gave an interpretable hand-contoured map.
- So I ended up with developing my own *Symbolic Addition* program, AUDICE, for centro-symmetric structures.
- AUDICE was locally rather successful since it also solved all other notoriously 'unsolvable structures' hanging around in the lab.
- Major calculations and program testing were done once a week during the '*nightshift*', not '*overnight*', on the Utrecht University mainframe. (A major social event in those days in view of the presence of most group members between 6PM and 8 AM the next morning)

~1966, Electrologica X8 ALGOL60 'Mainframe' (<1MHz)


Flexowriter for the creation and editing of programs and data

Times and Mainframe Changed

- MULTAN (FORTRAN/PUNCHCARD) came and replaced my Direct Methods program AUDICE (ALGOL60/Papertape) in the early 70's, when the single user university computer was replaced by a real multiuser mainframe (CDC6400).
- MULTAN was superseded in the 80's by the even more powerful SHELXS, SIR & DIRDIF software.
- No big improvements in small molecule DM since then ?
- In the 90's S&B, SHELXD entered the field, coming down from Macro-crystallography.

Direct Methods Meetings

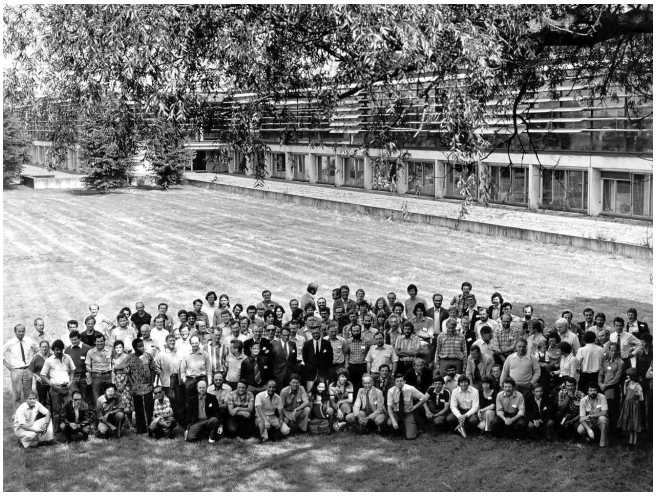
- Many past meetings and schools were organized with Direct Methods (software and theory) as a major subject.
- Important one's were the CECAM workshops on Direct Methods (5 weeks!, bringing together people working in the field to work on current issues) in the early 70's in Orsay around a big IBM-360 with lectures by Hauptman.
- Launch of MULTAN, many personal contacts – Viterbo
- NATO schools on Direct Methods in Parma and York in the 70's.
- Direct Methods schools in Erice in 1974 & 1978.
- Photo of the participants of the 1978 Erice School next :

Direct Methods Now

- Direct Methods appear to be currently no longer a major topic at meetings.
- Some years ago there was a morning lecture by Herbert Hauptman with the message that the tangent formula was what really mattered. That afternoon there was as lecture by Carmello Giacovazzo with the message that there was no need for the tangent formula…
- George Sheldrick will give us his perspective on '*The Future of Direct Methods'* at the end of this meeting.

IUCr Computing Schools

- ….. Mostly held jointly with IUCr Assemblies – Examples
- 1963 – Rollett, Algorithms (black book)
- 1969 - Least-Squares & Absorption Correction (SHELX76 - code)
- 1978 - Program systems (SHELX, XTAL, NRCVAX etc.)
- 1996 - Macro-crystallography
- 1999 - Macro-crystallography
- 2002 – (None)
- 2005 - Siena (again: Small, Macro, Powder)
- Photo 1978 school in Enschede (Netherlands)



Motivations for this Crystallographic Computing School

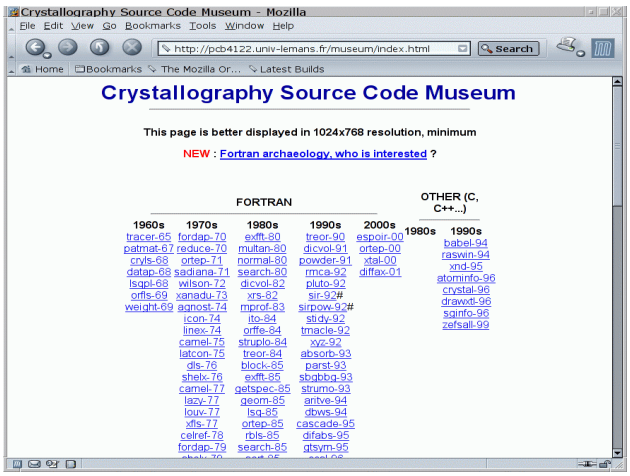
- A general feeling within at least the small-molecule community:
  'The current generation of software developers is phasing out, where is the new generation to keep things running in the future'
- There exists a growing community of push-button users
  (What is not behind a button can not be done…)
- Major funding and software development is currently in macro crystallography
  (possible useful spin-off to the small-molecule world)
- It is sensed that things well known in the small-molecule world are reinvented in the macro world and presented as new.
- Black Box and Proprietary Software as opposed to Open Source.
  (lack of info about the algorithms used and options to modify)

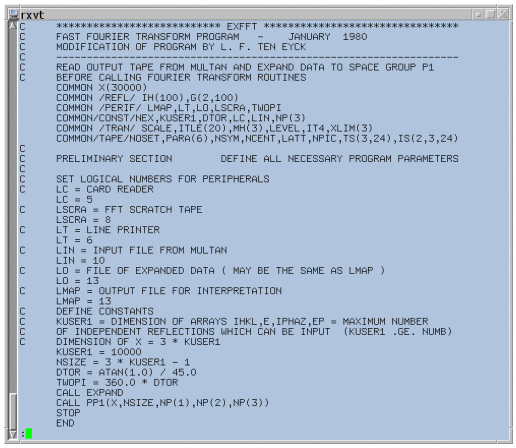
Hardware Platforms

- MS-Windows:
  - Small-Molecule Crystallography
  - Powder crystallography
- UNIX/LINUX/(OSX):
  - Macro Crystallography
  - (Small-Molecule Crystallography)

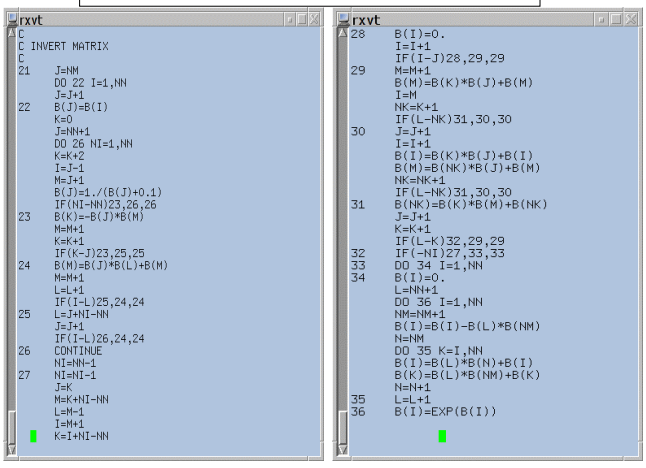
Software Languages

- Crystallographic software has been written in *machine language, assembly language, algol60, (turbo)basic, (turbo)pascal, Fortran, C, C++* and various scripting languages such as *python*
- 'Stone-age' *Fortran* based software is still ubiquitous in the small-molecule world (ORTEP, SHELX, CRYSTALS, PLATON etc.)
- New (commercial) software development mainly in C++ and scripting languages.
- A project just started in the UK to Rethink & Rewrite old Fortran based software to C++ (Durham, Oxford project).
- Old software saved in 'The Crystallographic Source Code Museum' by Armal LeBail, supposedly interesting to look for useful algorithms.





SHELX76-STYLE FORTRAN



Alternative Algorithms for the Implementation of the same Task

- Tasks can usually be programmed in a variety of ways with widely ranging claims on memory and CPU resources.
- It is important to know the actual application to make the relevant decisions.
- Following is a simple, though somewhat extreme, example from the 1960's where a theoretical idea in Direct Methods was given to a professional programmer to implement.
- The final program was nicely written and documented.
- However, the calculation didn't terminate within hours even for a trivial application ….(my mystery structure)

Problem from Symbolic Addition Method

P+ for triple H,K,H+K depends on

|E(H)E(K)E(H+K)|

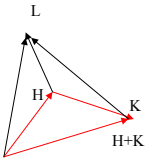
'Correlation Method' ➔ Improved P+

on the basis of P+ of three adjacent triples

|E(H)E(L)E(H+L)|

|E(K)E(L-K)E(L)|

|E(H+K)E(L-K)E(H+L)|



I.e. Strengthening of P+(|E(H)E(K)E(H+K)|

when in addition E(H+L),E(L-K),E(L) strong

(Note: Theoretically formalized in terms of neighbourhoods, Hauptman)

Two Implementations

- Implementation I: (Professional Programmer)

  1 - Search and store all triple products found with E > E(min)

  2 - Find from this list quartets of triples forming a tetrahedron

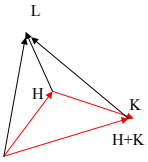
  Problem with 1: The number of triplets explodes with increasing size of the structure at hand and so memory requirements (limited to 16kW in those days)

  Problem with 2: Multiple nested loops with large range

Implementation II (by Young Student)

Generate 'correlations' on the fly during triple relation search by looping with L with E(L) > E(min) and testing for large E(L-K) and E(H+L).

Result: Completion of the search in minutes rather than hours.

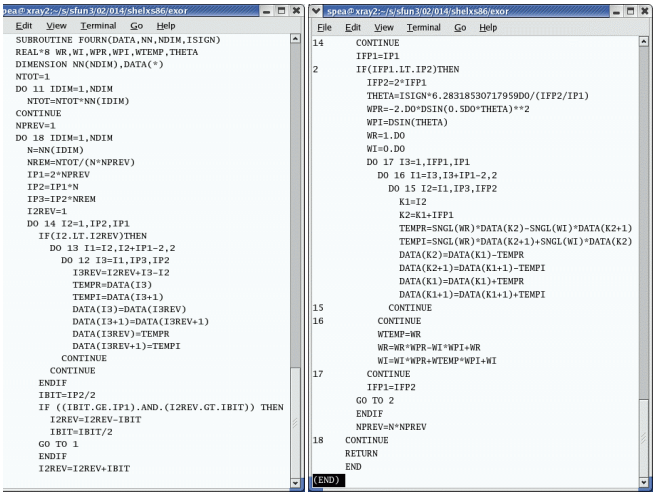


Numerical Recipes

- An excellent and rich source of numerical routines for sorting, optimisation, FFT etc. with associated background is the book *Numerical Recipes* by W.H. Press et al., that has separate Fortran and C versions

Numerical Recipe Example

- A very nice routine from NR is code with the name 'FOURN.FOR'.
- Forward and Backward FFT in N dimensions.
- In our crystallographic application: N = 3
- Code = 69 Fortran lines ! Next ….

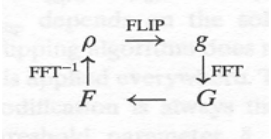


```
SUBROUTINE FOURN(DATA,NN,NDIM,ISIGN)
REAL*8 WR,WI,WPR,WPI,WTEMP,THETA
DIMENSION NN(NDIM),DATA(*)
NTOT=1
DO 11 IDIM=1,NDIM
  NTOT=NTOT*NN(IDIM)
CONTINUE
NPREV=1
DO 18 IDIM=1,NDIM
  N=NN(IDIM)
  NREM=NTOT/(N*NPREV)
  IP1=2*NPREV
  IP2=IP1*N
  IP3=IP2*NREM
  I2REV=1
  DO 14 I2=1,IP2,IP1
    IF(I2.LT.I2REV)THEN
      DO 13 I1=I2,I2+IP1-2,2
        DO 12 I3=I1,IP3,IP2
          I3REV=I2REV+I3-I2
          TEMPR=DATA(I3)
          TEMPI=DATA(I3+1)
          DATA(I3)=DATA(I3REV)
          DATA(I3+1)=DATA(I3REV+1)
          DATA(I3REV)=TEMPR
          DATA(I3REV+1)=TEMPI
        CONTINUE
      CONTINUE
    ENDIF
    IBIT=IP2/2
    IF ((IBIT.GE.IP1).AND.(I2REV.GT.IBIT)) THEN
      I2REV=I2REV-IBIT
      IBIT=IBIT/2
    GO TO 1
    ENDIF
    I2REV=I2REV+IBIT
```

```
14      CONTINUE
        IFP1=IP1
2       IF(IFP1.LT.IP2)THEN
          IFP2=2*IFP1
          THETA=ISIGN*6.28318530717959D0/(IFP2/IP1)
          WPR=-2.D0*DSIN(0.5D0*THETA)**2
          WPI=DSIN(THETA)
          WR=1.D0
          WI=0.D0
          DO 17 I3=1,IFP1,IP1
            DO 16 I1=I3,I3+IP1-2,2
              DO 15 I2=I1,IP3,IFP2
                K1=I2
                K2=K1+IFP1
                TEMPR=SNGL(WR)*DATA(K2)-SNGL(WI)*DATA(K2+1)
                TEMPI=SNGL(WR)*DATA(K2+1)+SNGL(WI)*DATA(K2)
                DATA(K2)=DATA(K1)-TEMPR
                DATA(K2+1)=DATA(K1+1)-TEMPI
                DATA(K1)=DATA(K1)+TEMPR
                DATA(K1+1)=DATA(K1+1)+TEMPI
15            CONTINUE
16          CONTINUE
            WTEMP=WR
            WR=WR*WPR-WI*WPI+WR
            WI=WI*WPR+WTEMP*WPI+WI
17        CONTINUE
          IFP1=IFP2
        GO TO 2
        ENDIF
        NPREV=N*NPREV
18    CONTINUE
      RETURN
      END
```

Application of FOURN.FOR

- *Ab Initio structure solution by charge flipping*
- See G.Oszlanyi & A. Suto, (2004) Acta Cryst A60,134.
- Procedure: cycle between reciprocal space to direct space and back after modification of the density map until convergence using forward and backward FFT.



So, No More D.M. ?

- Preliminary results on real structures, including incommensurate structures look interesting.
- There will be a lecture on this in Florence (MS20).
- Faster FFT: (Free C-library) FFTW
  However, with greater implementation and portability complexities.

Other Computing Areas

- Powder (indexing, solution, refinement)
- RDF (Billinge)
- Macro Xtal (Phasing, Building, Refinement)
- Charge Density Studies (XD)
- Least Squares and other optimisation techniques.

Other Computing Areas

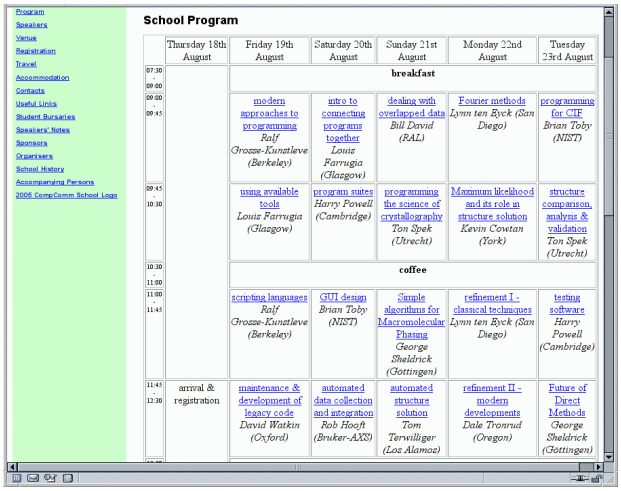
- Incommensurate Structures (solution, refinement) (Keynote lecture by Petricek in Florence).
- Graphics (GUI's and presentation)
- Data collection and data reduction.
- Databases, Structure analysis and Validation

The Program of the School

- There has been some discussion in the program commission on whether there should be two largely parallel sessions in view of a perceived growing diversion of interest.
- Eventually this path was not pursued, resulting in the current program that involves a mix of small-molecule, macro-molecule and powder interests.
- This format should provide a fruitful platform to pick up and discuss ideas from each others field.

The Program of the School

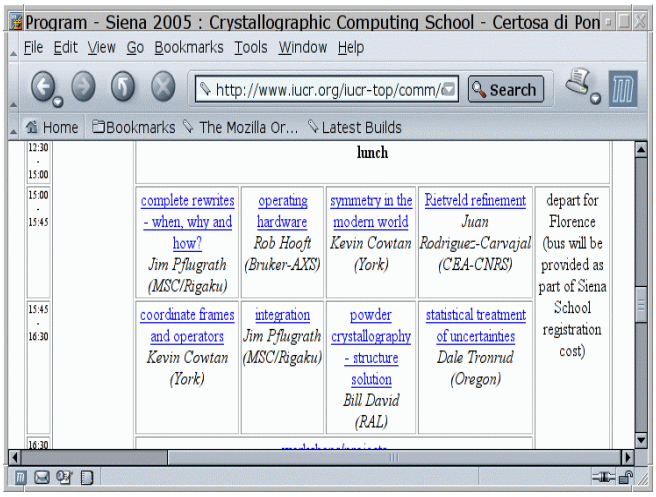
- Lecturers were asked to focus on software development and internals rather than presenting the latest science or user instruction to their software.
- Not a school to learn basic programming.
- An introduction to current software development techniques (scripting languages, toolboxes etc.)
- Hands-on projects and workshops on personal notebooks.
- Bringing together representatives of the older and a next generation interested in software development.



Program
Speakers
Venue
Registration
Travel
Accommodation
Contacts
Useful Links
Student Bursaries
Speakers' Notes
Sponsors
Organisers
School History
Accompanying Persons
2006 CompComm School Logo

School Program

	Thursday 18th August	Friday 19th August	Saturday 20th August	Sunday 21st August	Monday 22nd August	Tuesday 23rd August
07:30 - 09:00		**breakfast**				
09:00 - 09:45		modern approaches to programming *Ralf Grosse-Kunstleve (Berkeley)*	intro to connecting programs together *Louis Farrugia (Glasgow)*	dealing with overlapped data *Bill David (RAL)*	Fourier methods *Lynn ten Eyck (San Diego)*	programming for CIF *Brian Toby (NIST)*
09:45 - 10:30		using available tools *Louis Farrugia (Glasgow)*	program suites *Harry Powell (Cambridge)*	programming the science of crystallography *Ton Spek (Utrecht)*	Maximum likelihood and its role in structure solution *Kevin Cowtan (York)*	structure comparison, analysis & validation *Ton Spek (Utrecht)*
10:30 - 11:00		**coffee**				
11:00 - 11:45		scripting languages *Ralf Grosse-Kunstleve (Berkeley)*	GUI design *Brian Toby (NIST)*	Simple algorithms for Macromolecular Phasing *George Sheldrick (Gottingen)*	refinement I - classical techniques *Lynn ten Eyck (San Diego)*	testing software *Harry Powell (Cambridge)*
11:45 - 12:30	arrival & registration	maintenance & development of legacy code *David Watkin (Oxford)*	automated data collection and integration *Rob Hooft (Bruker-AXS)*	automated structure solution *Tom Terwilliger (Los Alamos)*	refinement II - modern developments *Dale Tronrud (Oregon)*	Future of Direct Methods *George Sheldrick (Gottingen)*



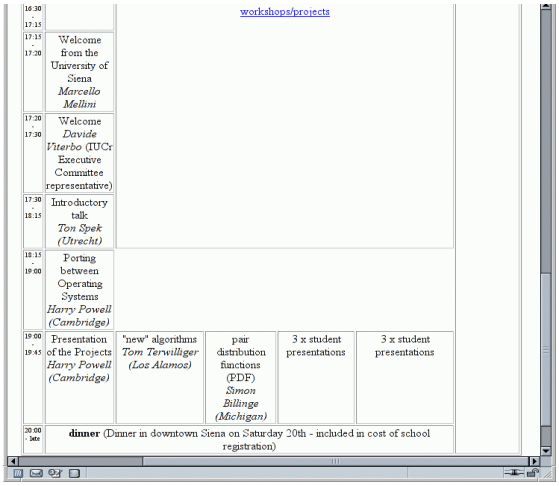
Program - Siena 2005 : Crystallographic Computing School - Certosa di Pon

File Edit View Go Bookmarks Tools Window Help

http://www.iucr.org/iucr-top/comm/ Search

Home Bookmarks The Mozilla Or... Latest Builds

12:30 - 15:00	**lunch**				
15:00 - 15:45	complete rewrites - when, why and how? *Jim Pflugrath (MSC/Rigaku)*	operating hardware *Rob Hooft (Bruker-AXS)*	symmetry in the modern world *Kevin Cowtan (York)*	Rietveld refinement *Juan Rodriguez-Carvajal (CEA-CNRS)*	depart for Florence (bus will be provided as part of Siena School registration cost)
15:45 - 16:30	coordinate frames and operators *Kevin Cowtan (York)*	integration *Jim Pflugrath (MSC/Rigaku)*	powder crystallography - structure solution *Bill David (RAL)*	statistical treatment of uncertainties *Dale Tronrud (Oregon)*	
16:30					



16:30 - 17:15		workshops/projects			
17:15 - 17:30	Welcome from the University of Siena *Marcello Mellini*				
17:30 - 17:30	Welcome *Davide Viterbo* (IUCr Executive Committee representative)				
17:30 - 18:15	Introductory talk *Ton Spek (Utrecht)*				
18:15 - 19:00	Porting between Operating Systems *Harry Powell (Cambridge)*				
19:00 - 19:45	Presentation of the Projects *Harry Powell (Cambridge)*	"new" algorithms *Tom Terwilliger (Los Alamos)*	pair distribution functions (PDF) *Simon Billinge (Michigan)*	3 x student presentations	3 x student presentations
20:00 - late	**dinner** (Dinner in downtown Siena on Saturday 20th - included in cost of school registration)				

Thanks to our Sponsors !

- Bruker-Nonius AXS
- Cambridge Crystallographic Data Center
- CCP4
- IUCr
- Max-Inf2
- Merck & Co., Inc, USA
- Oxford Diffraction
- Rigaku/MSC
- Universita degli Studi di Siena