Getting the Most Out of *SnB*

Russ Miller

Hauptman-Woodward Med. Res. Inst.

Principal Contributors:

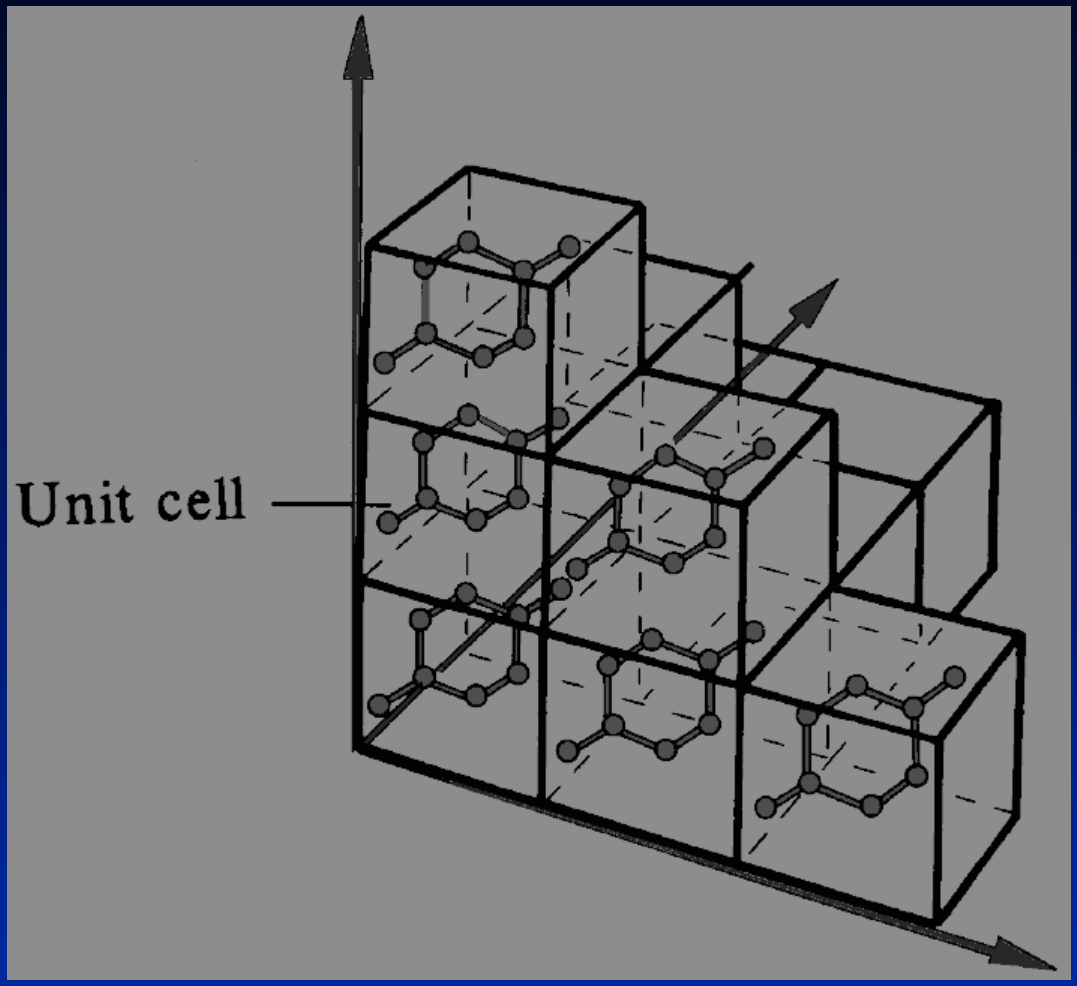
C.-S. Chang
G.T. DeTitta
S.M. Gallo
H.A. Hauptman
D.A. Langs
R. Miller
S. Potter
C.M. Weeks

Partial funding from NIH and NSF.

Outline of Talk

- X-Ray Crystallography
  - The Phase Problem
- *Shake-and-Bake*
  - The Minimal Function
- *SnB*
  - Results
- Summary
- ***Demonstration***

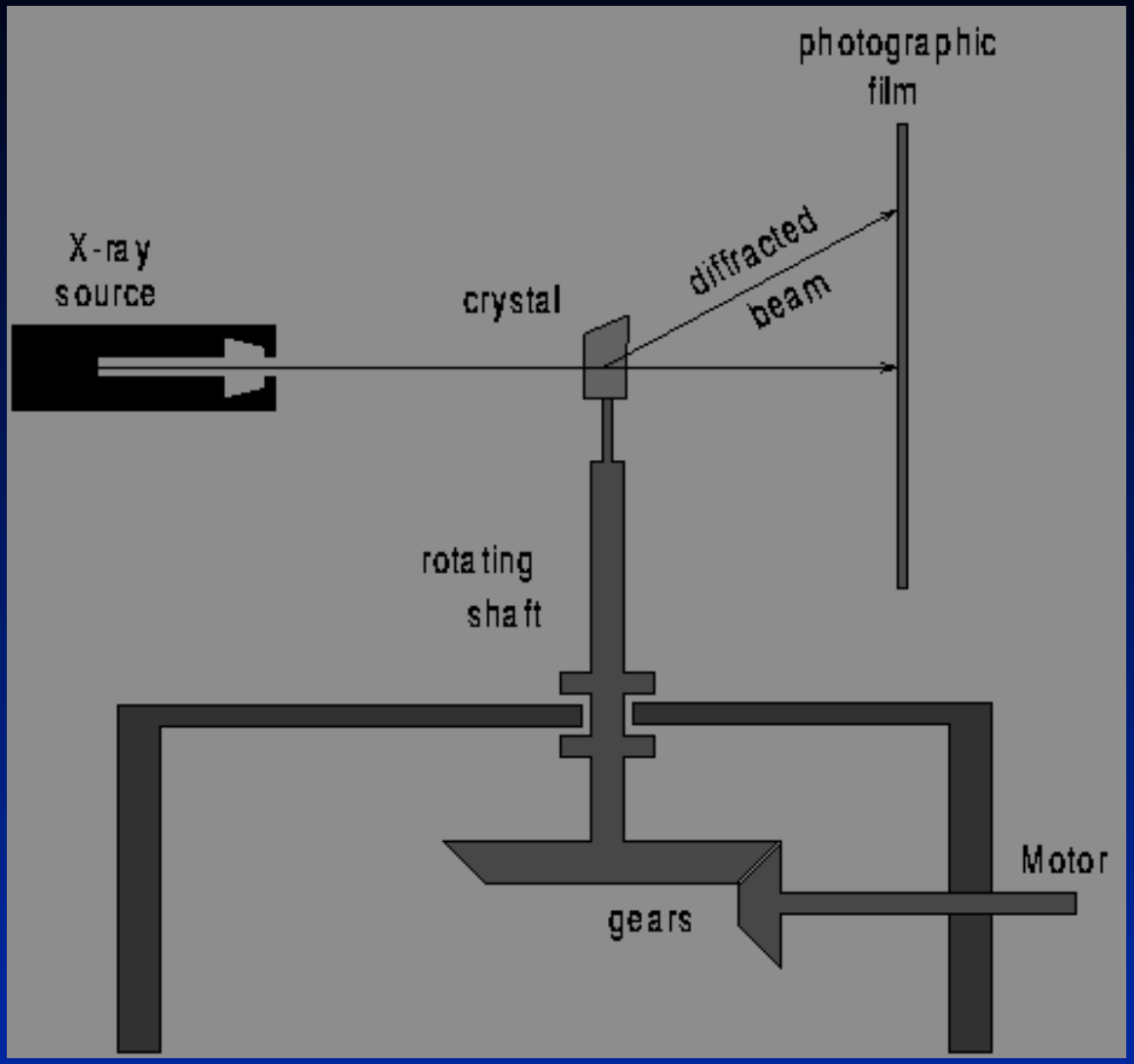
A Crystal Structure



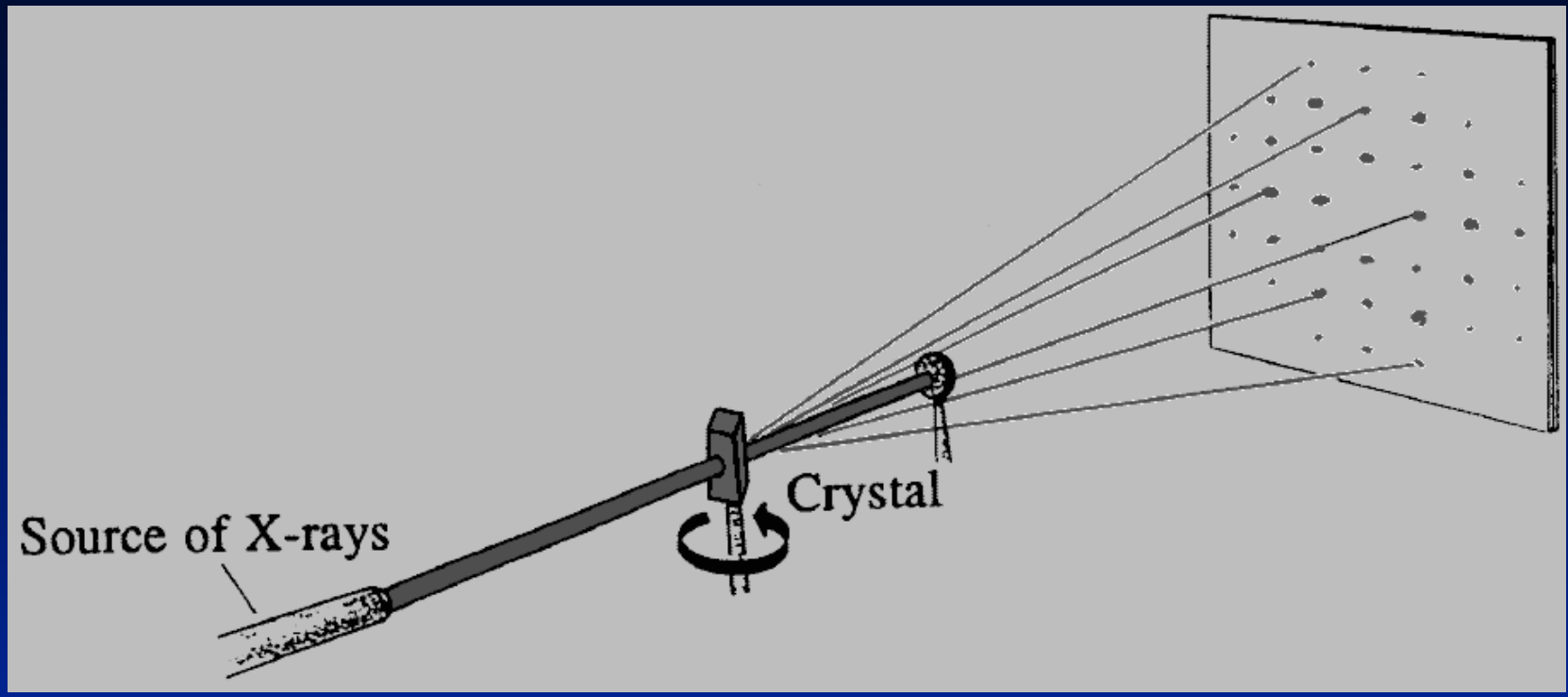
X-Ray Crystallography

- Objective: Provide a 3-D mapping of the atoms in a crystal.
- Procedure:
  - Isolate a single crystal.
  - Perform the X-Ray diffraction experiment.
  - Determine molecular structure that agrees with diffraction data.

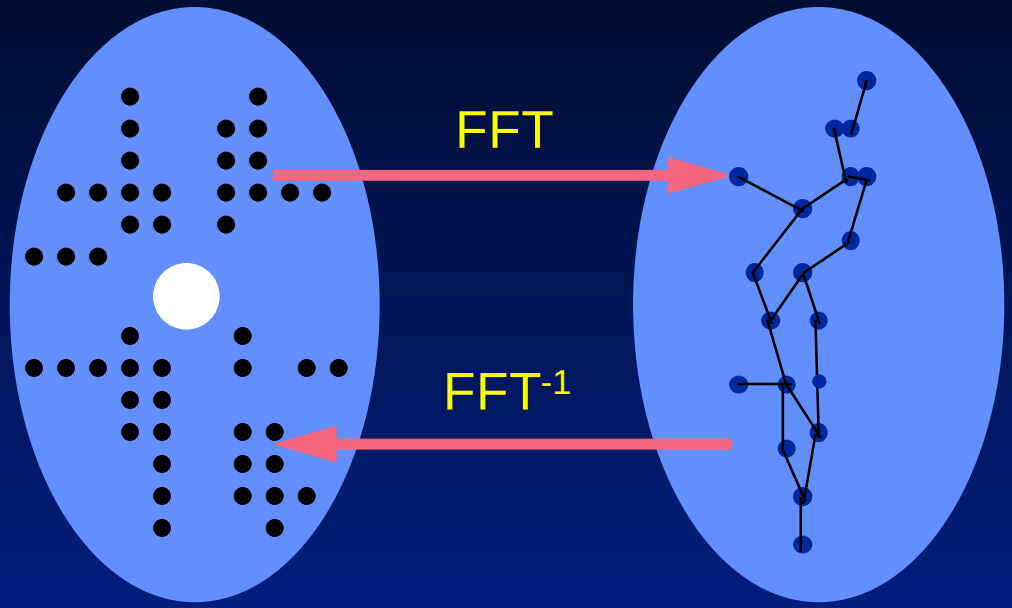
Diffractometer



The Diffraction Pattern



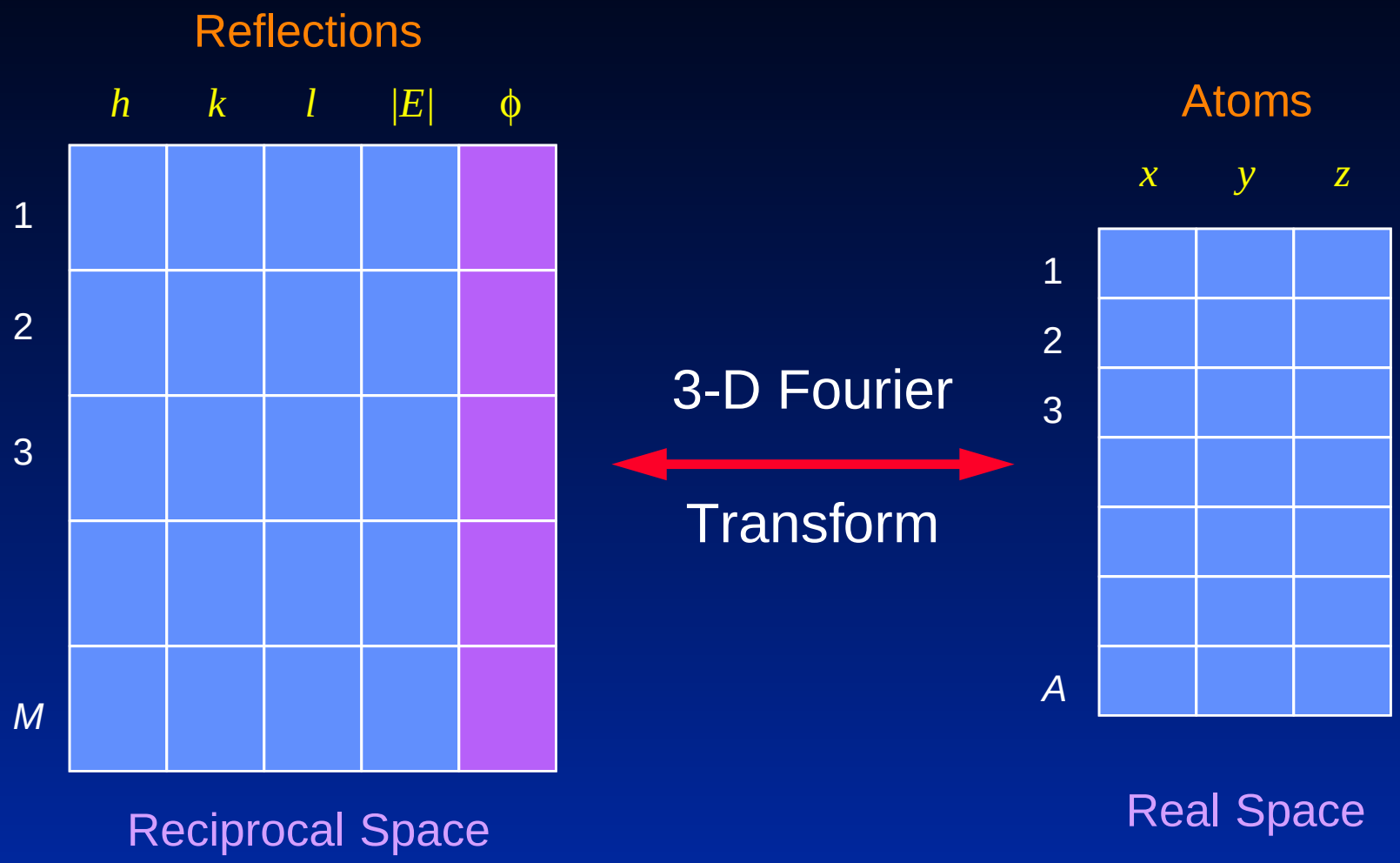
X-Ray Data & Molecular Structure



The Phase Problem

- Experiment yields:
  - reflections
  - associated intensities
- *Phase angles are lost in experiment.*
- Underlying atomic arrangement is related to the reflections by a 3-D Fourier transform.
- ***Phase Problem***: determine the set of phases corresponding to the reflections.

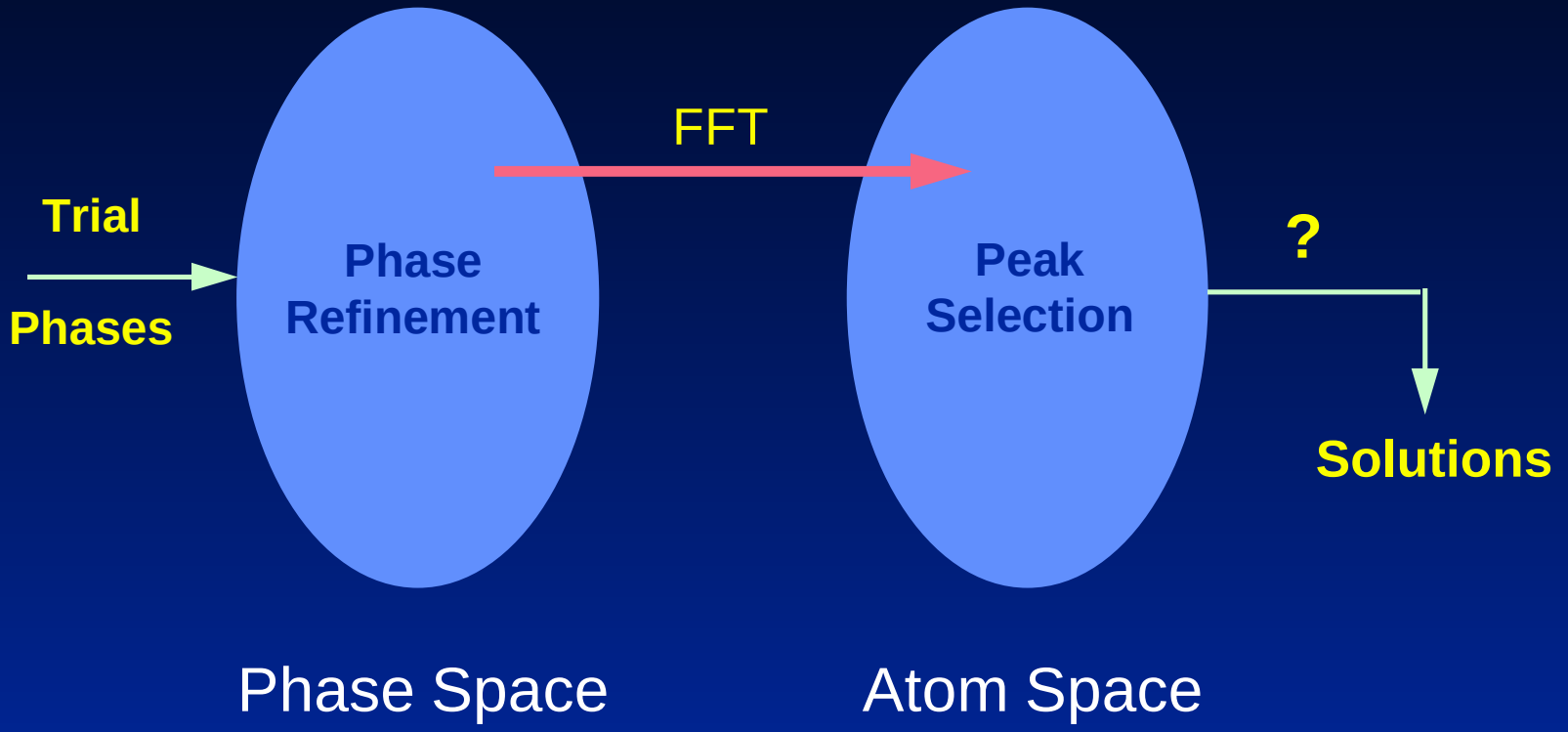
Data Structures



Structure Invariants

- *Direct Methods* exploit probabilistic theories to relate linear relationships among phases.
- A *triplet* $(\phi_h + \phi_k + \phi_{-h-k})$ has a most probable value of 0 mod 2π, given that h and k are distinct reciprocal vectors.

Conventional Direct Methods



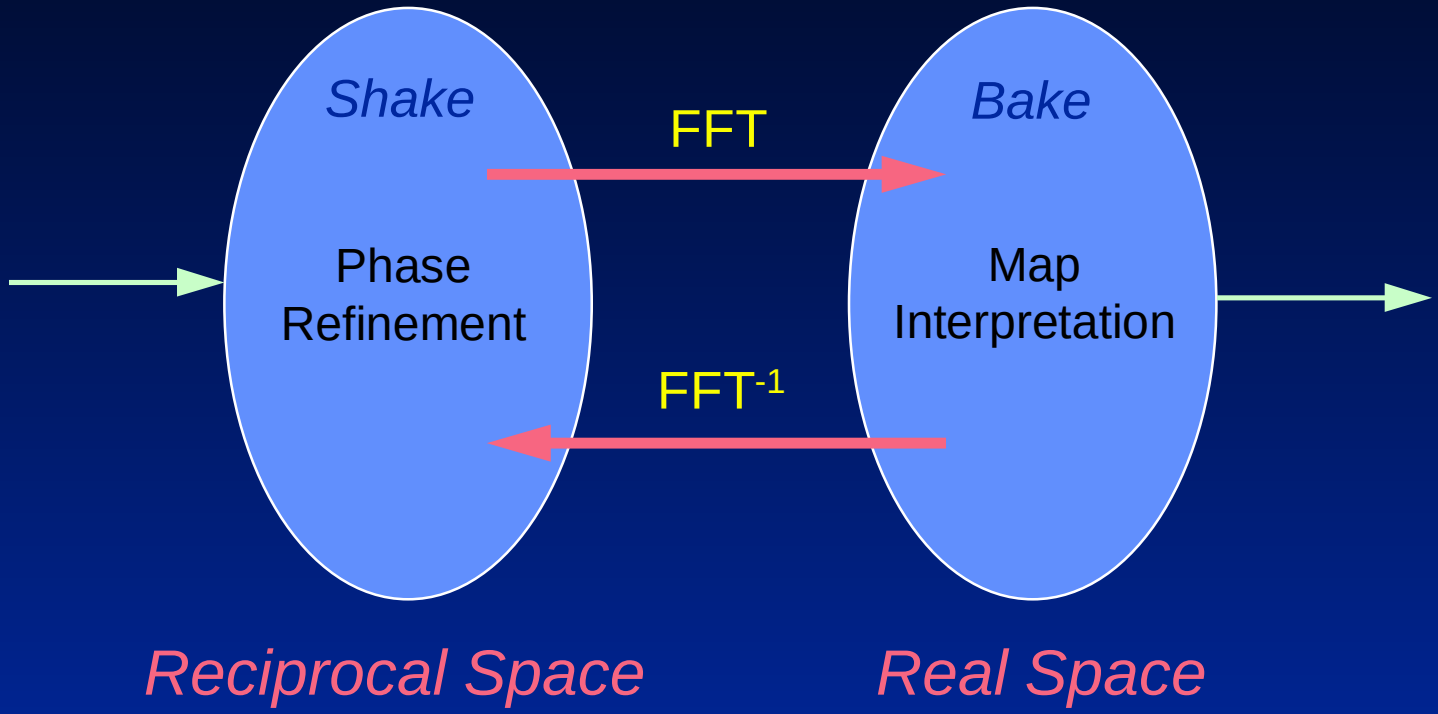
The Minimal Function

$$R = \frac{\sum_T W_T \left(\cos\phi_T - est_T\right)^2}{\sum_T W_T}$$

Triple: $\phi_T = \phi_h + \phi_k + \phi_{-h-k}$

est_T is the known expected value of $\cos\phi_T$

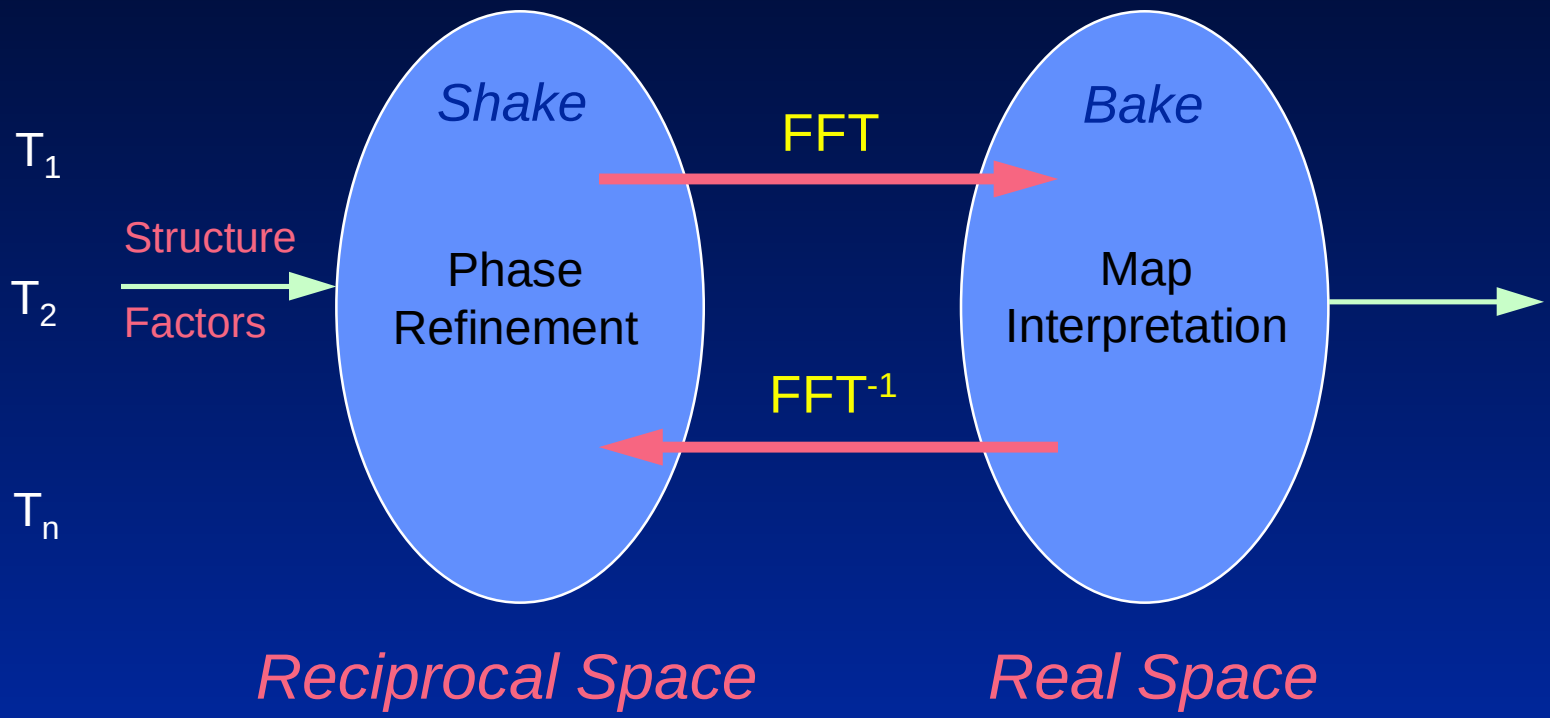
Shake-and-Bake



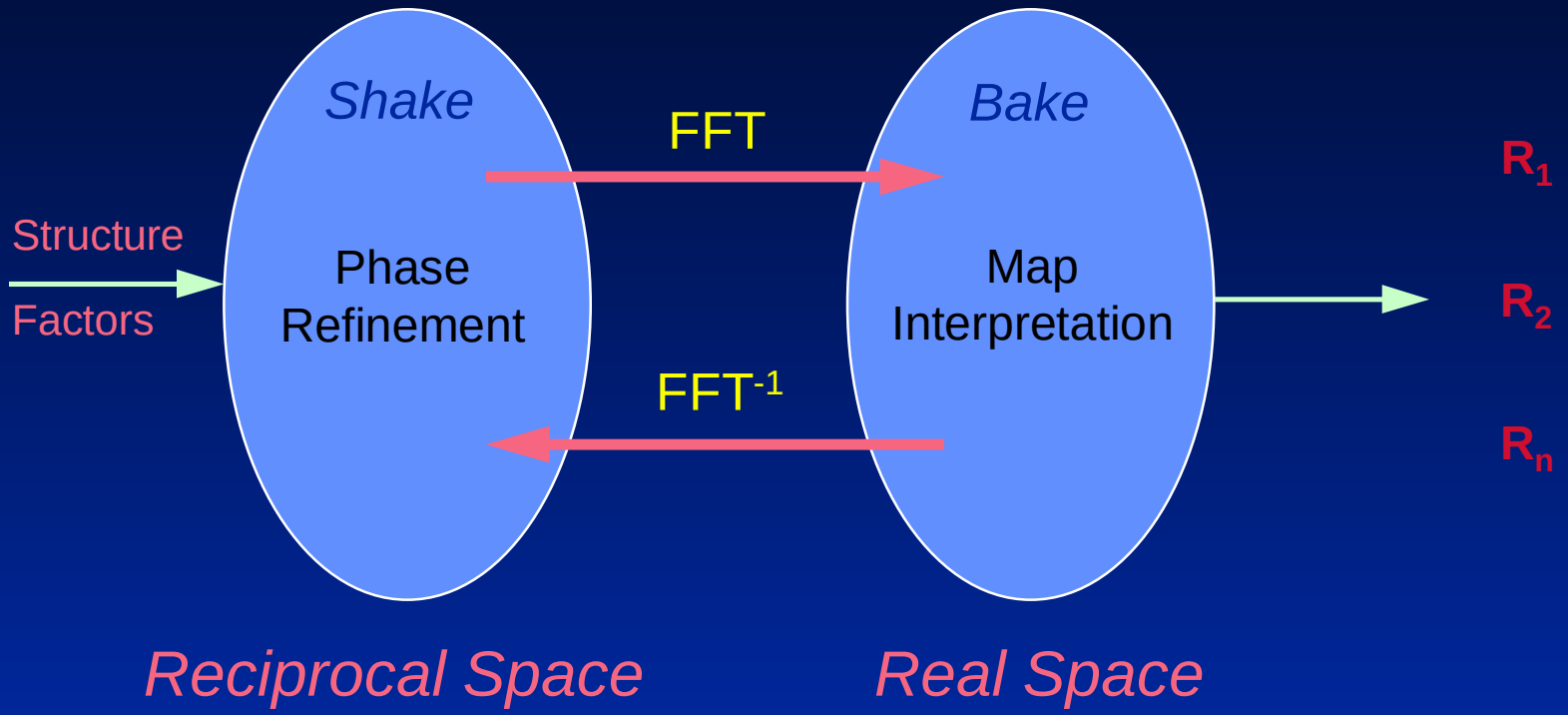
Shake-and-Bake

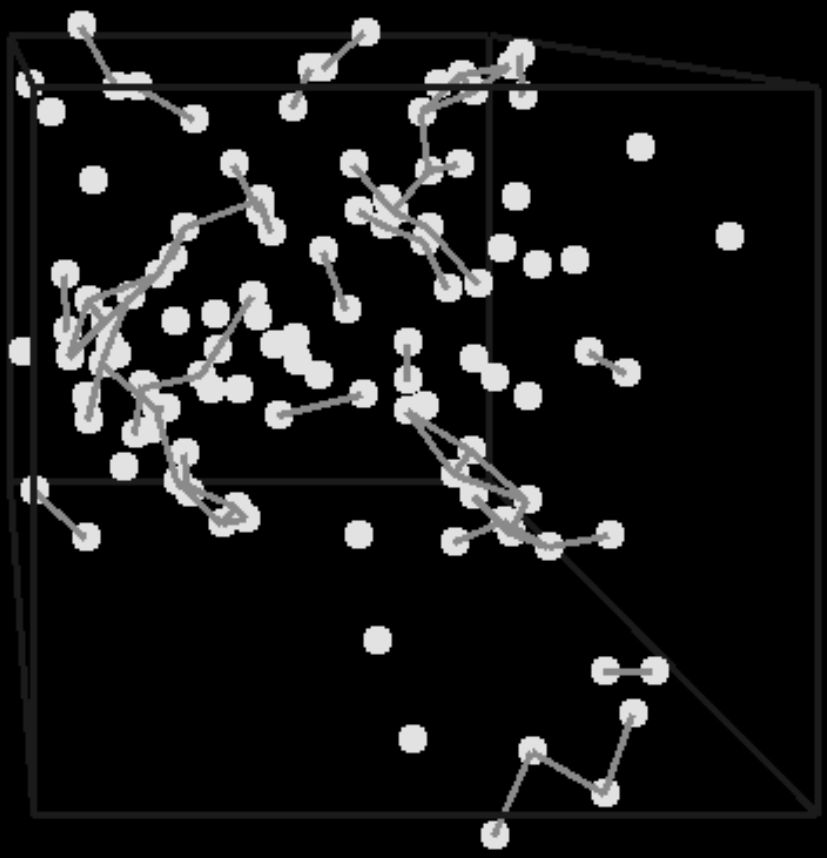
- Direct Methods Optimization Technique
- Multiple Trial Structures
- Real Space ⇔ Reciprocal Space
- Phase Refinement Techniques
  - Parameter Shift
  - Tangent Formula
  - Simulated Annealing
  - Genetic Algorithms

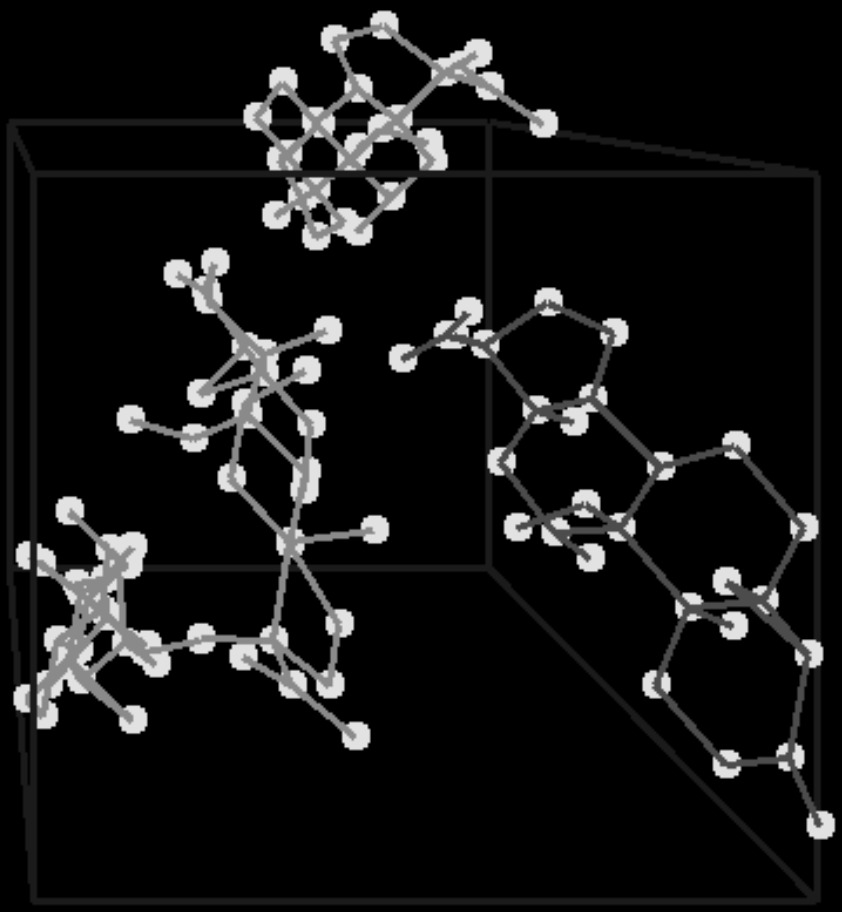
Shake-and-Bake



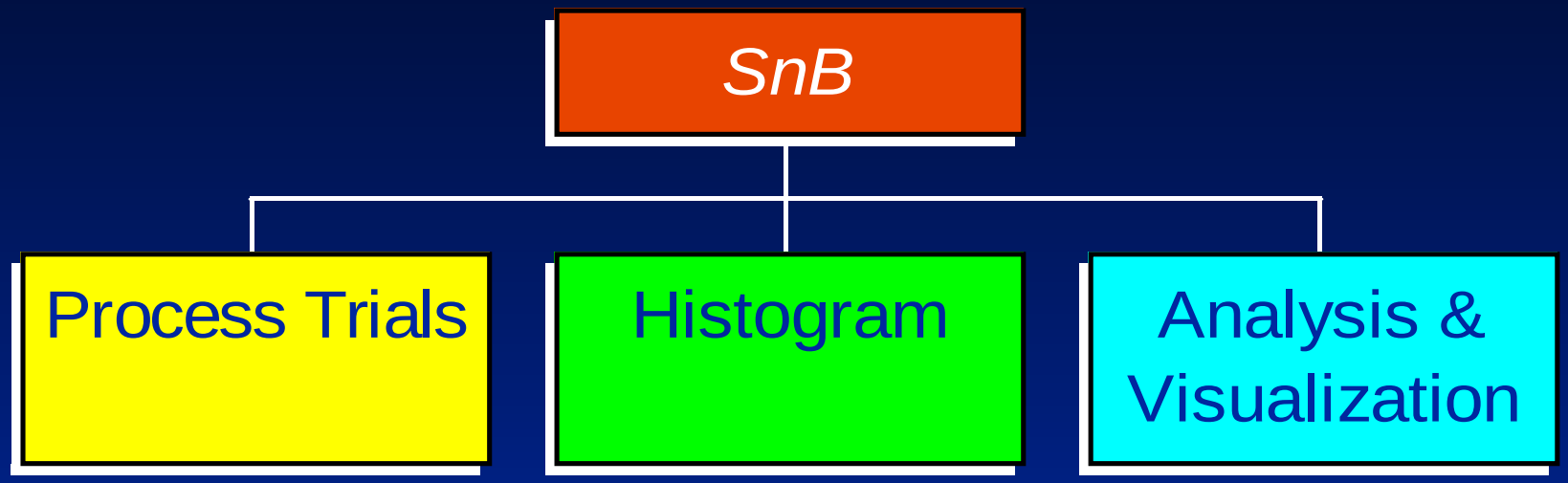
Shake-and-Bake

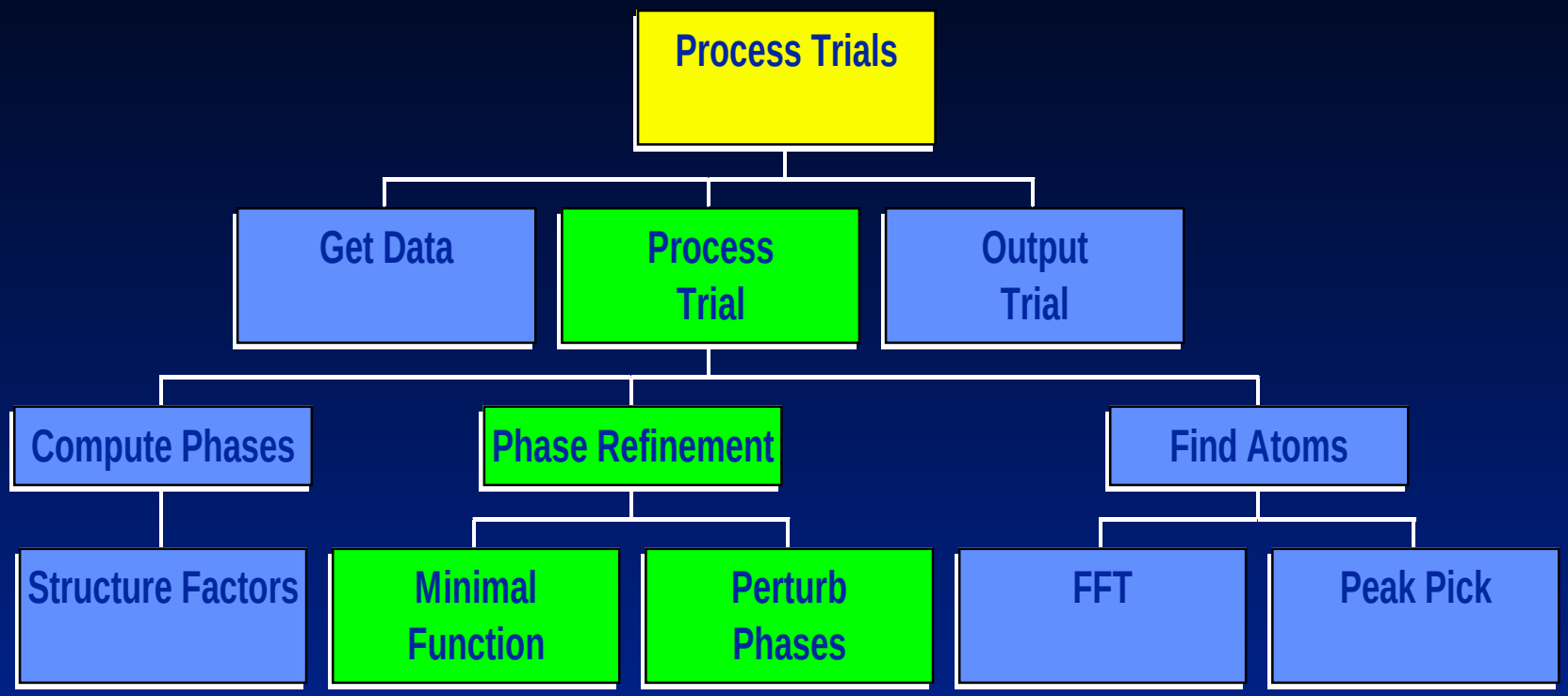






Structure of *SnB*




Process Trials
Get Data
Process Trial
Output Trial
Compute Phases
Phase Refinement
Find Atoms
Structure Factors
Minimal Function
Perturb Phases
FFT
Peak Pick

SnB Default Parameters

Atoms in asymmetric unit	n
Number of Phases	$8n$ - $10n$
Number of Triples	$70n$ - $100n$
Number of Quartets	0
Number of Cycles (Parameter Shift, Cent. Search)	$n/2$
Number of Cycles (Tangent Refinement)	$n/4$
E-Fourier Recycling Steps	C

Crambin

Atoms in asymmetric unit	*400*
Number of Phases	4,000
Number of Triples	40,000
Number of Quartets	0
Number of Cycles (Parameter Shift)	*200*
E-Fourier Recycling Steps	*5*

Structures

Structure	Atoms	Space Group	Success Rate
Ternatin	**110**	**$P2_12_12_1$**	**2%**
Scripps	**144**	**P1**	**??**
MSC	**180**	**P1**	**(after 8 years failure)**
UCLA	**302**	**C2**	**??**
Gramicidin A	**317**	**$P2_12_12_1$**	**0.3%**
OCI / U. of T.	**408**	**P1**	**5%**
Rubredoxin	**500**	**$P2_1$**	**2.7%**
ToxII	**580**	**$P2_12_12_1$**	**1/1600**

Cost-Effectiveness

- Tangent Formula in *SnB*
  - Smaller structures
  - Carefully chosen parameters
- Space Group
  - P1
  - P-1
- Simulated Annealing
- Genetic Algorithms

Running Times

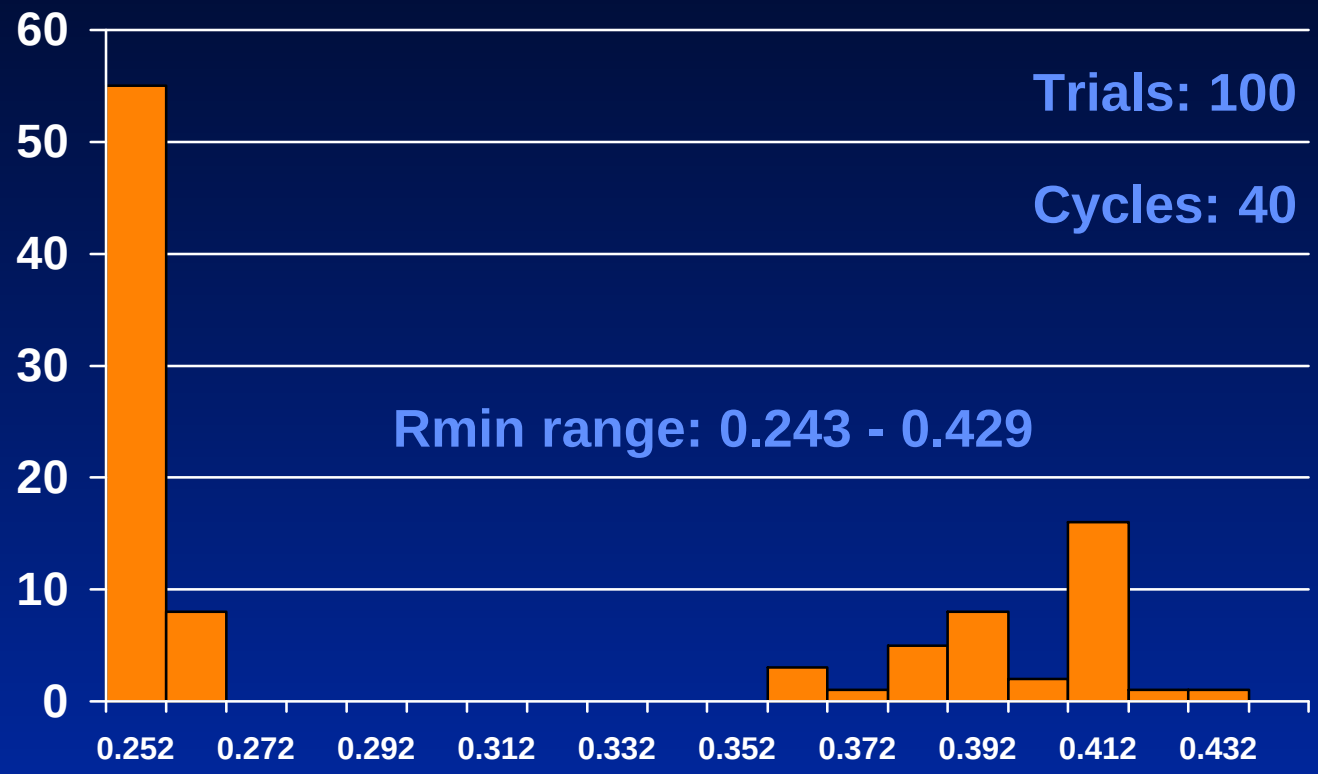
(Average Seconds/Cycle on Iled)

Machine	Total	StrFac	Refine	FFT	Pk Pk
DEC/Alcor 266	0.672	0.160	0.248	0.196	0.029
Sun UltraSparc	0.972	0.299	0.308	0.276	0.044
IBM RS6000	1.233	0.327	0.400	0.323	0.106
SGI Pwr Ind 2	1.247	0.210	0.426	0.485	0.068
DEC/Alpha 150	1.822	0.437	0.682	0.520	0.075
SGI Indigo	2.975	0.881	0.760	1.037	0.163
Sparc 10	3.140	0.821	0.919	1.111	0.162
Pentium 75	3.535	1.181	0.995	1.205	0.154
i486dx2-66	11.396	2.175	2.423	5.467	0.711

Sample Histogram

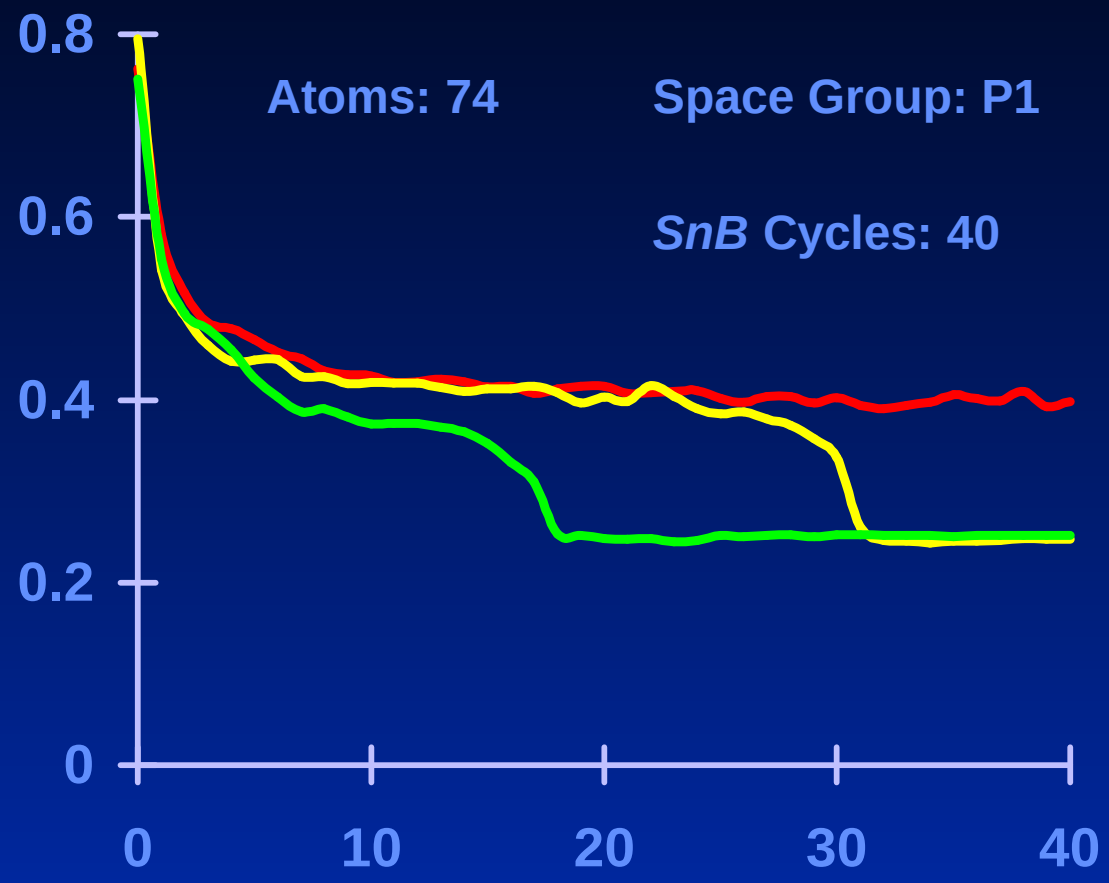
Structure: ph8755

Atoms: 74 Space Group: P1



The Minimal Function is Diagnostic

Structure: Ph8755



Additional Applications of *SnB*

- Fragment Recycling (D.A. Langs)
  - Large radius of convergence
  - Superior to traditional tangent formula (RANTAN)
- Electron Diffraction Data (Dorset & McCourt)
  - High success rates
  - Small % of data out of principle plane

Fragment Recycling

CBT: 132 atoms, P21, 20-atom fragment

Base: 20-atom random trials yield 2.5% success rate

	Deformed RMS Model	Deformed RMS Model	Angular RMS Model	Positional RMS Model
RMS A	**Soln (%)**	**Soln (%)**	**Soln (%)**	**Soln (%)**
0.0	19.80			
0.1	7.20			
0.15	2.20			
0.2	0.15			100
0.3		100	100	96
0.4		86	96	72
0.5		50	72	22
0.6		12	66	0
0.7		2	57	
0.8		0	40	
1.0			32	
1.2			6	
1.4			0	

Rantan

Shake-and-Bake

Electron Diffraction Data (Known) (with D. Dorset and M. McCourt)

Structure	Atoms	Space Group	Success Rate	3-D
Copper Chloride	5	$P2_1/m$	21%	52%
Poly-Butene-1	8	$P2_12_12_1$	28%	37%
Poly-ε-Caprolactone	8	$P2_12_12_1$	41%	15%
Copper Perchloro-phthalocyanine	16	Cmm2	Yes	0%
n-Paraffin	36	$Pca2_1$	Yes	50%

Computing Platforms

- Unix Workstations
  - SGI, Sun, DEC/Alpha
  - IBM RS6000
- Parallel Computers
  - Thinking Machines Corp. CM-5
  - Cray T3D
  - IBM SP2
- Cray C90

Availability of *SnB*

- MSC's teXsan 1-7.2
- MacScience
- Pittsburgh Supercomputing Center
- Hauptman-Woodward Inst.
- Cornell Theory Center

SnB Sites

Abbott Labs	Clemson	IVIC (Venezuela)	Pittsburgh	Windsor
Adelaide	CUHK (Hong Kong)	Keene State	Roche	Yale
Ames	Dalhousie	Kentucky, Louisville	Scripps	York
Ariad	Delaware	Loughborough	Shell	Zurich
ANU	Dow	McMaster	Southampton	Kodak
Berkeley	Dupong	Michigan State	South Carolina	LLNL
Birmingham	Duquesne	Minnesota	St. Andrews	SmithKline Beecham
Boehringer Ingelheim	East Anglia	MIT	Stanford	NIH
Boston College	Genetics Institute	NIU - Illinois	Sydney	Cornell
Brandeis	Griffith	Northwestern	SYSNU (Taiwan)	UCSF
BYUH	Guelph	Oak Ridge (ORNL)	UBC	Imerial College (London)
Calgary	Helsinki	Pavia	UCLA	Caltech
Cambridge	Hoffman Laroche	Univ. Penn.	UGA	Madras
Chinoin (Hungary)	Hong Kong University	Pfizer	USC	Univ. Queensland (Australia)

SnB Addresses

- http://www.hwi.buffalo.edu/SnB/
- snb-requests@hwi.buffalo.edu
- snb-comments@hwi.buffalo.edu

Summary

- Alternative structure determination method
- Targeted at 100-800 atom structures
- Publicly available
- Recommend: LEVY / EVAL (Bob Blessing)
- Current Research
  - Parameter tuning / Code Optimization (2X)
  - Additional FOMs
  - Lower resolution data
  - SIR/SAS data