SnB: A STATUS REPORT

Russ Miller[1,2] and Charles M. Weeks[2]

[1] *Dept. of Comp. Sci., SUNY-Buffalo*
[2] *The Medical Foundation of Buffalo*

{miller,weeks}@mfb.buffalo.edu
(716) 856-9600

Partial funding from NIH and NSF.

Computing from TMC, PSC, Intel, and NIH.

Principal Contributors

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

History

1988: Hauptman develops *Minimal Function*

1989: Simulated Annealing in Reciprocal Space

1991: *Shake-and-Bake* (Parameter Shift)

1992: *SnB*

1992: Genetic Algorithms

1993: Simulated Annealing in *SnB*

1993: Tangent Formula in *SnB*

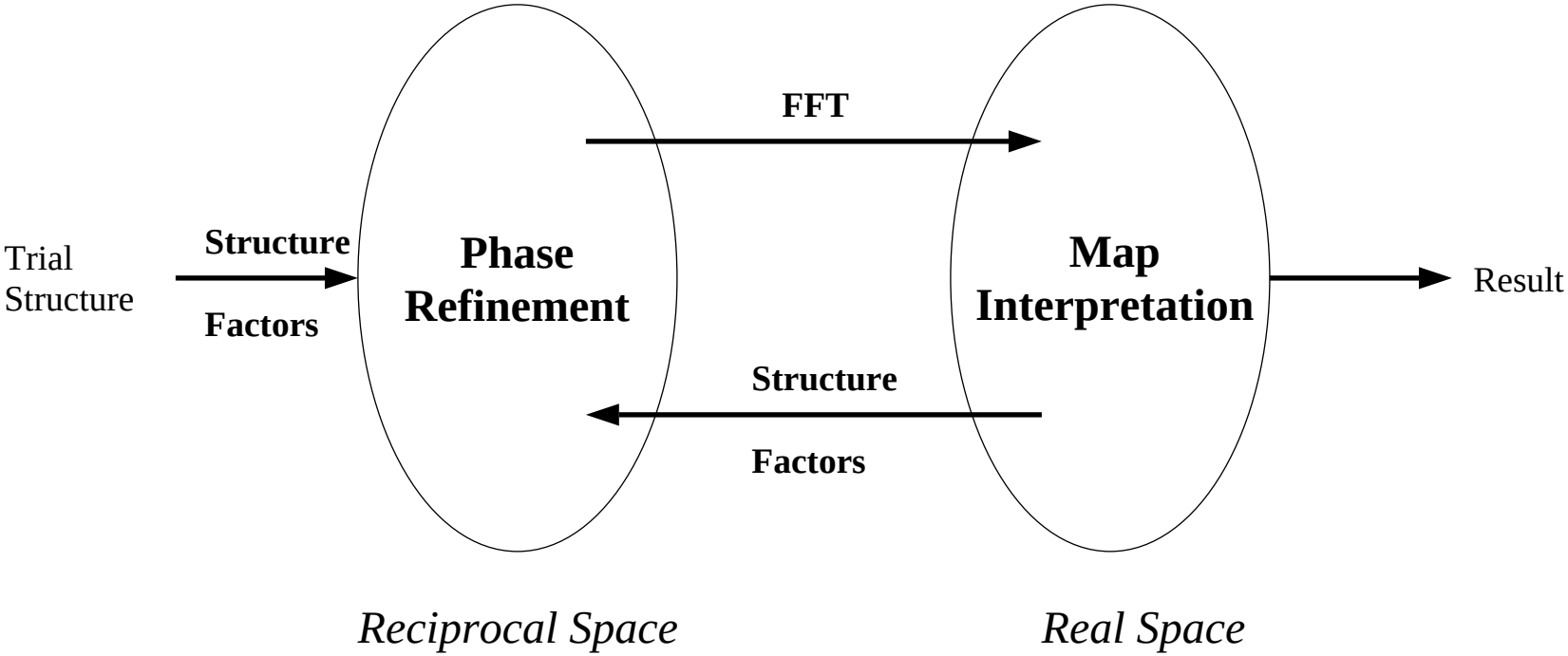
Computing Platforms

1. Intel iPSC/2 (UB)
2. Intel iPSC/860 (NIH/Intel)
3. Thinking Machines Corporation CM-2/200 (TMC/PSC)
4. Network of Sun Workstations (UB/CIT)
5. Thinking Machines Corporation CM-5 (TMC/PSC)
6. Network of SGI Workstations (MFB)
7. Cray C90 (PSC)
8. Cray T3D (PSC)

Overview

1. *Shake-and-Bake*
   - Direct Methods Optimization Approach
   - Real Space $\Longleftrightarrow$ Reciprocal Space
   - *Minimal Function* also as FOM
   - Sets of Random Trial Structures
2. *SnB* is the program based on *Shake-and-Bake*.
3. Phase refinement techniques considered:
   (a) parameter shift
   (b) modified global binary search
   (c) tangent formula
   (d) gradient descent
   (e) simulated annealing
   (f) genetic algorithms

Shake-and-Bake



The Minimal Function

$$R = \frac{\sum_T W_T(\cos\phi_T - I_T)^2 + \sum_Q W_Q(\cos\phi_Q - I_Q)^2}{\sum_T W_T + \sum_Q W_Q}$$

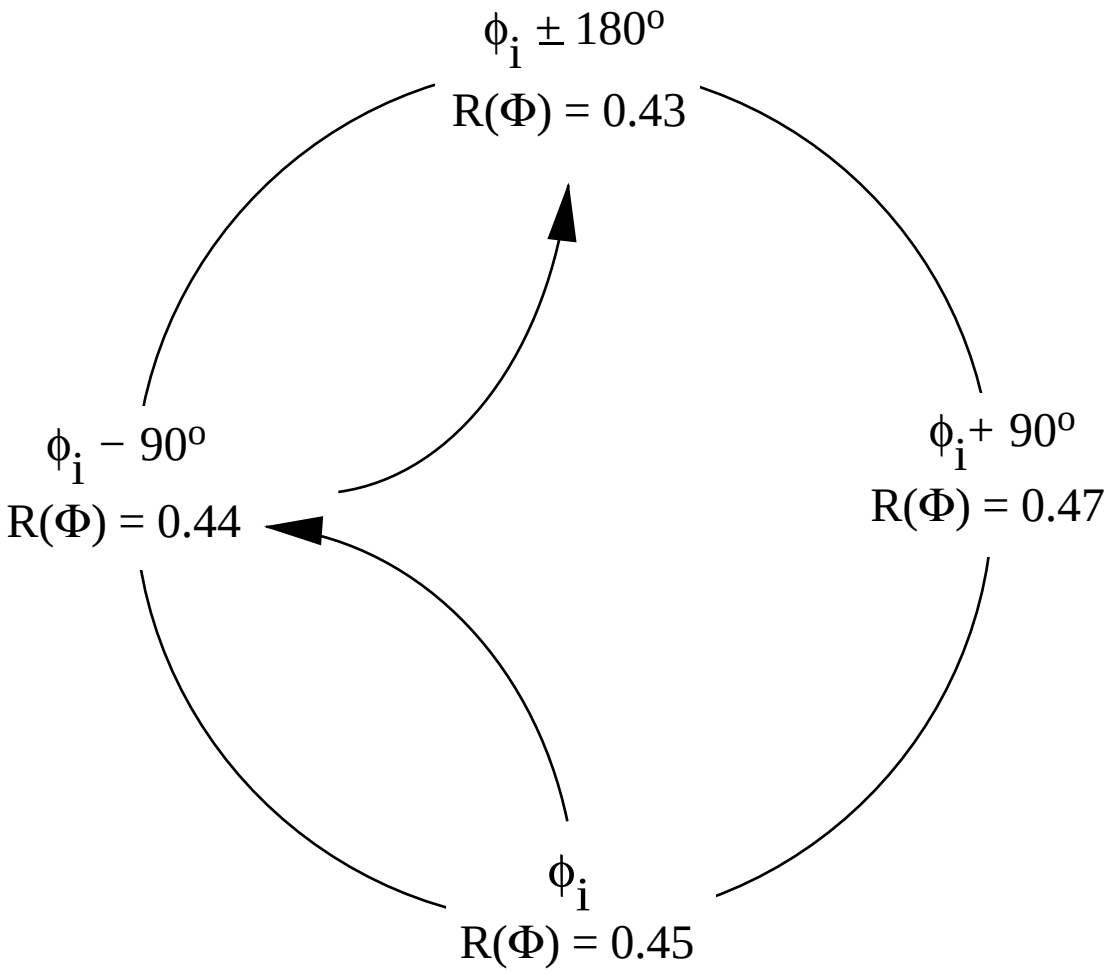
$$\phi_T = \phi_h + \phi_k + \phi_{-h-k}$$

$$\phi_Q = \phi_l + \phi_m + \phi_n + \phi_{-l-m-n}$$

I_T, I_Q — Known expected value of cosine of corresponding structure invariant.

W_T, W_Q — Weight factor inversely proportional to the variance.

Phase Refinement by Parameter Shift



Default Parameters for n Atom Structure

1. Random Atoms per Trial Structure: n
2. Phases: $10n$
3. Triples: $100n$
4. Quartets: 0
5. Refinement Routine: Parameter Shift
   - P1: (π/2, 2, 1)
   - Centrosymmetric: (π, 1, 1)
   - Other: (π/2, 2, 3)
6. *SnB* Cycles: $n/2$
7. Restricted Phases: no
8. Heavy Atoms: yes

SnB Structure Determination Procedure

```
 1.  Search path: ./
 2.  Reflection file prefix: ph8755
 3.  Space group: P1
 4.  Cell constants:
       A:    9.1300     ALPHA:   96.7300
       B:   10.5080     BETA :   99.2900
       C:   15.0770     GAMMA:   89.0900
 5.  Contents of the asymmetric unit: O14,N9,C51,H83
 6.  Generate new invariant set: Yes
       Number of phases to use: 740
       Number of triples to use: 7400
       Number of negative quartets to use: 0
       Save invariants to file: ./ph8755.inv
 7.  Generate random trial structures: Yes
       Number of trials to generate: 1000
       Random number seed: 11909
       Starting atoms per trial: 74
       Save random trials to file: ./ph8755.trials
 8.  Trial processing information
       Number of trials to process: 64
       Beginning at trial number: 1
       Number of Shake-and-Bake cycles: 10

 9.  Exploit knowledge of heavy atoms: No
10.  Refinement method: Parameter Shift
       Exploit knowledge of restricted phases: No
       Number of complete passes through phase set: 1
       Number of attempted phase shifts per pass: 2
         Phase shift used in pass #1:  90
11.  Number of peaks to select: 74
12.  Optional information storage
       Keep trace file containing Minimal Function values: Yes
       Store all final structures in file: No

Would you like to make any changes [n]:
```

Recognizing a Solution

Structure Name: ph8755

Number of Atoms: 74
Number of trials processed: 64
Lowest Rmin value: 0.244
Highest Rmin value: 0.471

No. of Shake-and-Bake cycles: 10
Number of Phases: 740
Number of Triples: 7400
Number of Quartets: 0

Rmin range			Num in range	
0.244	to	0.255	3	***
0.256	to	0.267	0	
0.268	to	0.279	0	
0.280	to	0.291	0	
0.292	to	0.303	0	
0.304	to	0.315	0	
0.316	to	0.327	1	*
0.328	to	0.339	2	**
0.340	to	0.351	2	**
0.352	to	0.363	5	*****
0.364	to	0.375	8	********
0.376	to	0.387	10	**********
0.388	to	0.399	8	********
0.400	to	0.411	4	****
0.412	to	0.423	10	**********
0.424	to	0.435	4	****
0.436	to	0.447	5	*****
0.448	to	0.459	1	*
0.460	to	0.471	1	*

Hit RETURN go back to main menu.

Previously Known Structures

Structure	Atoms	Space Group	Success Rate
Prostaglandin E_2	25	$P1$	20%
Prostaglandin $F_1\beta$	25	$C2$	8
Aldosterone	27	$P2_1$	14
9α-Methoxycortisol	28	$P2_12_12_1$	16
AZET	48	$Pca2_1$	21
Tetrahymanol	63	$P2_1$	3
APAPA	69	$P4_12_12$	1
Antibiotic A204A	71	$C2$	4
Emerimycin	74	$P1$	91
Isoleucinomycin	84	$P2_12_12_1$	12
Meso-valinomycin	84	$P\bar{1}$	1
Synthetic Pyrrole	96	$P1$	33
Sea Urchin Peptide	115	$P2_12_12_1$	1
Hexadeca Isoleucinomycin	127	$P2_12_12_1$	2
Cholesterol Butanoate	132	$P2_1$	3
Gramicidin A	317	$P2_12_12_1$	0.3
Crambin	400	$P2_1$	4
Rubredoxin	500	$P2_1$	2.7

Previously Unknown Structures

Structure	Atoms	Space Group	Success Rate
Zomepirac (Weeks/Miller)	24	$P2_1/c$	2%
Tolmetin Sodium Salt (Weeks/Miller)	40	$P\bar{1}$	1%
Boc-Gln-D-EtA-Hyp-Ala-Phol (Ciszak/Smith)	47	$P2_12_12_1$	9%
7-KetoCholesterol (McCourt/Dorset)	58	$P2_1$	1%
Full-Retro Valinomycin (Langs)	78	$P1$	1%
Cholesteryl Myristate / Cholesteryl Pentadeconoate (McCourt/Dorset)	87	$C2$	1%
Ternatin_E (Miller/Langs)	105	$P2_12_12_1$	1%
Ternatin_D (Miller/Langs)	110	$P2_12_12_1$	2%

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%

SnB – **Tangent Refinement**

1. Traditional Tangent Formula

2. Hull-Irwin Weights

   - Improvement for Small Structures
   - Worse for Larger Structures

3. Feedback

4. Optimal SnB Cycles: $n/5$

5. Optimal Iterations: 1 or 2

Cost-Effectiveness

Structure	Atoms	S.G.	T.R.	P.S.
Nordihydroquaiaretic Acid	22	$P2_1/c$	213.48	123.15
Prostaglandin E_2	25	$P1$	194.60	80.70
9α-Methoxycortisol	28	$P2_12_12_1$	101.44	37.43
3β-hydroxy-16-methyl-5,16-pregnadien-20-one	48	$P1$	405.3	209.3
Tetrahymanol	63	$P2_1$	3.01	2.91
Emerimycin	74	$P1$	65.51	46.41
Isoleucinomycin	84	$P2_12_12_1$	4.10	1.88
Meso-valinomycin	84	$P\bar{1}$	1.03	0.67
Synthetic Pyrrole	96	$P1$	12.09	9.27
Ternatin	105	$P2_12_12_1$	0.03	0.06
Hexadeca Isoleucinomycin	127	$P2_12_12_1$	0.05	0.06
Cholesterol Butanoate	132	$P2_1$	0.43	0.30
Gramicidin A	317	$P2_12_12_1$	No	Yes
Crambin	400	$P2_1$	0.02	0.02

Genetic Algorithms

Initialize P, a population of n randomly generated individuals.

Evaluate $F(i)$, $\forall\, i \in P$, where F is the pre-specified fitness function.

while (another generation is required) *do*

Selection: Create P'_s from P based on F.

Crossover: Randomly mate members of P'_s, interchanging a contiguous substring for each of the $\frac{n}{2}$ such pairs, to create P'_c.

Mutation: From P'_c, randomly flip bits in a subset of the n members to create the next generation P.

Evaluation: *Compute* $F(i)$, $\forall\, i \in P$.

end while

Genetic Algorithms

- Parallel Genetic Algorithms
- Demes (isolated subpopulations)

  Evaluation $\rightarrow$ Exchange $\rightarrow$

  Selection $\rightarrow$ Crossover $\rightarrow$ Mutation

- Combination of Fine- and Coarse-Grained
- Achieve Convergence
- Extremely Poor Performance
- Low Cost-Effectiveness

Simulated Annealing

begin
  Initialize:
    configuration $\mathbf{x} = \mathbf{x_0}$
    cooling parameter $c = c_0$
    objective function $C = C_0$
    cooling step $m = 0$
  repeat
    repeat
      Generate new configuration $\mathbf{x_i'}$ from $\mathbf{x_i}$
      Compute C_i'
      if $(C_i' - C_i < 0)$ then **accept** else
          if $e^{(C_i - C_i')/c} > random[0,1)$ then **accept**
      if **accept** then $x_i = x_i'$
    until **Equilibrium**
    $c_{m+1} = f(c_m)$
    $m = m + 1$
  until **Convergence**
end

Simulated Annealing

- Initial Temperature vs. Structure Size
- Rate of Cooling is Fixed
- Results Encouraging

			Cost-Effectiveness	
Structure	Atoms	S.G.	S.A.	P.S.
9α-Methoxycortisol	28	$P2_12_12_1$	75.5	29.7
Emerimycin	74	$P1$	69.8	38.3
Isoleucinomycin	84	$P2_12_12_1$	3.5	1.7

Fragment Recycling

CBT: 132 atoms, $P2_1$, 20-atom fragment

Base: 20-atom random trials yield 2.5% success

	Tangent Formula (RANTAN)	Shake-and-Bake (*SnB*)		
	Deformed RMS Model	Deformed RMS Model	Angular RMS Model	Positional RMS Model
RMS Å	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	

Availability of *SnB*

- Cray C90 at Pittsburgh Supercomputing Center
- Cray T3D at Pittsburgh Supercomputing Center
- Thinking Machines Corporation CM-5
- Molecular Structures Corporation

Concluding Remarks

- Available / Unix
- Success Rate: 100% (6/94)
- Manual / Help Screens
- Parameter Tuning
- Space Group Tuning (e.g., P1)
- Evaluating Simulated Annealing
- Evaluating Tangent Formula
- Electron Diffraction Data