

# 遺伝的交叉を用いた並列シミュレーテッドアニーリング - 分散メモリ型並列計算機への実装モデル -

小掠 真貴<sup>†</sup>, 廣安 知之<sup>††</sup>, 三木 光範<sup>††</sup>, 角 美智子<sup>‡</sup>

本研究では、遺伝的交叉を用いた並列シミュレーテッドアニーリング(PSA/GAc)の分散メモリ型並列計算機への実装モデルの検討を行う。PSA/GAcは著者らが提案しているアルゴリズムである。複数のプロセス上で並列にシミュレーテッドアニーリングの操作を行い、一定期間のアニーリングを行った後、2プロセス間で遺伝的アルゴリズムのオペレータである交叉により情報交換を行う。これにより、大局的にはいくつかの準最適解が存在し、局所的には無数の準最適解を有するような問題に特に有効である。しかし、PSA/GAcは共有メモリ型計算機を対象として開発されたため、PCクラスタのような分散メモリ型並列計算機に実装するためには、新たな実装モデルが必要である。本研究ではそのようないくつかの実装モデルを考案した。対象として最適化問題の一つであるタンパク質の構造解析をとりあげ、それぞれのモデルの性能について検討を行った。タンパク質の構造解析ではSAを中心に解析が行われているが、タンパク質のエネルギー関数は複雑であるため、SAでは膨大な計算時間が必要になるという問題点がある。PSA/GAcによってタンパク質の構造解析を行った場合、従来用いられていたSAと比較して、計算時間の短縮と高い解探索能力を示すことが明らかとなっている。本研究では、考案した数種のPSA/GAcの並列実装モデルを、計算時間と解探索能力の面からそれぞれ比較し、有効性の検討を行った。PSA/GAcの並列実装モデルによる構造解析は、並列計算機を用いることでPSA/GAcよりもさらに計算の高速化が図れ、また解探索能力も向上した。

## Parallel Simulated Annealing using Genetic Crossover - Models for distributed memory parallel machines -

Maki OGURA<sup>†</sup>, Tomoyuki HIROYASU<sup>††</sup>, Mitsunori MIKI<sup>††</sup>, and Michiko SUMI<sup>‡</sup>

### 1 はじめに

シミュレーテッドアニーリング(Simulated Annealing: SA)は最適化問題の手法の一つであり、多くの組合せ最適化問題に対して有効な手法である<sup>1)</sup>。SAの探索は最適解へ収束するという保証を持つが、解を得るまでの計算量が非常に多いという短所を持つ。特に連続最適化問題を対象とした場合、最適解を得るまでに多くの計算量が必要となり実用的ではない。そのため逐次処理であるSAを並列化し高速化を図る研究がされている<sup>2)</sup>。

そこで我々は、新たな並列SAアルゴリズムとして遺伝的交叉を用いた並列SA(Parallel Simulated Annealing using Genetic Crossover: PSA/GAc)を提案している。本手法は、並列に複数実行しているSAの解の伝達時に、遺伝的アルゴリズム(Ge-

netic Algorithm: GA)のオペレータである遺伝的交叉を用いたものである。

本研究では、PSA/GAcを分散メモリ型並列計算機に実装するための新たなモデルを数種考案し、それぞれのモデルについて性能の比較を行った。

### 2 遺伝的交叉を用いた

#### 並列シミュレーテッドアニーリング

遺伝的交叉を用いた並列シミュレーテッドアニーリング(PSA/GAc)は図1に示すように、複数のSAを並列に実行する。さらに、一定期間ごとのSAの解の伝達時に、GAのオペレータである遺伝的交叉を用いたものである。GAのオペレータを用いたSAであるため、SAの探索点の総数(SAの並列数)を個体数と呼ぶこととする。

PSA/GAcでは、一定期間nに達するとその時点での解を個体とし、個体をランダムに2つ選択して遺伝的交叉を行う。ここで用いる遺伝的交叉は一点の設計変数間交叉である。設計変数間交叉とは各設計変数の間でのみ交叉を行うことをいう。もとの親と生成した子との4個体間のうち評価値の高い2個体を選択して、選択された2個体から

<sup>†</sup> 同志社大学大学院  
Graduate School of Engineering, Doshisha University  
E-mail:ogu@mikilab.doshisha.ac.jp

<sup>††</sup> 同志社大学工学部  
Knowledge Engineering Dept., Doshisha University  
E-mail:tomo@is.doshisha.ac.jp

<sup>‡</sup> 同志社大学工学部(学)(現富士通(株))  
Undergraduate student, Doshisha University

次の探索を続ける。ある設計変数の最適値が求まっている場合、遺伝的交叉によってその設計変数の最適値を他の SA 探索に伝達することができるため、アニーリングの収束を早めることができる。これらの処理を繰り返し、終了条件を満たすまで探索を続ける。

本手法をいくつかのテスト関数に適用したところ、高い解探索能力を示すことを確認した。また、連続最適化問題の一つであるタンパク質の構造解析を行い、従来の手法より有効な結果を得た<sup>3)</sup>。

### 3 並列実装モデル

2 節で述べたように、遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) は高い解探索能力を示しているが、PC クラスタのような分散メモリ型並列計算機に実装する場合には新たな実装モデルが必要となる。本研究では、そのような実装モデルを数種考案した。考案した実装モデルについて以下に説明する。

#### 3.1 モデル 1 - 2 個体モデル -

モデル 1 の概念図を図 2 に示す。探索の初めに個体の集団(島)を作り、各島に 2 個ずつ個体を生成する。1 プロセッサ当たりには、全島数をプロセッサ数で分割した数分の島が存在する。交叉周期  $n$  まで各個体は SA の処理を行い、交叉周期  $n$  になると島内の 2 個体を親として設計変数間交叉を行う。交叉終了後、すべての島で同期をとり、島内の 2 個体のうちランダムに選択した 1 個体を隣の島に送信する。これを移住という。各島において、送信した個体数と受信する個体数は共に 1 であるから、各島の中には常に 2 個体である。プロセッサ内の島へ移住する場合に通信は行われないが、異なるプロセッサの島への移住にはプロセッサ間通信が必要となる。すべての島が個体情報の受信を完了した時点ですべての個体が SA の処理を再開する。この操作を温度が十分低くなるまで繰り返し、終了条件を満たせば探索を終了する。

#### 3.2 モデル 2 - 集合交叉モデル -

モデル 2 の概念図を図 3 に示す。このモデルでは交叉処理を行うプロセッサを 1 つ設け、他のプロセッサでは SA の処理を行う。SA の処理を行うプロセッサ 1 つ当たりに 1 つの島が存在する。各島には、全個体数を SA 処理を行うプロセッサ数で分割した数分の個体が存在する。交叉周期  $n$  まで各個体は SA の操作を行う。SA 処理を行っているプロセッサは、交叉周期  $n$  に達すると島に存在する 1 個体をランダムに選択し、交叉処理を行うプロセッサに送信する。

交叉処理を行うプロセッサは、受信した個体から 2 個ずつを親として設計変数間交叉を行い、生成された子の 1 個ずつを送信されてきた SA 処理を行うプロセッサに返す。子を受信したプロセッサから SA の処理を再開する。この操作を温度が十分低くなるまで繰り返し、終了条件を満たせば探索を終了する。SA の処理を行うプロセッサと交叉を行うプロセッサが異なる場合、交叉周期  $n$  に達するたびにプロセッサ間通信が必要となる。

#### 3.3 モデル 3 - 島モデル -

モデル 3 の概念図を図 4 に示す。探索の初めに島を作り、各島に個体を生成する。1 プロセッサ当たりに島は 1 つしかなく、全個体数をプロセッサ数で分割した数分の個体が各島に存在する。交叉周期  $n$  まで各個体は SA の処理を行い、交叉周期  $n$  になると島に存在する 2 個ずつをランダムに親として選択し、設計変数間交叉を行う。移住周期  $m$  までこの処理を繰り返す。移住周期  $m$  に達するとプロセッサ間で同期をとり、各島に存在する 2 個体をランダムに選択し、隣の島に移住させる。各島において、送信した個体数と受信する個体数は共に 2 であるから、各島の中に存在する個体は常に一定数である。移住の際にはプロセッサ間通信が必要となる。すべての島が個体情報の受信を完了した時点で各個体が SA の処理を再開する。これらの操作を温度が十分低くなるまで繰り返し、終了条件を満たせば探索を終了する。

### 4 数値実験

#### 4.1 実験環境

本実験で用いた IBM RS/6000SP は、分散メモリ型の超並列プロセッサ (Massively Parallel Processors : MPP) で、表 1 に示す 3 タイプのハードウェア、13 ノードから構成されている。本実験では、科学技術計算(浮動小数点演算)に適した設計である Type1 の 10 ノードを用いた。通信媒体には SP スイッチ (150MByte/Sec) を用いた。1 ノード当たり 1 プロセッサであるため、プロセッサ間通信では SP スイッチを用いたノード間通信を行うものとする。使用した OS は AIX4.2 であり、コンパイラーとして XL Fortran Compiler5.1.1 を用いた。

#### 4.2 タンパク質の構造解析

タンパク質は生命現象に直接関わる重要なものであり、構造を解明することは生命現象の仕組みを説明することにもつながる。タンパク質の立体構造はエネルギーの最小状態に対応しており、エネルギーを最小とするような構造を最適化手法を用

いて求めることで、構造解析が可能である。これまで、タンパク質の構造解析においては SA が主として使用されてきた<sup>4)</sup>。岡本らは、小規模なタンパク質 (Met-enkephalin) を対象として構造解析における SA の有効性を確かめている<sup>5)</sup>。我々は、提案手法である遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) を Met-enkephalin の解析に適用し、従来用いられていた SA よりも高い探索能力を示すことを明らかにした<sup>3)</sup>。

本実験では PSA/GAc の並列実装モデルを用いて Met-enkephalin の解析を行い、各モデルの解探索能力および計算時間の比較を行った。用いたパラメータを表 2 に示す。Met-enkephalin は 19 個の設計変数を持っており、1MCsweep によって 19 回のアニーリングの処理が行われるものとする。

#### 4.3 結果

それぞれのモデルを用いた解析を 40 回ずつ試行し、最適構造が得られた確率を Success rate とした。また、ジョブ開始からジョブ終了までを計算時間とする。

図 5、図 6 に各モデルの解探索能力、計算時間を示す。なお、モデル 1 は各島に存在する個体数が 2 である必要があるため、24 個体を用いた場合に 8 プロセッサでの計測は行えない。

図 5 では、どのプロセッサ数においてもモデル 2 の解探索能力が高いことが示されている。図 6 の計算時間の面ではモデルによる大きな差はなく、すべてのモデルが高い並列化効率を示している。

### 5 結論

本研究では、我々が提案している遺伝的交叉を用いた並列シミュレーテッドアニーリング (PSA/GAc) を分散メモリ型並列計算機に実装するモデルを数種考案し、解探索能力や計算時間の面から比較を行った。小規模なタンパク質である Met-enkephalin の構造解析に適用したところ、モデル 2 は解探索能力が高く、計算時間も少なく、最もよいふるまいをした。常に 100% 近い確率で最適構造を得ることができたため、モデル 2 は特に探索能力の面で優れていることが明らかとなった。

今後は、現在用いられている手法では解析が困難である大規模なタンパク質の構造解析にモデル 2 を適用し、良好な解が得られることを確認する。

### 謝辞

本研究の実施にあたり、岡崎国立共同研究機構の岡本祐幸先生に多大なご指導をいただいたことに深く感謝する。IBM RS/6000SP の利用にあたり、ご指導くださいました青山幸也様、吉野弘文様、越智英昭様をはじめとする IBM や通信放送機構 (TAO) の皆様に心より感謝申し上げる。

### 参考文献

- 1) Bruce E. Rosen, 中野良平. シミュレーテッドアニーリング -基礎と最新技術-. 人工知能学会誌, Vol. 9, No. 3, 1994.
- 2) E. H. L. Aarts, J. H. M. Korst. *Simulated annealing and Boltzmann machines*. John Wiley & Sons, 1989.
- 3) 小掠真貴、廣安知之、三木光範、角美智子、岡本祐幸. 遺伝的交叉を用いた並列シミュレーテッドアニーリングの検討. 情報処理学会研究報告, Vol. 2001, No. 27, pp. 57-60, 3 2001.
- 4) 木寺詔紀. コンピュータ解析 -最適化をめぐって. 蛋白質 核酸 酶素, Vol. 39, No. 7, 1994.
- 5) 岡本祐幸. モンテカルロシミュレーションで探るタンパク質の折り畳み機構. 物性研究, Vol. 70, No. 6, pp. 719-742, 7 1998.

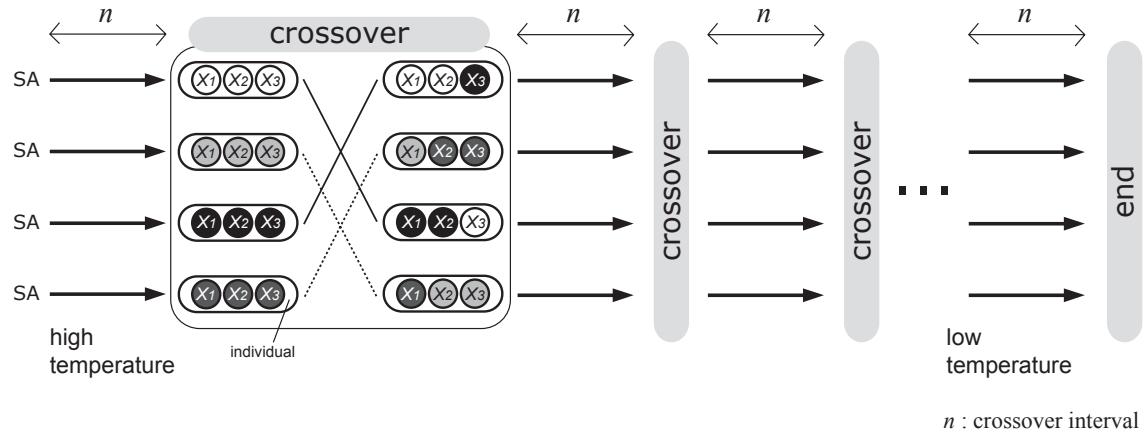


図 1: 遺伝的交叉を用いた並列 SA(PSA/GAc) の概念図

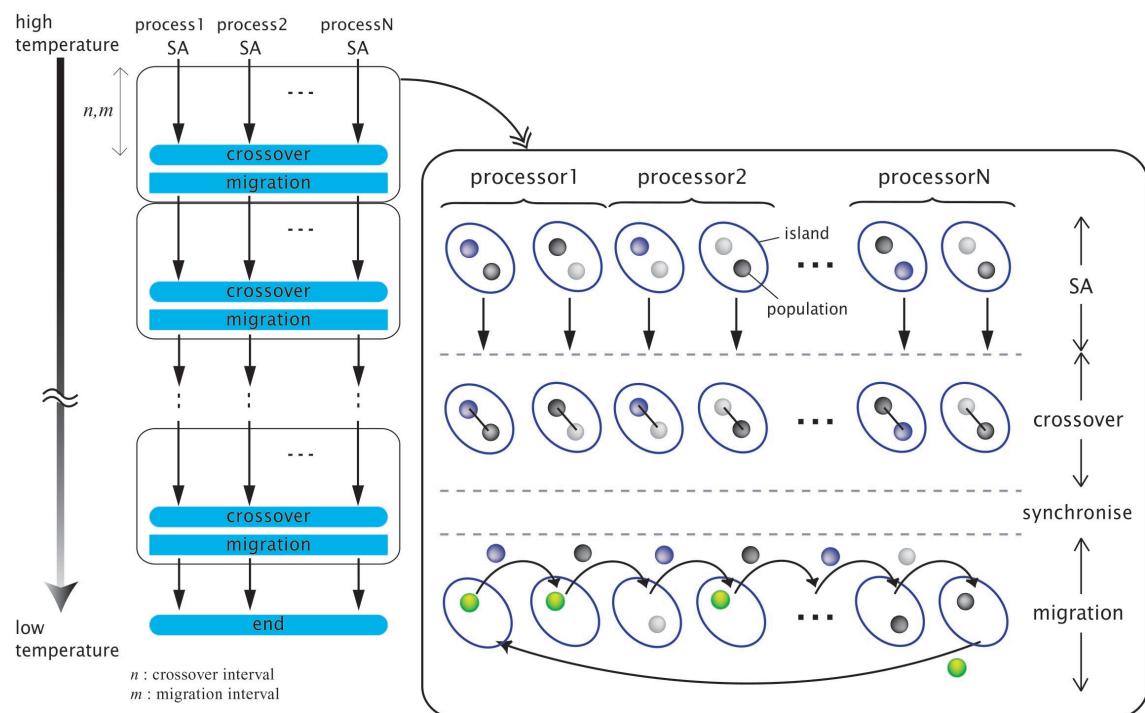


図 2: モデル 1 の概念図

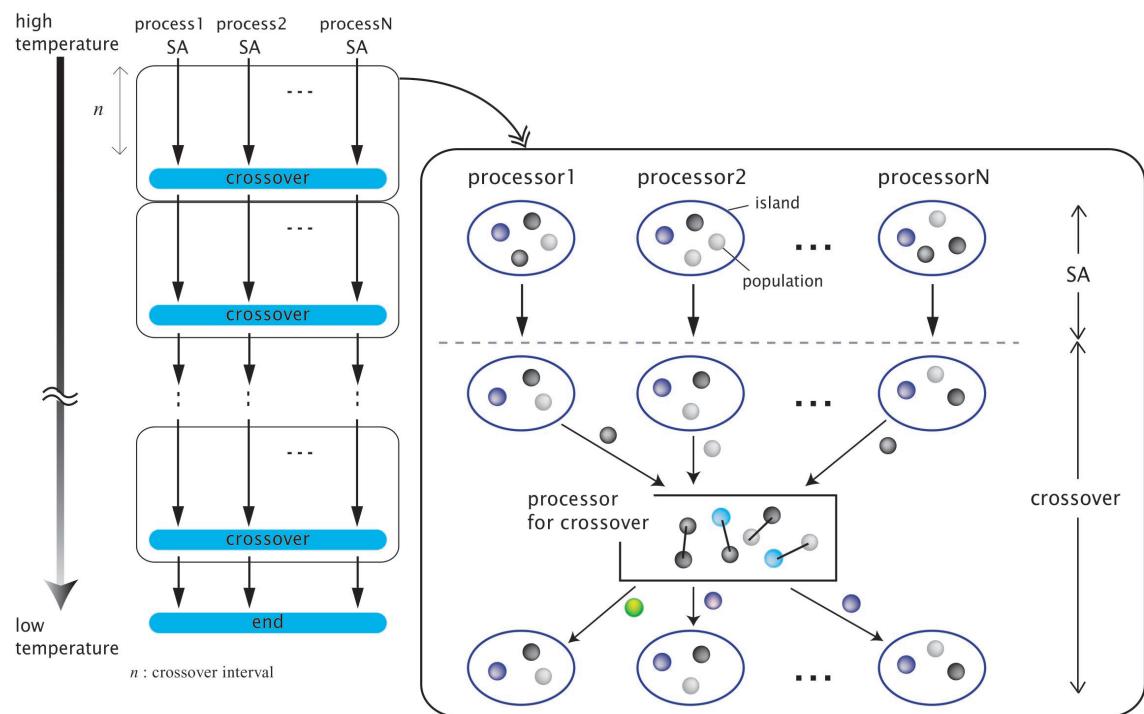


図 3: モデル 2 の概念図

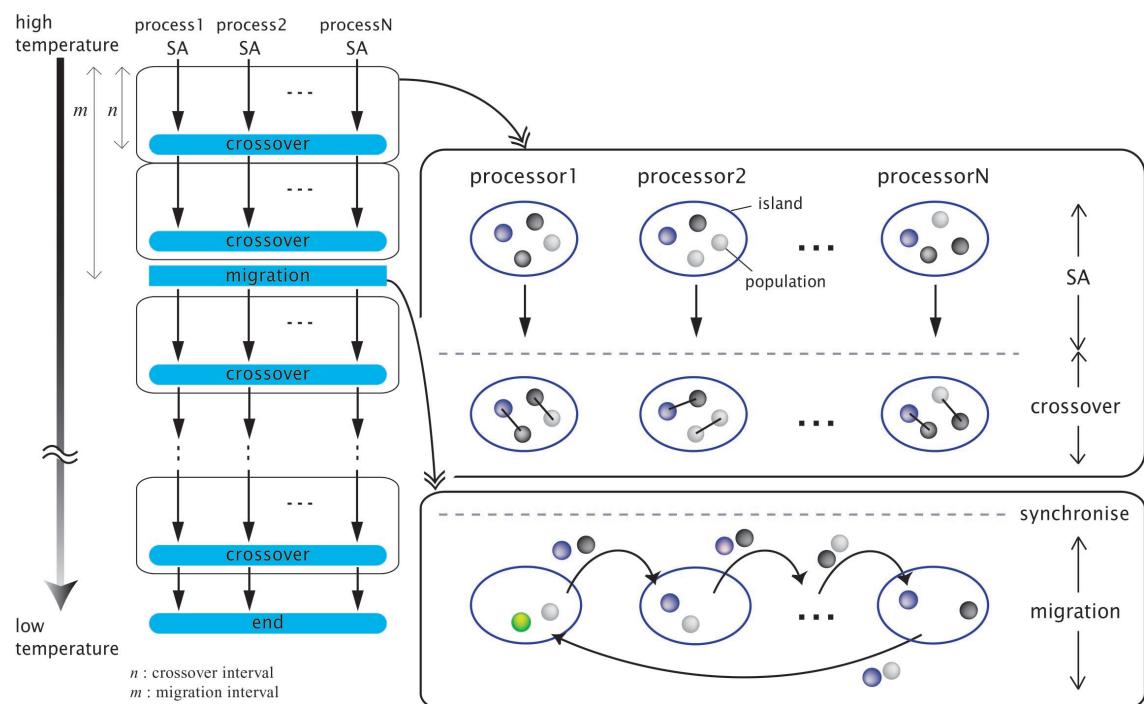


図 4: モデル 3 の概念図

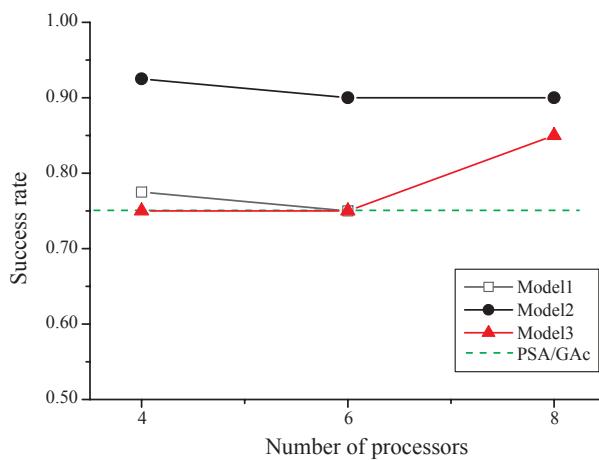


図 5: 解探索能力

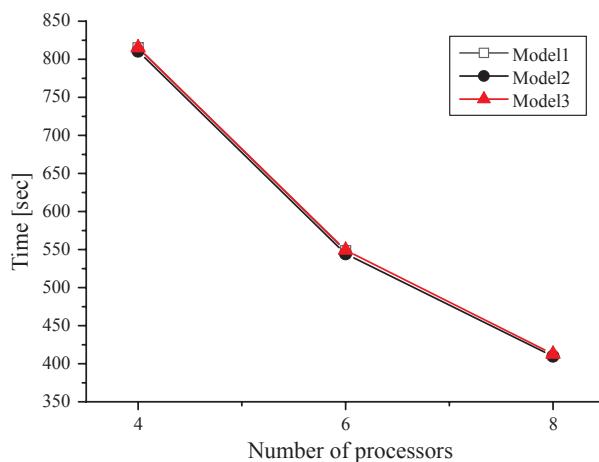


図 6: 計算時間

表 1: IBM RS/6000SP

Type	Node	CPU	Main Memory
Type1	Node4 ~ 13	Power2 SuperChip 160MHz (SPECfp 26.6)	1GB
Type2	Node2 ~ 3	PowerPC604e 233MHz 4Way SMP (SPECint95 14.0)	3GB
Type3	Node1	PowerPC604e 233MHz 2Way SMP (SPECint95 14.0)	1GB

表 2: パラメータ

Parameter	Value
Population size	24
MCsweeps	4000
Initial temperature	2.0
Cooling rate	0.999
Crossover interval	32
Migration interval	32