数 値 解 析 42－3
(1992. 8. 20)

AP1000におけるN次元通信ライブラリの試作，検討

柴田 一哉*，藤崎 正英**，金澤 宏幸**，奥田 基**

*㈱富士通青森システムエンジニアリング，
**富士通株式会社

ユーザのN次元計算モデルに対応した疑似的なプロセッサ要素空間上で，並列アプリケーションの開発ができるメッセージ・パッシングタイプの通信ライブラリを，並列計算機ＡＰ１０００で動作するN次元通信用ライブラリとして試作した。さらに，最近接参照を行う3次元格子モデルを2次元トーラスにマッピングする場合を例に，本通信ライブラリを評価した。評価関数としてＰＥ間の距離を用い，最適化手法としてシミュレーテド・アニーリング法を採用した。これによって経験的マッピングと比較して，良い結果が得られる場合があることを示した。本報告では，アプリケーション並列化の観点から本ライブラリの試作と実例による評価結果について述べる。

A TRIAL MANUFACTURE AND ITS EXAMINATION OF A MESSAGE PASSING COMMUNICATION LIBRARY ON AP1000

Kazuya Shibata* Masahide Fujisaki** Hiroyuki Kanazawa** Motoi Okuda**
*FUJITSU AOMORI SYSTEM ENGINNERING LIMITED
** FUJITSU LIMITED

We made the message passing communication library on AP1000 which users can develop applications on the pseudo processor element (PE) space just corresponding to the user's calclation model, on AP1000 as a trial manufacture. And then, we evalute this library using an example of 3 dimensional lattice model with referring to the nearest neighbor PEs, on 2 dimensional torus network of AP1000. we adopt the distance between PEs as a evaluation function, and adopt the simulated annealing method as a optimization method. Using this example, This library provides the effective performance compared with a method based on the experience. This report describes the development and eveluation of the library from the view point of the parallelization of application program.

1. はじめに

近年，並列計算機の開発が進み，並列化を行う環境が次第に整い，実用的なアプリケーションが適応可能となってきた。しかし，並列化の過程で必要となる作業が自動化されていない。特に，データ分割や，機能分割のアルゴリズム（以降，分割アルゴリズム）を考える時に，並列計算機のネットワーク・アーキテクチャが様々である事から，効率の良いプログラムを開発するためには，一般ユーザが，マシンのアーキテクチャを十分に考慮する必要がある。この様にして完成した並列アプリケーション・プログラムでは，並列計算機のアーキテクチャの変化に追いついて行けないであろう。これでは，適応性の高い並列アプリケーション・プログラムの蓄積は望めない。

一部の先進ユーザは，現状の環境でアプリケーションの高速化を達成するために，アーキテクチャを十分に考慮して高速化を行っている。一方並列計算機を普及させるには，一般ユーザに分割アルゴリズムを意識させないで，並列アプリケーションを開発させる環境が必要である。それには，使い勝手だけを考えれば，ユーザが自分のN次元計算モデルのイメージのままで並列アプリケーション・プログラムの開発が出来ればいい。また，アプリケーションの高速化という観点では，先進ユーザの計算モデルに対応したマッピングが必要になる。

そこで，アプリケーションによく見られる，データ参照パターンについての，ユーザに使い易くかつ，高速であるインターフェースが必要となる。今回は，高並列計算機ＡＰ１０００ [1]を対象に行った研究について，このユーザ・インターフェース実現の方法を述べるとともに，典型的なデータ参照パターンについての，最適なマッピング方法について，これまでに得られた成果について述べる。

2. ＡＰ１０００について

並列計算機ＡＰ１０００について概要を説明する。

ＡＰ１０００は，各プロセッサが，2次元トーラス状に接続されているＭＩＭＤ (Multi Instruction stream Multi Data stream) 型の並列コンピュータである。特徴として，３つの通信ネットワークを装備していて，ＰＥ間の通信はトーラスネットワークを用いる。ＰＥ間通信の最適化にあたっては，ネットワークの特性を考慮しなければならない。

図１にＡＰ１０００のアーキテクチャ構成図を示す。

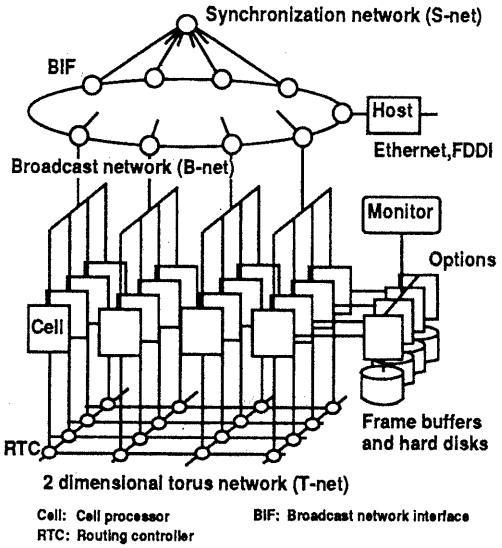

図１　ＡＰ１０００のアーキテクチャ構成図

3. ユーザ空間の重要性

ユーザ・インターフェースの必要性

ユーザが計算させようとする計算モデルは様々であるが，適応する問題には独特のデータ参照パターンが存在していて，さらに計算モデル構成が３次元以上の高次元であるものが多い。これを反映させる必要がある。

これらを踏まえ，ユーザに，アーキテクチャを意識させず，計算モデルのイメージのままで作業させるためには，あたかも，物理プロセサ構成がユーザの計算モデルと同じ構成をしているかのごとく，ユーザにイメージさせて，その空間のイメージのままでプロセッサ間通信をさせるインター

フェースを定義しなければならない。

ユーザ・マッピングの必要性

最適なマッピングとは，アプリケーションのデータ参照パターンに対し，最も効率的なプロセッサ間通信を実現するものである。

本来，ユーザ空間から，プロセッサ空間へのマッピング作業は，自動的に最適な形が選ばれるのが理想である。しかし，アプリケーション独自のデータ参照パターンは，ユーザしか知らないもので，これまで多くの並列計算機システムでは，満足のいくマッピングを行ってはいなかった。また，最適なマッピング手法自体分からない部分が多い。そこで，まず各通信パターンに対応した，最適なマッピング規則を見つけることに目標をおいた。そのためには，ユーザが自由にマッピング規則を操作出来るものが必要になる。

4. インターフェースの試作

以上のような，ユーザ・インターフェースの必要性から，ユーザがN次元ユーザ空間のイメージで，並列アプリケーション・プログラムを開発することができ，システムではなくユーザが自由にマッピング規則を操作できるものを試作した。

実現方法の概要は，図2のように，最初に，ユーザが分割したN次元モデルの計算ユニットと，物理プロセッサに通し番号をつけておき，その対応をアドレス変換テーブルでもつ方法である。すると，これにユーザがアクセスするだけで，自由にマッピング操作をすることが可能になる。その後は，プロセッサ間通信はユーザが分割ユニットの通し番号を用いて行うことが出来る。これで，ユーザ空間での作業が実現し，N次元計算モデルに応じてマッピングを自由に操作できるようになる。これを，従来のライブラリ群に加えて，N次元サブルーチン・ライブラリとして拡張，実現した。

5. 計算モデルに対応したマッピング

データの参照パターン

データ参照パターンは，アプリケーションによってさまざまである。最初から全ての場合について考えるのは無理であろう。そこで，アプリケーションによく見られるデータ参照パターンによって分類してみることにする。

・最近接格子：

応用例

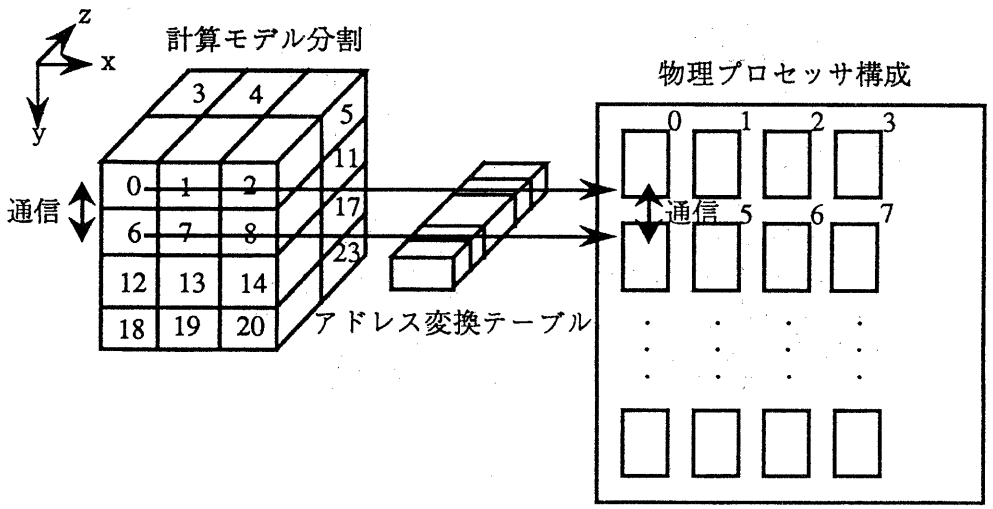

図2　ユーザ・インターフェース

構造解析，熱伝動，流体解析
（有限要素法，中心差分）
物性（イジングスピン）
ＭＤ

・近接格子：
応用例
ＱＣＤ

・N対N：全ての分割ユニットについて
応用例
希薄流体

・完全独立：通信が無い

計算モデルの次元数

データの参照は，問題が何次元であるかによっても考え方が異なってくる。例えば図３の様に，データ参照が最近接方向で２次元の計算モデルを，ＡＰ１０００のような２次元トーラスにマッピングする場合は，そのままマッピングすればいいが，３次元の計算モデルの場合だと，途中で変換が必要になる。計算モデルの次元が増える程，各計算ユニットが通信する方向は増えてゆき，高次元の計算モデルになるにつれて，難しくなることが想像される。例えば，３次元の最近接格子の場合は通信方向が６方向で，４次元だと８方向である。これを２次元にマッピングするには，残りの方向に当たる部分をどこかに持ってゆかなければならない。

今回は，ＡＰ１０００（２次元トーラス）での３次元の最近接格子の通信の最適化問題を考えることにする。これは，アプリケーションで最も応用範囲の広いケースである。

6. 通信コストの評価関数

ＡＰ１０００（２次元トーラス）でのマッピングを考える前に，通信の評価基準として評価関数を決めなければならない。

一般的に，並列計算機で通信コスト（時間）の削減を考える時に，通信回数，通信量，通信の混

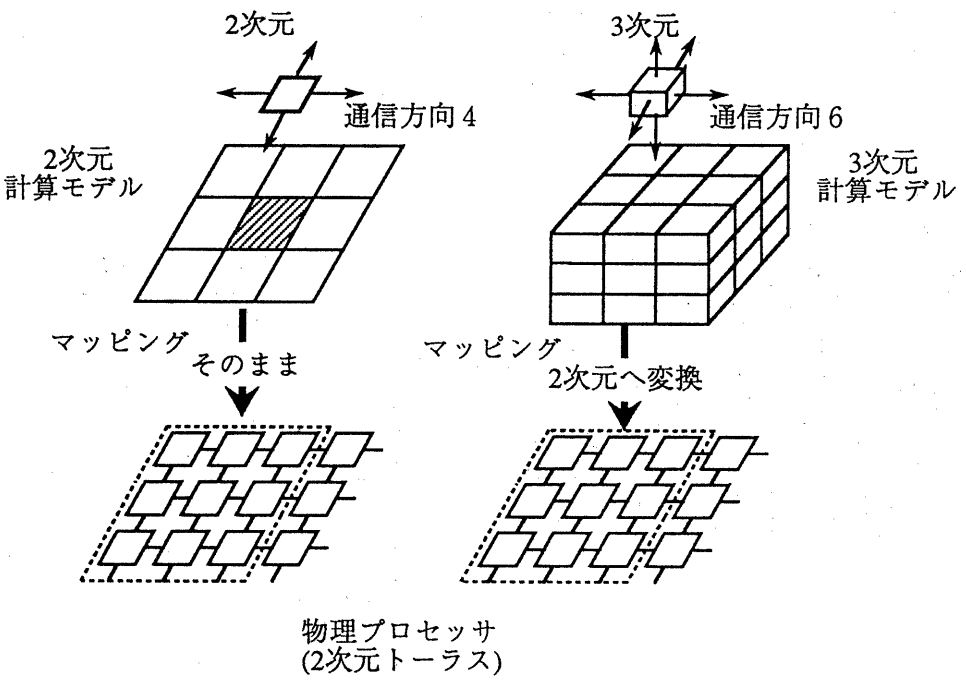

図３　モデルの次元数によるマッピング方針

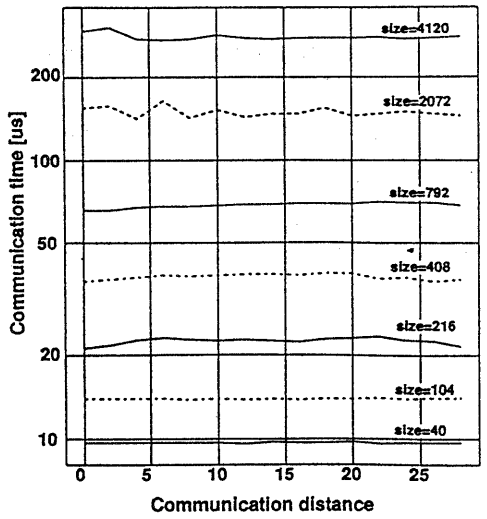

図4　AP1000に於けるPE間距離対通信時間

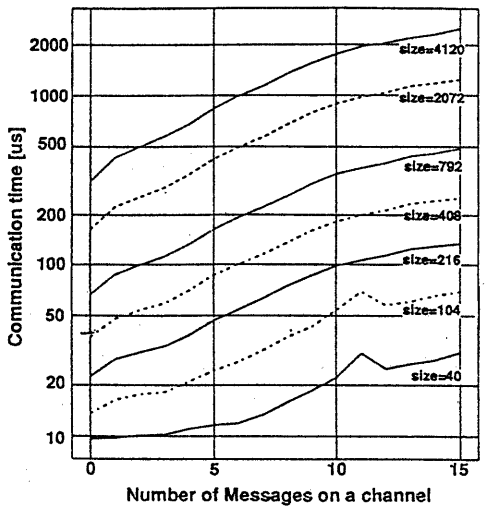

図5　AP1000に於ける通信ｺﾝﾃﾝｼｮﾝ 対通信時間

み方，ＰＥ間距離を考えねばならない。どれが，どのくらい通信コストに，効いてくるかの比率は，その並列計算機のアーキテクチャに大きく依存している。ＡＰ１０００の場合は，プロセッサ間通信コストは，図4，図5に示される様に，[2] ＰＥ間の距離にはあまり依存せず，通信の混み方に大きく依存している。

そこで，ＡＰ１０００の場合，今回の評価では，距離が最小であるなら通信の衝突も起こりにくく混み方が少なくなり，通信コストが最小に近い値が得られるだろうと仮定した。評価関数として各プロセッサ間の距離の総和を導入した。これについては，今後検討の余地がある。

7. マッピング

今回は，サンプルとして，3^3 の物理体系を，8^2 の2次元ネットワーク構成の計算機に，4^3 を8^2 にマッピングする場合について考える。しかし，全ての組み合わせはそれぞれ，10^{48}通り，10^{90}通り存在するため，これら全てを調べるのは無理である。そこで，経験的に規則的な並びを考える方法と，シミュレーションにより求める方法を行った。

経験的マッピング

3次元の計算ユニットの，6つの通信方向全てに対し均等な通信コストは考えにくいので，まず，ある方向を，優先的に隣になるようにおき，残りの方向についてなるべく近く，規則的になるように置くことを考えた。

ＸＹＺ各方向に対して，立法体に分割し，それをマッピングするモデルを考える。

図6は3^3 を8^2 にマッピングした場合

図7は4^3 を8^2 にマッピングした場合

図6，7に於いて，

Ａのマッピング方法は，ＺＸ平面で4つに切っておき，Ｙ方向の各計算ユニットが正方形で隣合うように並べる方法で，ＺＸ方向の通信がひとつ飛びに並ぶ恰好になる。しかし，Ｙ方向が隣あう為には，ＺＸ平面を4つに分割しなければならない。

Ｂのマッピング方法は，ＺＸ平面でスライスしたものを順番に置いていったものである。これで，ＺＸ方向が隣合いＹ方向が均等（スライス数分離れる）に並ぶ。

シミュレーションによるマッピング

次に，シミュレーションによって最適なマッピ

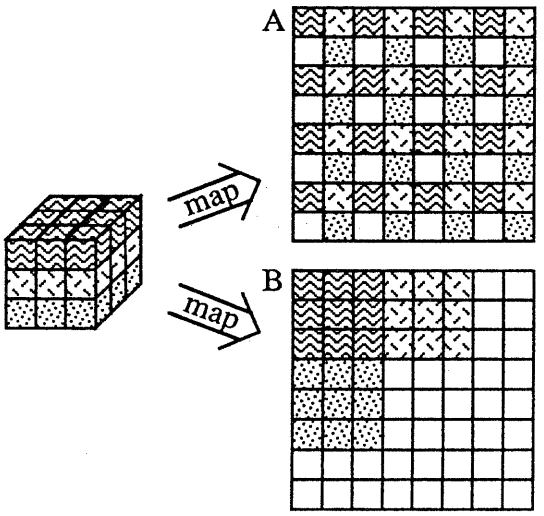

図6　3^3 を 8^2 にマッピングした場合

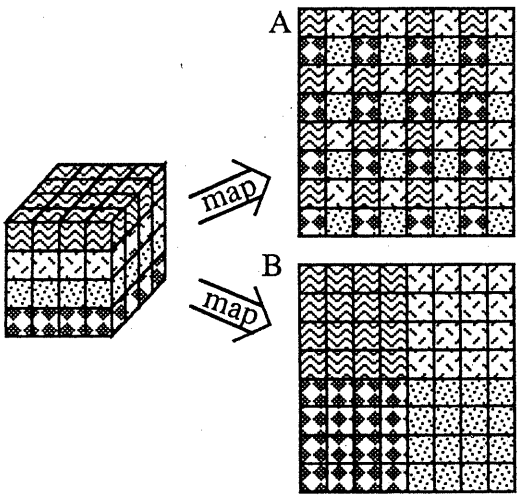

図7　4^3 を 8^2 にマッピングした場合

ングを求めようという方法を用いた。つまり，今回のマッピングの問題を距離の総和を最小化する組合わせの最適化問題として考えた。

マッピングによる通信コストは，すべての最近接間格子間の距離の総和にほかならないとした。

方法は，初期状態としてランダムにマッピングしておき，経路の総和を求めておく。どこか一組を取り替えてみて，経路の総和が減ったら交換を採用する。交換を繰り返していき，総和の変化が少なくなった状態を近似解として採用する方法である。

具体的には，隣接格子点P_i P_jがそれぞれ2次元格子点上のある点

(X_i, Y_i) (X_j, Y_j) にマッピングされたとする。P_i P_jの距離l_{ij}は

$$l_{ij}=|X_i-X_j|+|Y_i-Y_j|$$

よって評価関数は $L=\Sigma l_{ij}$になりこれが最小になった時の，マッピング状態を最適なものと近似する。（最小なのは全ての格子点が一つになったときであるが，交換で格子点が重なることは考えない）ところが，このような緩和法[3]には，局所的最小の存在が確認されており，ΔLが最小だからといって必ずしも，通信コストが最小であるとは限らない。

アニーリング法

そこで，組合わせの最適化の近似解法として，カークパトリックら（S. Kirkpatrick et al. 1983）によって提案された，シミュレーテドアニーリング法（simulated annealing method）[4]を用いた。これは，評価関数をより小さくするように変化させていく過程に，局所的最小を脱出するような確率を導入した方法である。その確率をω，目安をT（温度）とする。

$\omega(\Delta L)=\exp(-\Delta L/T)$，$\Delta L>0$

$\omega(\Delta L)=1$　　，$\Delta L<0$

ΔL：距離の変化

T（温度）を段々下げてゆく（確率を下げてゆく）と評価関数が最小に近くなる。

Tの下げかた（アニーリング・スケジュール）は，ギーマン兄弟（S. Geman and D. Geman 1984）による，

$T(t)=B/\ln t$

$T(t)\rightarrow 0$　（$t\rightarrow$無限大）とした。

B：評価関数の障壁の高さ

図8に，アニーリング法のアルゴリズムを示す。

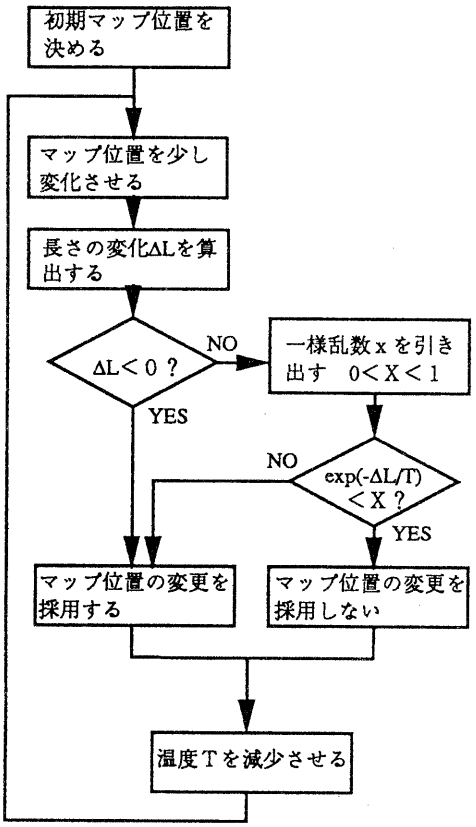

図８　アニーリング法アルゴリズム

結果

これによって得られた，マッピング結果，$3^3 \Rightarrow 8^2$ の場合（B=1.0,1Mstep）を図９，$4^3 \Rightarrow 8^2$ の場合（B=1.0,1Mstep）を図１０に示す。（図中の番号はＸＹＺ各座標を表す）

図９に於いて，一は３次元で隣合っているペアを結んでいる。4^3 の場合は並び方に殆ど規則性が見られない。

8. マッピングの評価

シミュレーションで，得られたマッピングと経験的に得られたマッピングとを距離の総和（Ｌ）で比較してみた。（表１参照）

$3^3 \Rightarrow 8^2$ の場合は，経験的に得られたマッピングより，本手法でのマッピングの方が距離の総和が少なく良い結果が得られた。ただし，最良なマッピングであるとは限らない。　$4^3 \Rightarrow 8^2$ の場合は，経験的に得られたマッピングの方が良い結果であった。

本最適化手法では，常に最良の結果が得られるとは限らないため，今後の改善が必要である。

9. 結論

評価関数として距離をとり，３次元格子の最近接参照モデルを２次元のトーラス．ネットワークにマッピングする時，アニーリング法を用いる事で，結果を比較的容易に経験的マッピングより良い結果が得られる場合があることを示した。今後，評価関数を正確化し改良することで，経験的手法が働かない様な，高次元の問題のマッピングを考えるのに有効と考えられる。

10. まとめ

適応性の高いアルゴリズムを考えるために，ユーザが分割アルゴリズムを意識せずに作業出来るインターフェースを，ＡＰ１０００上で動作する，Ｎ次元通信ライブラリとして試作した。また，その可能性を示すために，高次元のマッピングを考えるためのひとつの手法として，３次元計算モデルを例にとり，アニーリング法を用いて評価した。今後，実用アプリケーションの経験に基づいて，よく使われるデータ参照関係について評価関数を正確化し，最適化手法の改善を行う事で，適応性の高い分割アルゴリズムを開発してゆく予定である。

表1　距離の総和（L）による比較

手法	マッピングパターン	系	コストL
頭で考えたマッピング	A	$3^3 \rightarrow 8^2$	180
		$4^3 \rightarrow 8^2$	320
	B	$3^3 \rightarrow 8^2$	180
		$4^3 \rightarrow 8^2$	336
アニーリング	C	$3^3 \rightarrow 8^2$	154
	D	$4^3 \rightarrow 8^2$	380

C

...	022	...	...	...	...	...	...
021	222	122	020	120	121	220	221
011	212	112	010	110	111	210	211
001	012	102	000	100	101	200	201
...	202	002	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...

図9　$3^3 \Rightarrow 8^2$ にマッピングする場合

D

201	302	312	211	012	111	101	002
202	213	212	113	112	103	102	203
232	332	322	222	022	122	132	032
230	231	223	221	123	121	131	233
333	323	220	023	120	133	130	033
330	331	320	321	020	021	030	031
300	313	310	013	010	003	000	303
200	301	210	311	110	011	100	001

図10　$4^3 \Rightarrow 8^2$ にマッピングする場合

謝辞

サブルーチンライブラリ作成に御指導頂いた，富士通研究所　池坂主任研究員に感謝します。

参考文献

[1] 石畑，稲野，堀江，清水，加藤，高並列計算機CAP-Ⅱ構成とメモリシステム，第1回SWooP計算機アーキテクチャ，P217-222，1990年 7月

[2] Hiroaki Ishihata, Takeshi Horie, Satoshi Inano, Toshiyuki Shimizu, Sadayuki Kato, and Morio Ikesaka : Third Generation Message Passing Computer AP1000: International Symposium on Supercomputing , june 1991

[3] 茨木俊秀：「組合せ最適化，分枝限定法を中心として」産業図書，1983

[4] Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. Optimaization by simulated annealing, Science, 220(1983) 671-680.