

CT3-1

スーパーコンピュータ・コンテスト 2002
Supercomputer Contest 2002松田裕幸[†] 太田元規[†]
MATSUDA Yuko OTA Motonori

1. コンテストの概要

1995年から東京工業大学では学術国際情報センター(当時は総合情報処理センター)が所有するスーパーコンピュータ(以下、「スパコン」と呼ぶ)を高校生に使わせたプログラミングコンテストを実施してきた。普段パソコン等でプログラムを組むことはあってもスパコンは別世界であり、その性能を最大限に発揮するプログラムを書くにはかなりの熟練した知識が必要であり、高校生には無理ではないかという声も当初はあった。しかし、今年で第8回を迎える本コンテストでは、過去、すぐれたアイデアがいくつか生まれてきたことも事実である。

本コンテストの当初の目的は、大学が所有するスパコンを高校生に使わせることで、理工系分野への興味を持ってもらう、あるいは、大学そのものへの興味(コンテストは実際に大学にて実施)を持ってもらうことにあった。しかし、現在では高度なプログラミング技術を競うことを第一義とするコンテストへと姿を変えてきている。それは、同様のコンテストが日本国内のみならず海外でもほとんどないことが一番の理由になっている。甲子園大会のようにスポーツの世界で優劣を競う高校生がいるならば、プログラミングの世界でその腕を競う場があってもよいはずである。

コンテストは3人以下で1チームを構成する。予選を行い、大学で行われる本選には10チームが参加できる。過去予選への参加は平均30チーム程度である。参加地域は北海道、沖縄を除くほぼ日本全国にまたがっている。また、1回だけであるが、海外からの参加を募ったこともあった。この時は、インターネットを利用した。

コンテストでは、当初CRAY C916を使用し、高速化は2重ループの内部におけるベクトル化と外側の並列化が中心であった。この手のプログラミングはある意味では慣れれば容易であった。これに対し、1999年の第6回大会以降リプレイスされたSGI Origin2000でのMPI並列ライブラリーを使った通信をベースにした並列プログラミングは、残念ながら高校生にはかなり荷が重たいようである。ただこれも事前の講習等でなんとかクリアできるのではないかと考えている。実際、今年は事前にMPIプログラミングを体験できる環境を用意し、インターネットでアクセスできるようした。

2. 過去の本選課題の紹介

過去にどんな問題が出たか簡単に紹介する。第1回大会では、「石運び問題」という名前で、ナップザック暗号問題について挑戦してもらった。与えられた石の重さが公開鍵に相当し、そこから秘密鍵を求める問題となっている。優勝チームはこれを総当りの力づくで解くアプローチをとったにも関わらず、ハッシュキーの中に石の組合せを埋め込んでしまうというウルトラテクニクを駆使し、一瞬にして問題を解いてしまった。驚きと共に、第1回コンテストは終了した。第2回大会は「カフェテリア問題」。いわゆる待ち行列問題である。並列化が難しいだけでなく、問題の仕様自体にも不備があり、この時は、逐次プログラムでの競争となった教訓深い大会であった。この大会をきっかけに、問題として、入力、出力サイズが小さく(その分、仕様に対する制約等の条件が少なくなる)、かつ、計算自体は入力の長さ(大きさ)Nに対し、指数的に増大するような問題を作るよう意識するようになった。第3回大会では「三角ビリヤード」という奇妙な問題を考えた。この問題の背景には、箱の中で自由運動する粒子の軌道を求めるという、純粋に物理学上の要請が元となっている。第2回大会の教訓があったにも関わらず、浮動少数点をあつかうために生じる誤差の問題、かつ、すべての解を求めよという制約のために、高校生たちはベクトル化、並列化には相当苦勞していたようである。第4回大会では、一転して「板詰め問題」と非常にとっつきやすい問題を選んだ。同じ形を持たない矩形(問題では長方形に限定)で領域を隙間なく埋めるという問題は日常、いたるところで経験する。全解出力を強制しなかったため、勝敗は入力例とヒューリスティックスの選び方とに依存するという問題点を残した。第5回大会はライフゲームの変形である。最初にライフゲームの規則とは異なる規則によってセルの初期配置を準備する。その後、ライフゲームの規則に従いセル配置を変化させる。Nが与えられた時、Nステップ後、1つ以上のセルが生きているために必要な初期配列を求める。この時は、素直に並列化、ベクトル化にすぐれたチームが上位に入賞した。第6回大会は、「配管問題」。Steiner treeのアイデアを利用して、

[†]東京工業大学学術国際情報センター, GSIC

与えられた接続点以外のポイントを追加することで最小の配管接続を求める。新たな節の導入が許される点が *minimum spanning tree* とは異なり、回路設計、道路網問題、DNA の進化過程予測モデル (ミッシングリンクの発見) など多数の興味深い応用が知られている。MPI による並列プログラミングを始めて導入した年で、並列化プログラミングに相当苦労していた。その分、興味深い解法を発見するに至る余裕はまったくなかったようである。この年は事前に本選課題を公開していたため、インターネット等を駆使し多くの情報を集めたチームが続出したが、逆にそれらの情報に引張られ、新しいアイデアが出なかったとも考えられる。しかし真相は不明である。第 7 回大会は「N 体問題」である。この時は、並列化によって優劣が出るよう、事前に逐次版のプログラムを公開した。文字通り並列化の競争となった。優勝チームは元の逐次プログラムを 3 倍に高速化し、32 プロセッサに対し 32 倍の並列化を実現し、ほぼ元の 100 倍の性能に達した。これはこれで見事な結果である。

3. 2002 年度の問題 「格子タンパク質の基底状態探索」

タンパク質は特異的な (ユニークな) 立体構造をとることによってその機能を発現する。特異的な立体構造はタンパク質の配列によってのみ決まる。ここではエネルギー最低の状態 (基底状態) だとして、タンパク質の理論モデルとしては H (疎水基) と P (親水基) からなる配列 (HP 配列) がよく利用される。例えば、

HPHHPHPHHPHP

といった配列である。こういった配列が二次元格子の上を運動して、基底状態に辿り着き、働く (生理学的機能をもった) 分子となることを想定する。このモデルをタンパク質の二次元格子模型と言う。HP からなるタンパク質では、HH, HP, PP の 3 組の相互作用が考えられるが、HH の隣り合った接触 (HH 接触) のみエネルギーをカウントする、つまり HH 接触が最大となる形がタンパクの基底状態となる。例の HP 配列では、図 1 のような形で HH 接触数が 5 となり、基底状態を与える。図 2 の点線は HH 接触を表している。

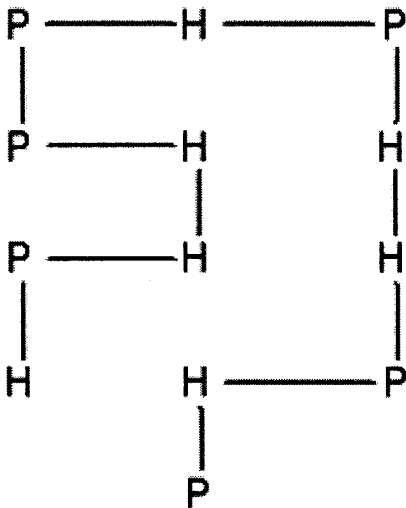


図 1 HP 格子

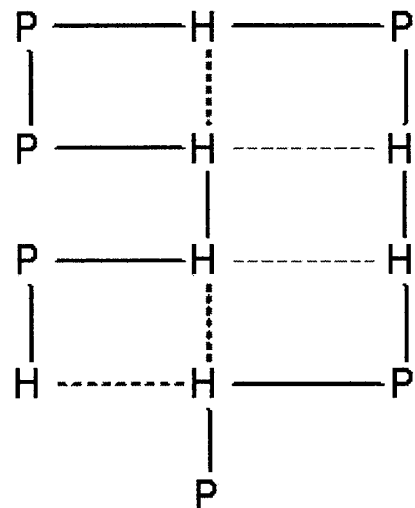


図 2 HH 接触 (点線部分)

与えられた HP 配列から、HH 接触数を最大にする形を求めるのが 2002 年度の問題である。

4. 資料

コンテストに関する情報は以下の URL から得ることができる。

<http://www.gsic.titech.ac.jp/supercon/>

また、次の雑誌にコンテストの報告が掲載されている。

- 『bit』 (共立出版) Vol27 No12, Vol28 No11, Vol29 No12
- 『数学セミナー』 (日本評論社) 1995.12, 1996.12, 1997.12, 1998.12, 1999.12, 2000.12, 2001.12