

C-034

FPGAによる動的再合成可能ハードウェアを用いた相同性検索専用計算機 A special-purpose computer for similarity search by the dynamic reconfigurable system

杉江 崇繁[†] 下馬場 朋禄[†]
Takashige Sugie Tomoyoshi Shimobaba

戎崎 俊一[†] 伊藤 智義[‡]
Toshikazu Ebisuzaki Tomoyoshi Ito

1. まえがき

バイオインフォマティクスにおいて重要な解析の一つに相同性検索がある。データベースに蓄積されたDNA配列やアミノ酸配列から、ある配列に類似した部分を検索するものである。相同性検索には一般にBLAST法[1][2][3]が用いられるが、類似配列を見落す可能性があるという欠点がある。これに対し、Smith-Waterman法[4]は完全に類似配列を検索することが可能であるが、BLAST法に対して計算量が非常に大きい。2つの配列間におけるSmith-Waterman法を用いた相同性検索では配列長を N とすると、計算量は $O(N^2)$ となる。例えばヒトのDNAは約30億の塩基から構成されていることを考慮すると、その検索時間は非現実的になる。

これまでに、我々はSmith-Waterman法を用いた相同性検索専用計算機Bioler[5][6]を開発してきた。特にBioler-2ではPentium4 3.2GHzに対して90倍の高速化に成功した。ここで、専用計算機とはある特定の回路で構成された計算機であり、汎用的な演算は一切行えない。しかし、その効果は絶大であり、これまでに重力多体問題専用計算機GRAPE[7]、ホログラフィ専用計算機HORN[8]などが開発されている。これらの専用計算機は論理ゲートで構成されていたり、ASIC(Application Specific Integrated Circuit)を用いていたりするが、FPGA(Field Programmable Gate Array)を使用しているものもある。我々の相同性検索専用計算機でも採用しているFPGAは、ユーザが内部回路を自由に構成することができ、また、複数の回路を再構成することができる。一般に、回路を構成するデータはPROM(Programmable Read Only Memory)に格納されているため、再構成するにはPROMのデータを交換しなければならない。本論文ではこの問題を回避することで、利便性と高速化を実現したBioler-3について報告する。

2. ハードウェアデザイン

ASICよりも柔軟性に優れているFPGAに注目してみると、一般に回路の再構成は可能であるがPROMデータの交換が必要であるという特徴を持つ。ここで、FPGAの回路構成に必要なデータをPROMに蓄えておくのではなく、ホストコンピュータからそのデータを直接通信することによりFPGAの回路を再構成することを考える。このときの通信路にはPCI(Peripheral Component Interconnect)バスを用いることにする。

ホストコンピュータからデータの通信を行うために、Bioler-3はPCIデバイスとして動作している必要がある。つまり、常にホスト側からの要求に応答しなければならないため、回路の動的再構成を禁止したFPGAや

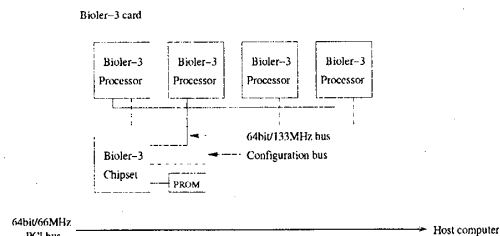


図1: Bioler-3 システム図。

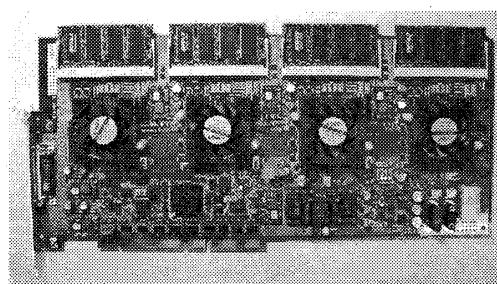


図2: Bioler-3 概観図。

ASICなどのデバイスを使用しなければならない。我々はこのデバイスのために、PROMからコンフィギュレーションデータを読み込む静的なFPGAを1つ用意した。図1に示すBioler-3チップセットがそれである。続いて、Bioler-3チップセットに演算用のFPGAであるBioler-3プロセッサを図1のように4つ配置する。Bioler-3プロセッサは動的再構成可能なFPGAであり、各々にPROMは接続されていない。Bioler-3チップセットはPCIバスから送られてきたコンフィギュレーションデータを用いて、適切なBioler-3プロセッサをコンフィギュレーションする。以上のような構成にすることで、Bioler-3プロセッサが動作、非動作にかかわらずホストコンピュータに対してBioler-3は動作し続けられる。したがって、Bioler-3プロセッサは任意の時に任意の回路に変更することが可能である。

図2は開発したBioler-3の概観図である。冷却ファンが取り付けられたBioler-3プロセッサを4つ搭載しており、それぞれにDDR-SDRAM(Double Data Rate-Synchronous Dynamic Random Access Memory)256Mbyteが接続されている。Bioler-3プロセッサにはXilinx社製XC2VP70-5FF1517C(700万ゲート相当)を使用している。隣接するBioler-3プロセッサ間は最大144bit/133MHzのLVTTTL(Low Voltage Transistor Transistor Logic)信号で接続されている。Bioler-3チップセットにはXilinx社製XC2V1000-5FG456C(100

[†]理化学研究所, RIKEN

[‡]千葉大学, Chiba Univ.

万ゲート相当)を使用している。Bioler-3 チップセットから各々の Bioler-3 プロセッサのコンフィギュレーションポートに配線されており、Bioler-3 チップセットからコンフィギュレーションデータを直接転送するようになっている。Bioler-3 プロセッサと Bioler-3 チップセット間は 64bit/133MHz バスで接続されている。外部とのインターフェイスは 32/64bit, 33/66MHz PCI バス接続と LVDS (Low Voltage Differential Signaling) 信号による 18bit バスがある。

3. アプリケーション

我々はバイオインフォマティクスにおける相同性検索専用計算機を開発してきた。Bioler-1 ではアミノ酸配列の類似度検索を行うことができる。Bioler-2 では配列データを核酸データで扱うことで、アミノ酸配列と核酸配列の両配列の類似度検索に対応している。ところが Bioler-2 では演算回路はアミノ酸配列と核酸配列のどちらも同一の回路を用いている。20 種類あるアミノ酸が計算可能であれば 4 種類の核酸も計算可能だからである。ここで、各々の回路を用意することができ、FPGA 内部の回路を動的に変更できれば、より高いパフォーマンスを期待できる。また、演算に使用するレジスタの語長を変更することで、低精度で高速演算したり、より高精度で演算したりすることも可能である。例えば、EDA (Electronic Design Automation) ツールに ISE6.2i sp3 を用いて、以下の結果が得られた。アミノ酸配列における類似度得点に 16bit を用いた場合、1 チップにパイプラインを 100 段配置でき、133MHz で動作する。一方、アミノ酸配列における類似度得点に 24bit を用いた場合、1 チップにパイプラインを 64 段配置でき、100MHz で動作する。このようにレジスタの語長を 8bit 拡張することで精度が向上するが、演算性能は半減する。このように要求される問題に対して最適な回路を選択することが可能である。また、遺伝子情報処理は相同性検索以外の解析も必要であることから、他の解析回路に変更することも有効な手段の 1 つである。

コンフィギュレーション用の PROM が不要であるため、固定パラメータ等の変更によるアップデートも容易である。これは非常に重要なことで、バグフィックスにも柔軟に対応できることを示唆する。ASIC などでは開発のミスが許されず、開発期間、コストともに非常に大きくなり、回路変更の場合にはチップ自体の交換に陥る。一般の FPGA が搭載されたシステムにおいても、PROM を書き換えなければならず、Bioler-3 は保守性、利便性の点で非常に優れている。以上のように、演算能力を維持しながら、アプリケーションの自由度を大幅に広げることが可能である。

4. まとめ

Bioler-3 は Linux-2.4 または Linux-2.6 がインストールされたホストコンピュータ上で動作する。演算可能な最大配列長は 15,000 程度であり、基本動作周波数は 133MHz である。10,000 配列長を持つ生物配列の Global-Global における類似度計算に 2ms を要した。これに対して汎用計算機である Pentium4 3.2GHz では Linux-2.4.26, gcc-3.4.1 を用いて 810ms を要した。したがって、約 400 倍

の高速化を達成することができた。

今後は動的再構成可能システムである Bioler-3 の特徴を有効に活用し、演算の高速化に活かせる技術を模索していきたい。

参考文献

- [1] Altschul S.F., Gish W., Miller W., Myers E.W. and Lipman D.J., A basic local alignment search tool, *Journal of Molecular Biology* **215** (1990) 403-410.
- [2] Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W. and Lipman D.J., Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res.* **25**(17) (1997) 3389-3402.
- [3] Zhang Z., Schaffer A.A., Miller W., Madden T.L., Lipman D.J., Koonin E.V. and Altschul S.F., Protein sequence similarity searches using patterns as seeds, *Nucleic Acids Res.* **26**(17) (1998) 3986-3990.
- [4] Smith T.F. and Waterman M.F., Identification of common molecular subsequences, *Journal of Molecular Biology* **147** (1981) 195-197.
- [5] T. Sugie, K. Horita, S. Hosono, T. Ito, T. Ebisuzaki, A special-purpose computer for exploring similar protein sequences by the dynamic programming method, *Computer Physics Communications* **157**(1) (2004) 1-8.
- [6] T. Sugie, T. Ito, T. Ebisuzaki, A special-purpose computer for exploring similar biological sequences: Bioler-2 with multi-pipeline and multi-sequence architecture, *Computer Physics Communications in print*.
- [7] Makino, J., Fukushige, T., Koga, M., and Namura, K., GRAPE-6: The massively-parallel special-purpose computer for astrophysical particle simulations, *Publications of the Astronomical Society of Japan* **55** (2003) 1163-1187.
- [8] T. Shimobaba, S. Hishinuma, and T. Ito, Special-purpose computer for holography HORN-4 with recurrence algorithm, *Computer Physics Communications* **148** (2002) 160-170.