

# ストリームデータプロセッサ SDP (1)

3N-9

— 基本アーキテクチャ —

早川 佳宏 伊藤 正雄 田村 登 上田 謙一  
松下電器産業(株) 東京研究所

## 1. はじめに

人工知能システム、データベースシステムにおける処理の一例として検索処理があり、これらの高速化を目的としたハードウェアの開発もさかんに行われている。

我々は、上記分野への応用を目的として大量のデータを対象とする全数検索処理を専用に行うストリームデータプロセッサSDP(Stream Data Processor)の開発に着手した。

本稿ではSDPの基本思想を述べ、データの扱い方を中心に本プロセッサのハードウェアアーキテクチャ、仕様及び目標性能の検討結果について報告し、同時に本プロセッサの応用例についても報告する。

## 2. 基本思想

第1図にSDPを核としたシステム構成の例を示す。

SDPではデータをストリームとして扱う。またSDPは低レベルでのデータ駆動型の概念を持つ。

SDPは必要に応じて複数個多段接続することを前提に考える。複数個多段接続することは複雑な検索条件を実現する場合に各SDPに容易に検索条件の分担を可能とし、システム全体の性能向上に不可欠なものである。

このためSDPは入力と出力を分離した構成とし、それぞれ独立に動作可能とする。入力は常にデータ入力待ち状態であり、ストリームが入力されると動作を開始する。出力は処理結果あるいは入力されたストリームを入力と同様のストリーム形式で次段のSDPなどへ出力する。

したがってSDPを汎用のbusに接続する場合には汎用のbusからストリーム型のbusに、またストリーム型のbusから汎用のbusに変換するユニットが必要となる。しかし、SDPに汎用性を持たせる観点から考えると、特定のbusに依存しない現在の方式が妥当であると

考える。

SDPへの検索条件の設定、命令の設定は通常のデータ同様にストリーム形式で入力、設定する。これは専用のコマンド入力ラインを持たないことを意味しており、多段接続の実現性を十分に考慮した結果である。

### ・データ形式

SDPで使用するストリーム型のデータ形式を表1に示す。SDPで扱うデータは40bitであり、データ以外の8bitは32bitのデータからbus変換ユニットにて生成するものとする。SDPで使用するストリームデータの基本単位はレコードであり、レコードはフィールドの集合である。レコードのサイズ、フィールドの数はレコード単位に独立に設定可能とする。さらに、フィールドのサイズはレコード間の対応するフィールドで同じサイズの必要はなく、またレコード内においてもフィールドはすべて独立にサイズを設定可能とし、柔軟なデータ構造を実現する。

表1. SDPデータ形式

Dn	内 容
D0~D31	32bitの文字列、数値などのデータ
D32~D33	2bitのSDP番号を表す。
D34~D36	3bitのSelect-codeであり初期設定に使用する。
D37~D39	3bitのタグデータである。

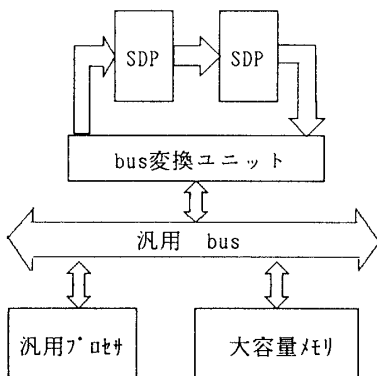
D32~D33、D34~D36はSDPに対して初期設定、すなわち検索条件の設定、命令の設定などに使用する。

D37~D39のタグデータは以下の内容である。

- 000: データ
- 001: ストリーム終了コード
- 010: ストリーム強制終了コード
- 011: 未使用
- 100: レコードスタート
- 101: レコードエンド
- 110: フィールドスタート
- 111: フィールドエンド

## 3. ハードウェアアーキテクチャ

第2図にSDPの内部構成を示す。SDPはストリーム入力部(SIE)、検索処理部(PME)、ストリーム出力部(SOE)の3ブロックより構成され、3個のレコードバッファがクロスバースイッチによりそれぞれのブロックに順次切り換え接続される。これより各ブロックの処理は各々独立に行うことが可能であり、3段のパイプライン構成を実現する。入出力はすべてストリーム形式で行い、SDPへの検索条件の設定、命令の設定もストリーム形式でSDPに入力、設定するためストリーム入出力アーキテクチャのみ存在し、汎用的なバスアーキテクチャは存在し



第1図 システム構成図

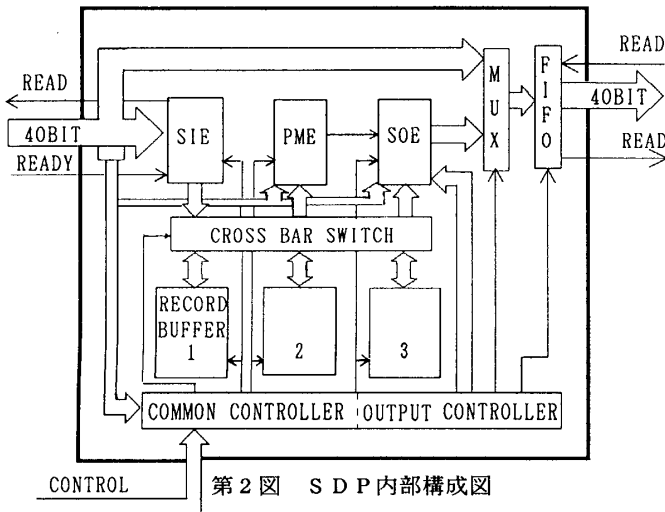
ない。

以下、SDPの概略動作について説明を行う。

SIEでは入力されてきたストリームデータをレコードバッファに格納し、同時に3bitのタグデータを基に各フィールドのインデックスを作成し、同様にレコードバッファに格納する。PMEでは格納されたストリームデータ、インデックスデータを基に検索処理を実行し、実行結果は直接SOEに出力する。SOEではPMEの実行結果を基にストリームデータを処理し、出力する。なお、各処理は独立に並行して実行している。

クロスバースイッチは各処理の終了を検出し、すべての処理が終了した時点でレコードバッファを切り換える。

SOEの出力に接続されているMUXはデータのパイパス用であり、SDPでの処理が発生しない場合に高速に次段のSDPにストリームを転送する場合に使用する。また、FIFOメモリは次段のSDPとの間での整合を取るためのものである。



第2図 SDP内部構成図

4. 仕様及び目標性能

SDPの仕様及び目標性能を表2に示す。

表2. 仕様及び目標性能

レコードサイズ	最大4Kbyte
フィールドサイズ	最大4Kbyte-タグ(3byte)
フィールド個数	最大32個
ストリーム速度	最大26Mbyte/sec
検索条件	正規表現をサポート

ストリーム型のデータを扱うプロセサにおいては、転送速度すなわちストリーム速度が一番重要であるとする。SDPでは各ブロック(SIE、PME、SOE)が並列に動作を行うがクロスバースイッチの切り替えを3ブロックの同期をとって行うため一番処理の遅いブロックの処理速度が全体のストリーム速度となる。現在開発を進めているSDPでは、ストリーム速度は最大26Mbyte/secであり実用的な速度に達していると考えている。

レコードサイズ及びフィールドサイズは現在最大4Kbyteであり2byte文字(漢字など)で約2000文字となっており、A4サイズ1ページを1レコードとして処理の対象とすることも可能となっている。

5. 機能

SDPの機能すなわち処理の単位はすべてレコード単位でありレコード間での演算を含む処理は実行出来ない。

以下にSDPで実現する機能の概略を示す。

検索：SDPの基本処理であり、結果により出力ストリームの抑制、内容変更等の処理を行う。

変更：フィールド単位に内容の変更を行う機能であり検索機能と結合することも可能である。

追加、削除：レコード単位の処理であり特定のレコード間に新しいレコードの追加を行ったり、特定のレコードの削除を行う。

フィールド操作：フィールドの切出し、フィールド単位の削除を行う。

6. 応用例

SDPの応用例としては以下のものが考えられる。

(1) AIシステム

AIシステムでは問題解決に必要な知識を膨大な知識情報から高速に選択することが必要である。一般的に知識情報はif-thenルール形式、あるいはフレーム形式などで表現されているが、完全に形式化されたデータフォーマットにはなっていない。SDPでは非常に柔軟なデータフォーマットをサポートしており、知識情報の検索処理にはもっとも適している。

(2) RDB

RDB(リレーショナルデータベース)では表形式のデータ処理が基本であり、SDPによる高速検索は可能である。RDBではSELECT、PROJECT、JOIN、SORTなどが代表的な処理であるがSDPでは各処理に共通して発生するパターンマッチング処理の高速支援を行う。

(3) 電子辞書

CD-ROMなどに格納された辞書を用いて処理を行う場合に目的とする辞書エントリの検索に本プロセサは有効である。辞書の検索では見出しによる検索、意味分類による検索、などがあり本プロセサによる柔軟な検索機能が有効に作用する。

(4) 全文テキストDB

本プロセサのレコードサイズは4Kbyte(漢字で約2000文字)であり、A4サイズ1ページ全てを1レコードとして扱う処理が可能である。このため、特許データベースの検索、あいまいな内容による書籍の検索にも有効である。

7. むすび

全数検索処理を専用に行うストリームデータプロセサSDPの基本アーキテクチャについて報告した。今後ハードウェアの開発を行い性能評価を進めていく予定である。また、試作したプロセサを実際にホスト計算機に組み込むことによりその効果を明らかにしたい。

末筆ながら、貴重な御意見を頂いた東京大学 喜連川助教授に感謝します。

【参考文献】

(1) K.C.Lee, G.Herman: "A High Performance VLSI Data Filter" 5th International Workshop on Database Machines P394~P411  
 (2) K.Takahashi, H.Yamada他: "Intelligent String Search Processor to Accelerate Text Information Retrieval" 5th International Workshop on Database Machines P440~P453