Scan-Directional Architectures

Dinu Coltuc$^1$ and Jean-Marie Becker$^2$

$^1$ Research Institute for Electrical Engineering,
Splaiul Unirii 313, Bucharest, Romania
$^2$ Ecole Supérieure de Chimie Physique Electronique de Lyon,
43, bd du 11 Novembre 1918, BP 2077 - 69616 Villeurbanne cedex - France

Abstract. A considerable number of algorithms, of interest for image processing and general matrix analysis problems, operate on certain pre-defined streams of data. This paper presents an architectural framework to tackle these algorithms, namely parallel/pipeline architectures built with "scan-directional" memories. The paper analyzes the realization of such memory modules for simultaneous fast access on various scan-directions, e.g., lines, columns, diagonals, Peano-curve, bit-reversed, etc. Some of the most important treatments (matrix transposition, multiplication, inversion, Schur complement, orthogonal transforms, linear, non-linear image filtering, etc.) fit into the scan-directional frame. Their implementation is investigated.

1 Introduction

Algorithms can roughly be divided into two main categories: data oriented or address oriented. This paper is concerned with the second class, of great interest for matrix oriented problems and notably for image/signal processing applications. A well-known example of a wholly address oriented algorithm is the FFT where 1) either the input or the output data needs bit-reversal ordering; 2) "butterflies" are computed for fixed positions of the data. In the sequel, such a predefined data-scanning order will be called a scan pattern. Usually, the scan patterns are derived from the "geometry" of the data set, e.g., rows, columns. Different algorithms may need different scan patterns. For such algorithms, the implementation efficiency depends not only on the processing speed, but also on the access time on the required scan patterns. This paper presents a scan-directional approach for the implementation of address oriented algorithms - parallel (or pipeline) architectures developed around multiple scan pattern memory modules.

2 Problem Setting

The scan directional approach copes with the implementation of address oriented algorithms. At a general level the proposed architectures are composed of memory modules and processing devices. Each memory module provides high speed
access on predefined streams of data (scan patterns) to feed on several processing devices. More precisely, a problem is considered to fit the scan directional approach when:

1. data processing order is fully known and it is independent of data values;
2. different kinds of scan patterns are required;
3. high speed data streams are needed.

Requirements 1 and 2 are algorithm-specific. Requirement 3, which seems to be application-specific, deserves some comments. It should be understood in the context of the memory module bandwidth: namely the memory module is supposed to be implemented with much slower memory chips than the access time required for its input/output. In fact, the main aspect of this approach is to provide predefined high speed streams of data by using slow memory chips. Even if nowadays memory chips become faster and faster the problem of memory bandwidth is and will remain an important issue. To give some arguments, we mention that: once, the processing speed increases as well; twice, the need of parallelism demands that multiple data streams are simultaneously available.

To summarize, the proposed scan directional approach provides general architectures to be tailored to each specific application. The core of such architectures consists of memory modules, called multiple scan pattern memory modules, that provide very high bandwidth on several data streams. While the memory module architecture is general, the mapping of the data into the memory chips strongly depends on the scan patterns at hand (application specific). Moreover, the very existence of the mapping should be investigated. An important result in our approach is the existence of a memory mapping for any 2 scan patterns. For more than 2 scan patterns no such existence theorem can be given and the mapping, if any, should be established by construction. A lot of work on memory mappings has been already done in [1], [2]. The aim of this paper is to provide a unified approach on this subject and to point out the architectural framework suitable for many applications (most of them classical) in the field of matrix computation and signal/image processing.

3 Multiple Scan Pattern Memory Modules: Architecture

Let memory modules be organized as square arrays of size $N \times N$. Each data element is specified by its location, i.e., the ordered pair $(x, y)$, $0 \leq x, y < N$, where $x, y$ are the row and the column position, respectively. If $n$ memory chips (banks) are used for the implementation of the memory module, one must determine an explicit mapping:

$$m : [0, N - 1] \times [0, N - 1] \rightarrow [0, n - 1]$$

Interleaved memory systems are often used for increasing the effective memory bandwidth (Fig. 1). Thus, $n$ data are simultaneously read from $n$ memory chips, then loaded in a parallel/serial converter and shifted serially. The serial