Consiglio Nazionale delle Ricerche

# ISTITUTO DI ELABORAZIONE DELLA INFORMAZIONE

## PISA

VLSI LINEAR TRANSFORMATIONS OF BANDED MATRICES

B. Codenotti, G. Lotti

Nota interna B84-17

Settembre 1984

# VLSI LINEAR TRANSFORMATIONS OF BANDED MATRICES

B. CODENOTTI[1] and G. LOTTI[2]

[1]Istituto di Elaborazione dell'Informazione - CNR, via S. Maria 46, PISA, ITALY.

[2]Dipartimento di Scienze dell'Informazione - University of Pisa, Corso Italia 40, PISA, ITALY.

## ABSTRACT

The complexity of the multiplication of a banded matrix by a vector is studied, with respect to the VLSI model.

Upper and lower bounds are showed, which improve the known results.

An upper bound, which is optimal up to logarithmic factors, is obtained for the case of fixed band.

Key Words - VLSI Design, Area-Time Complexity, Banded Matrix, Recursive algorithm.

# 1. INTRODUCTION

In this paper, the linear transformation of a vector by a $(2k+1)$-diagonal matrix with variable entries, is studied.

This problem has been investigated in [KUNG], where the systolic arrays structure has been proposed, producing $AT^2 = O(n^2 k)$.

By comparing this result with the lower bound obtained by using the well known tecnique of minimal information flow [THO], it seems possible to improve either the lower and the upper bound.

In fact, the minimal information flow results to be dependent on $k$ and not on the size $n$ of the matrix, while simple considerations based on the number of inputs, yield an $\Omega (nk)$ lower bound.

On the other hand, the above mentioned upper bound can be also improved, in the case of $(2k+1)$-diagonal matrices, with $k$ not increasing with $n$.

This assumption does not influence the validity of our result, since $(2k+1)$-diagonal matrices with fixed $k$ represent a very important class of banded matrices (e.g. matrices arising from differential problems).

In section 3, the upper bound $AT^2 = O(n \, k \, \log^3 n)$, is obtained, which meets the lower bound of section 2 up to logarithmic factors.

## 2. LOWER BOUNDS

In this section, the well known technique introduced by Thompson [THO] to give lower bounds to the minimal information flow across a VLSI circuit is considered. This technique has been succesfully applied to the case of a linear transform, whose matrix has constant entries (e.g. DFT transform).

Our problem can be seen as a linear transform defined by a $(2k+1)$-diagonal matrix $A(x)$, i.e. $A(x)=(a_{ij})$, $a_{ij} = 0$ for $|i-j|>k$, with varible entries.

Let

$$A(x) = \begin{bmatrix} A_{11}(x) & A_{12}(x) \\ A_{21}(x) & A_{22}(x) \end{bmatrix}$$ , the computation of the n-vector

$z(x)=A(x)y$ is now splitted as

$$z_1(x) = A_{11}(x)y_1 + A_{12}(x)y_2 ,$$

$$z_2(x) = A_{21}(x)y_1 + A_{22}(x)y_2 ,$$

where $z_1(x) \cap z_2(x) = \emptyset$, $z_1(x) \cup z_2(x) = z(x)$, $|z_1(x)|=|z_2(x)|=n/2$,

$y_1 \cap y_2 = \emptyset$, $y_1 \cup y_2 = y$, $|y_1|+|y_2|=n$.

The straightforward extension of Thompson approach yields the following lower bound to the minimal flow I:

$$I \geq \max_{x} \ rk \ A_{12}(x) + rk \ A_{21}(x) .$$

It is trivial to see that the above relation, in the case of banded matrices, gives a lower bound not depending on the size

- 3 -

n of the matrix. A more significant lower bound can be obtained by the following arguments.

If $p = O(nk)$ and $r = \Theta(nk)$ are respectively the number of input gates and of input variables, then $A = \Omega(p)$ and $T = \Omega(r/p)$, that is $AT^2 = \Omega(n^2k^2/p) = \Omega(nk)$.

The best known upper bound is $AT^2 = O(n^2k)$ [Kung] obtained by sistolic arrays. In the next section an area-time upper bound of order $O(n \log^3 n)$ is shown.

## 3. UPPER BOUNDS

Let $n = 2^h$, let A be a $(2k+1)$-diagonal matrix of size $n$, partitioned as

$$A = \begin{bmatrix} U & V \\ W & Z \end{bmatrix}, \quad \text{where U and Z are } (2k+1)\text{-diagonal square}$$

matrices of size $n/2$.

The computation of the matrix-vector product $z = Ay$ can be performed as follows.

Let $y = [\, y_1^T, y_2^T \,]$, $z = [\, z_1^T, z_2^T \,]$, where $y_1, y_2$, $z_1, z_2$ are $n/2$-vectors; then

$$
\begin{aligned}
z_1 &= U y_1 + V y_2 \\
z_2 &= W y_1 + Z y_2 .
\end{aligned}
$$

(3.1)

Formula (3.1) leads to the recursive design of fig.1, where gates of type (1), (2), (3) transmit respectively the entries
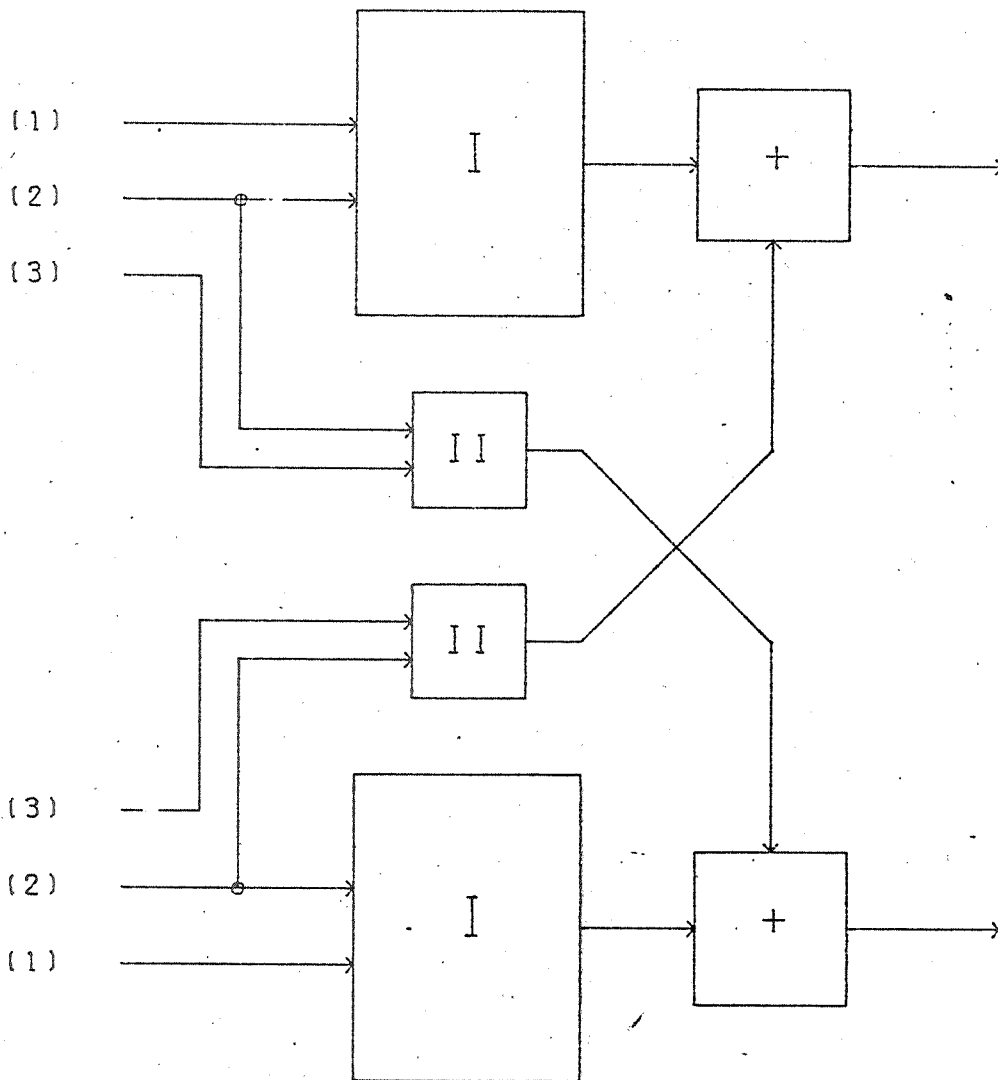
- 4 -

Fig. 1   Recursive module structure.

of $U$, $y_1$, $V$; modules of type I perform the transform of size $n/2$; modules of type II perform the matrix-vector product of size $k$, which requires $A = O(k^2 \log^2 k)$ and $T = O(\log k)$, using the well known structure of mesh of trees [LEI].

In order to allow a correct recursive implementation, the $r=(k+1)(2n-k)$ inputs have to be sent with the following order

$$a_{11} \ a_{21} \ \ldots \ a_{k1} \ b_1 \ a_{12} \ \ldots \ a_{k+1,2} \ b_2 \ \ldots \ a_{n-k+1,n} \ \ldots \ a_{nn} \ b_n ,$$

throught the $r$ input gates.

In fig.2, the recursive circuit is shown in more detail, in the case $n=4$, $k=1$.

The inner recursive step consists of $(k+1) \times (k+1)$ $(k \times k)$ matrix-vector product, for $k$ odd (even), which can be performed by using mesh of trees.

Let $T(n)$, $H(n)$, $L(n)$ be respectively the time, the height and the width required by the circuit related to the problem of size $n$.

It readily follows that:

$$T(n) = \max \left\{ T(n/2) , O(\log k) \right\} + O(1) = O(\log n),$$

$$H(n) = 2H(n/2) + O(nk) = O(nk \log n),$$

$$L(n) = L(n/2) + O(k \log k) = O(\log n),$$

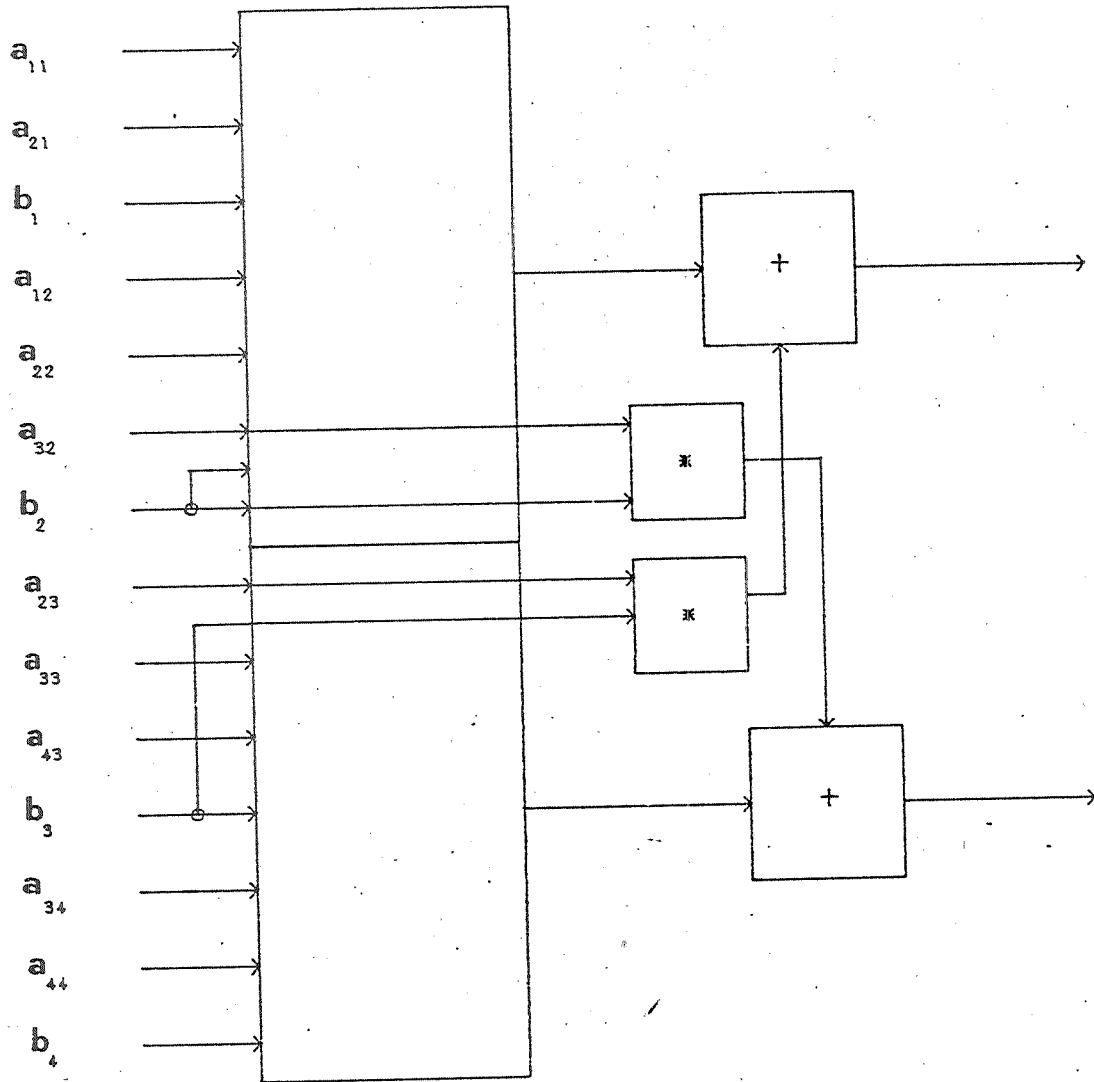from which an $AT^2$ upper bound of order $O(nk \log^3 n)$ is obtained.

Fig. 2. Detailed design for n=4 and k=1.

# REFERENCES

[1] Kung, H.T. and C.E. Leiserson, Algorithms of VLSI Processor Arrays, in: C.A. Mead and L.A. Conway, eds., Introduction to VLSI Systems (Addison-Wesley, Reading, MA, 1980) pp. 271-292.

[2] F.T. LEIGHTON, New Lower Bound Tecniques for VLSI. Proceedings of the 22nd Annual IEEE Symposium on Foundations of Computer Science (1981), 1-12.

[3] C.D. THOMPSON, Area-Time Complexity for VLSI. Proc. 11th Annual ACM Symp. on the Theory of Computing (SIGACT) (1979) 81-88.