Práctica 2

Prototipado de matrices sistólicas (II)

1. Objetivos

El objetivo de esta práctica es que el alumno comprenda el funcionamiento de las matrices sistólicas propuestas, calcule su coste computacional y sea capaz de diseñar un prototipo que demuestre el correcto funcionamiento de la solución prevista. Para ello se propondrá el diseño de una red lineal de multiplicación matriz/vector, mucho más sencillo que el anillo sistólico de la práctica anterior. Basándose en un EP muy similar al utilizado anteriormente se propondrá el diseño de una matriz sistólica de multiplicación de matrices. El alumno diseñará ambas matrices sistólicas, analizará sus ventajas en coste temporal, analizará la relación computación frente a entrada/salida y analizará el coste en área frente a un procesador secuencial.

Es fundamental que el alumno haya aprendido a utilizar las nociones básicas sobre OrCad, realización de patrones de test, etc., aprendidos en la primera sesión de prácticas.

2. Introducción

A continuación se explican las dos arquitecturas a diseñar en esta sesión de prácticas. En esta sesión no se explica la descripción VHDL de cada elemento de proceso por lo que el alumno habrá tenido que preparar esta descripción con antelación, al menos tenerlo esbozado. Tampoco se ponen explícitamente los vectores de test por lo que tendrán que estar preparados antes de empezar la sesión.

2.1 Red lineal de multiplicación matriz/vector

Consideremos el problema de multiplicar una matriz $A=(a_{ij})$ por un vector $Y=(x_1 \ldots x_n)^T$. Los elementos en el producto $Y=(y_1 \ldots y_n)^T$ pueden ser computados por las siguientes recurrencias.

$$y_i = y_i^{(n+1)}$$
$$y_i^{(1)} = 0$$
$$y_i^{(k+1)} = y_i^{(k)} + a_{ik}\, x_k$$

Supongamos que A es una matriz de $n \times n$ elementos con una banda $w=p+q-1$. Siendo p el índice j del último elemento a_{1j} cumpliéndose: $a_{1j} \neq 0$ y $a_{1k}=0 \ \forall k>j$. De la misma manera q será el índice i del último elemento a_{i1} cumpliéndose: $a_{i1} \neq 0$ y $a_{k1}=0 \ \forall k>i$. En estas condiciones se puede calcular el vector Y en $2n+w$ pasos sistólicos. Un ejemplo aclaratorio es el que se muestra a continuación, en este caso $p=2$ y $q=3$, con lo que $w=4$.

$$\begin{array}{cc} & \overbrace{\qquad\qquad}^{p} \\ q & \left[\begin{array}{cccccc} a_{11} & a_{12} & 0 & 0 & \ldots & 0 \\ a_{21} & a_{22} & a_{23} & 0 & \ldots & 0 \\ a_{31} & a_{32} & a_{33} & a_{34} & \ldots & 0 \\ 0 & a_{42} & a_{43} & a_{44} & & \\ 0 & 0 & a_{53} & a_{54} & & \\ .. & .. & & & & \\ 0 & 0 & .. & & & a_{nn} \end{array}\right] \end{array} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ .. \\ \\ .. \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ .. \\ \\ .. \\ y_n \end{bmatrix}$$

Con este supuesto las ecuaciones recurrentes se pueden resolver con una red lineal sistólica, que para el caso de la matriz anterior ($w = 4$) se muestra en la figura 1.

En cada paso la simulación entra en la red una línea de índices de manera que en primer lugar entra a_{11}, en el siguiente golpe de reloj entrarán a_{12} y a_{21} para de esta manera sucesivamente entren el resto de coeficientes. Simultáneamente por la izquierda de la red van entrando los valores de x y por la derecha van entrando los valores de y. Inicialmente los valores de y serán 0, pero se van actualizando en cada transición con el valor calculados por el EP, que será $y_i^{(k+1)}=y_i^{(k)}+a_{ik}\, x_k$.

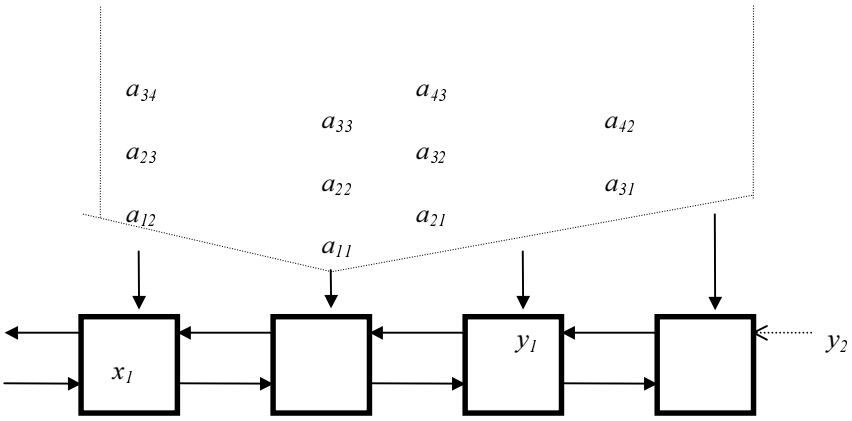

Figura 1: Red lineal sistólica de multiplicación matriz/vector.

Para una mejor comprensión del funcionamiento de la red lineal en la figura 2 se muestran los primeros pasos de su funcionamiento.

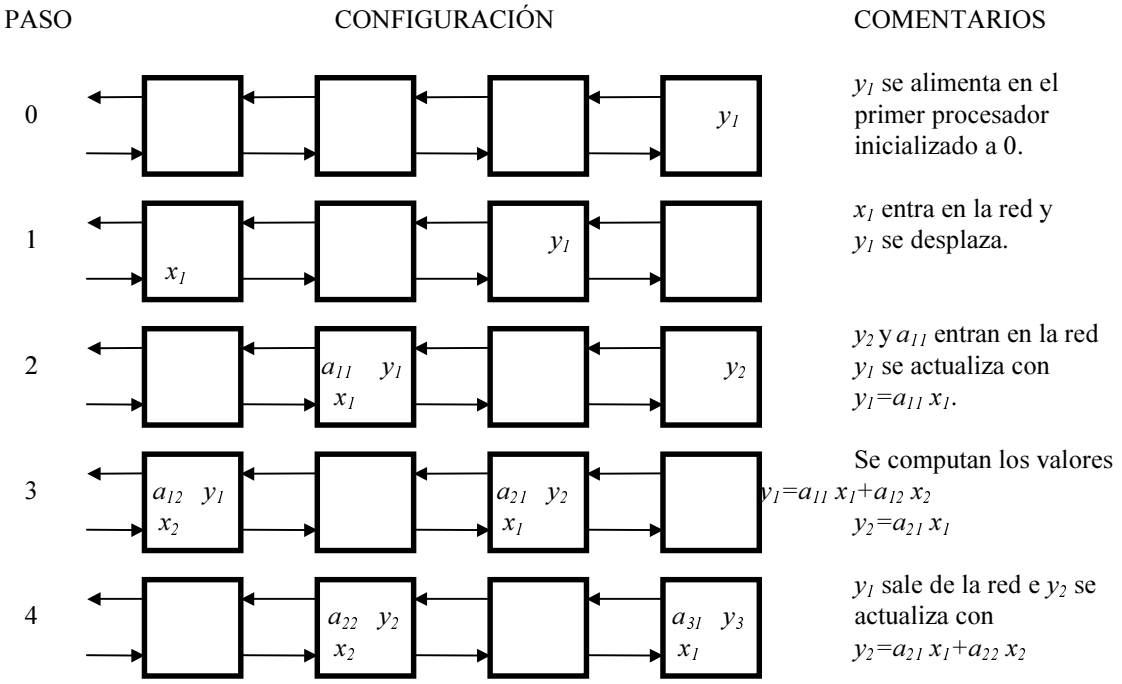

Figura 2: Primeros pasos de simulación.

Realiza un prototipo con OrCad que muestre el correcto funcionamiento de la red sistólica. Para ello se debe diseñar el EP que realiza el producto y la suma y a partir de ahí montar la red lineal. Es importante el orden en que deben entrar los coeficientes en la red, así como el instante exacto en que la coordenada de x entra también en la red.

2.2 Multiplicación matriz/matriz en redes hexagonales.

Siguiendo el ejemplo expuesto de multiplicación matriz vector es posible ampliar la matriz sistólica para realizar la multiplicación de matrices. Sean 2 matrices $n \times n$, $A=(a_{ij})$ y $B=(b_{ij})$. Los elementos del producto de ambas matrices $C=(c_{ij})$ pueden ser computados por las siguientes recurrencias.

$$c_{ij}^{(1)} = 0$$
$$c_{ij}^{(k+1)} = c_{ij}^{(k)} + a_{ik}\, b_{kj}$$
$$c_{ij} = c_{ij}^{(n+1)}$$

Supongamos que A es una matriz $n \times n$ con una banda w_1 y B es una matriz $n \times n$ de banda w_2. El producto (la matriz C $n \times n$) se puede calcular con una red de $w_1 \cdot w_2$ procesadores sistólicos conectados hexagonalmente en $2n+\min\{w_1,w_2\}$ pasos sistólicos. En la figura 3 se muestra el caso particular en el que $n=3$ y $w_1=w_2=3$. Se muestra el camino de los índices de las matrices. Se deja como ejercicio la especificación del orden del resto de los índices a_{ij} y b_{ij}, y el diseño del EP. Se deberá de mostrar el correcto funcionamiento de la red con un ejemplo numérico.

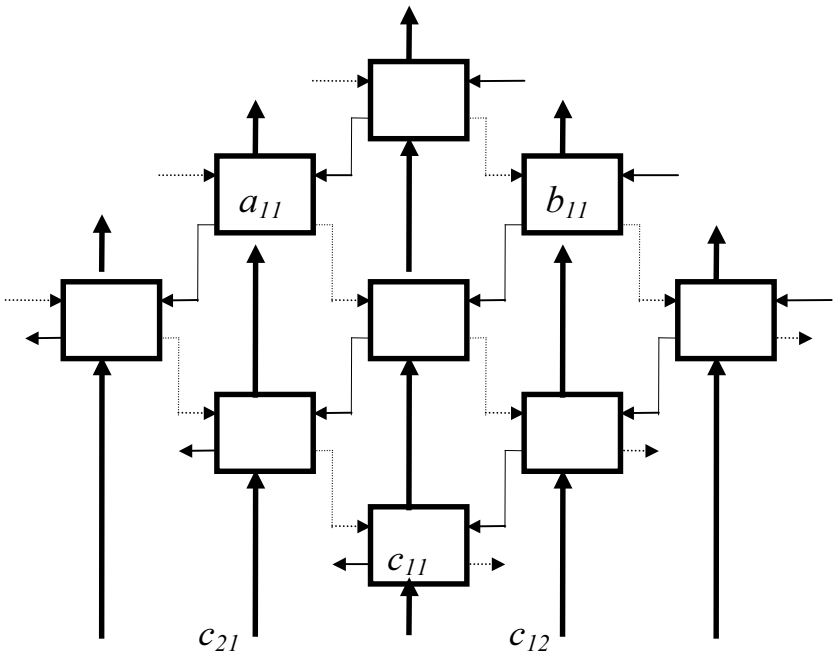

Figura 3: Primeros pasos de simulación.

Nota: Estas matrices sistólicas son estructuralmente más complejas, por lo que el símbolo generado por OrCad es inadecuado para cablear toda la matriz. En este caso conviene modificar el símbolo antes de usarlo y colocar las entradas y salidas tal y como se han venido mostrando en las figuras precedentes. Esto se realiza de forma sencilla dentro de Capture, una vez generado el símbolo, acudiendo a la ventana del proyecto a la carpeta *Outputs,* luego a *./proc.olb* que es la biblioteca con los componentes y luego seleccionar el elemento de proceso clicando dos veces, momento en el que se abre una ventana donde modificar el símbolo, mover los pines, etc.

3. Resultados

Se debe realizar el diseño de la red lineal en el laboratorio y demostrar el correcto funcionamiento con un ejemplo numérico concreto. Se debe también realizar la matriz hexagonal de multiplicación de matrices y mostrar su correcto funcionamiento.

La memoria que se realice de esta práctica deberá incluir la descripción del trabajo realizado y un análisis de los costes (en tiempo de proceso) de la resolución de estos problemas con las arquitecturas sistólicas comparándolas con otras arquitecturas paralelas o no. También deberá incluirse un resumen de las principales características del procesamiento sistólico, sus ventajas y desventajas.