

Systolic Architecture Design

(b) Map the same DG with different mapping vectors:

i. $\mathbf{d}^T=[1 \ 1]$, $\mathbf{p}^T=[1 \ -1]$, $\mathbf{s}^T=[1 \ 0]$.

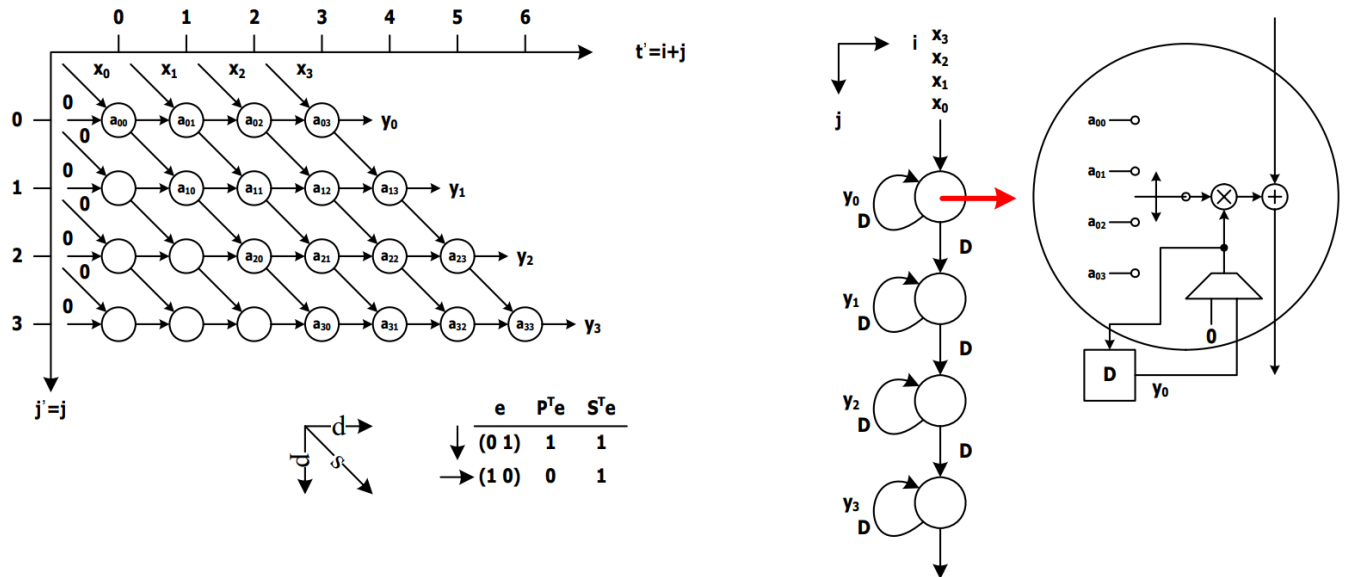
ii. $\mathbf{d}^T=[1 \ 1]$, $\mathbf{p}^T=[1 \ -1]$, $\mathbf{s}^T=[0 \ 1]$.

We have three architectures now, including these two architectures and the one in (a). Do you think which one is better? Why?

(c) Use the mapping vector in (a). If the elements in matrix \mathbf{A} are not constant.

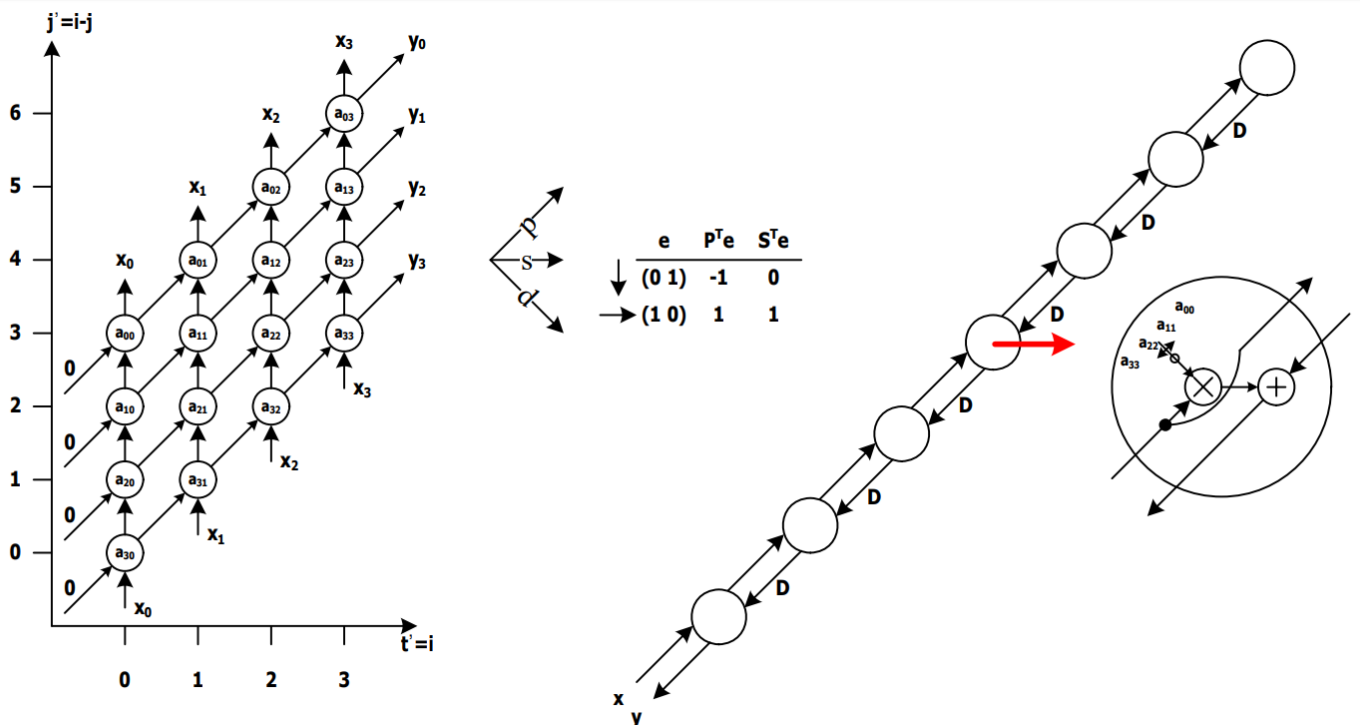
Derive the systolic architecture.

(a)

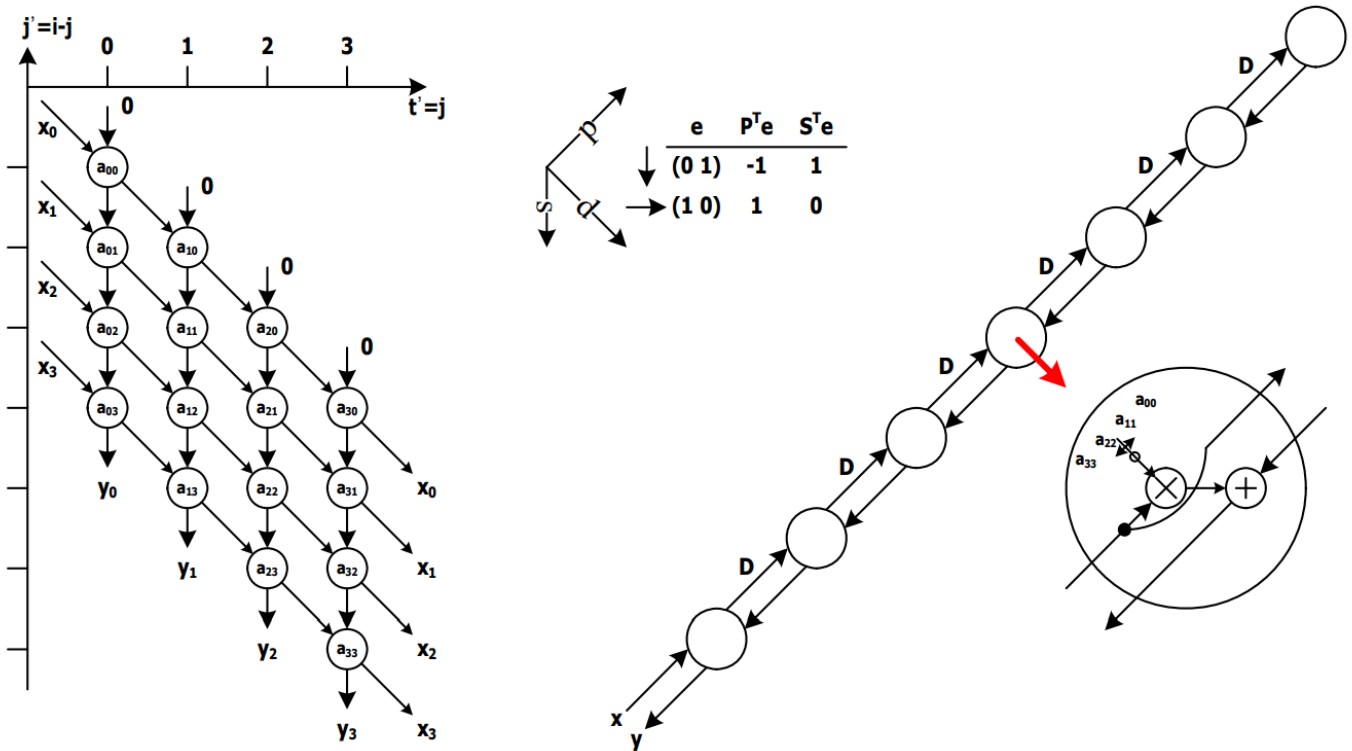


(b)

(i)



(ii)



After folding (b-i) and (b-ii), the resources used by them are less than that used by (a). Besides, critical paths of (a) and (b-i) are shorter than (b-ii). From the two points of view, (b-i) is better. But we should choose one of them not only according to the performance but also to our design specifications like I/O constraints.

(c) The only difference between this one and architecture in (a) is the registers put on the path of a_{xx} to store arbitrary data.

2. Remember in homework 1. You have drawn the DG of a 2D moving average filter for an image, which can be shown as the following equation:

$$y(i, j) = \sum_{m=-1}^1 \sum_{n=-1}^1 x(i+m, j+n),$$

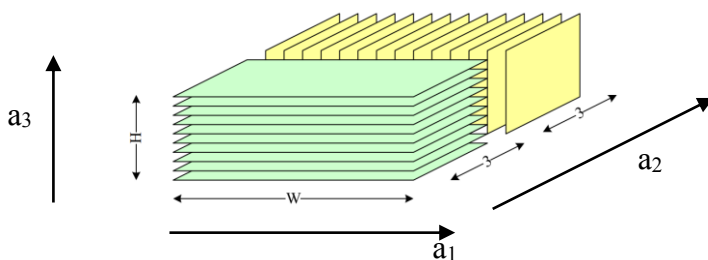
$$i, i+m \in [0, W]$$

$$j, j+n \in [0, H],$$

where $x(i, j)$ is the input image, $y(i, j)$ is the filtered image, and W and H are the width and height of the image, respectively.

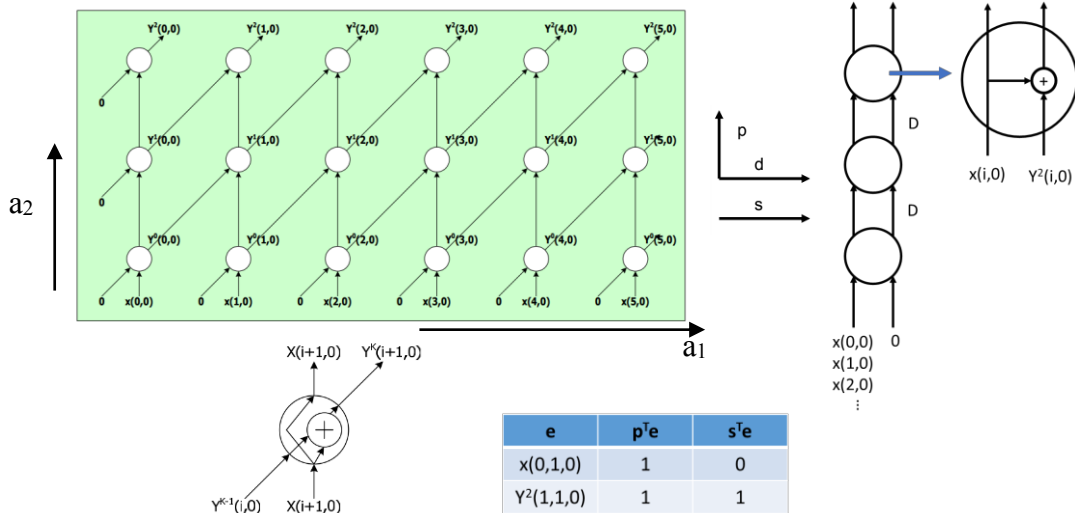
(a) Now, please choose two set of mapping vectors (projection vector and scheduling vector) and draw the associated hardware architecture.

In homework1, we separate the DG to 2 directional DG.



Step 1

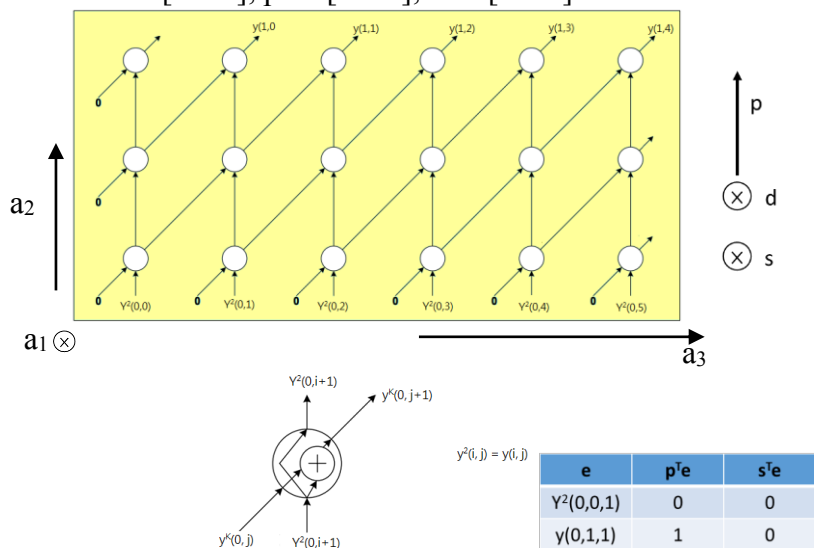
$$d^T = [1 \ 0 \ 0], p^T = [0 \ 1 \ 0], s^T = [1 \ 0 \ 0]$$



Step 2

Similar to step 1

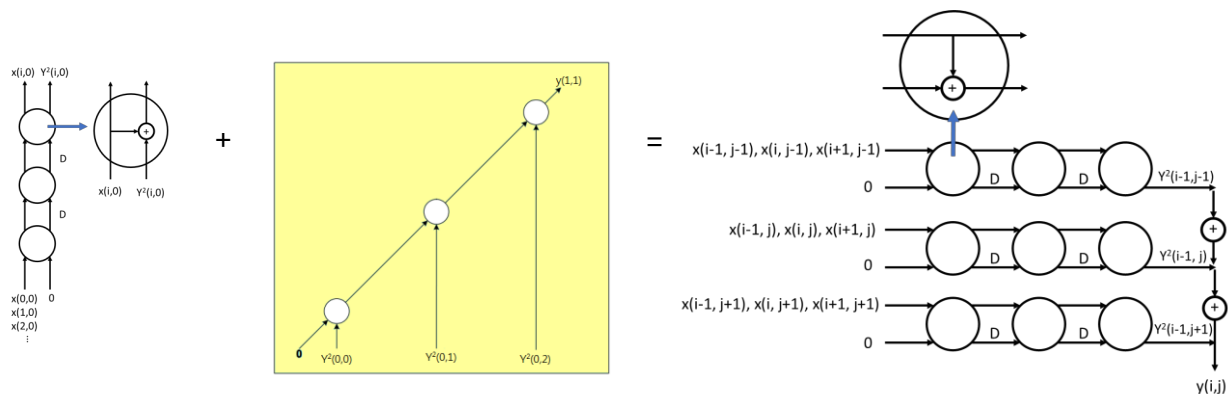
$$d^T = [1 \ 0 \ 0], p^T = [0 \ 1 \ 0], s^T = [1 \ 0 \ 0]$$



The systolic architecture is the same.

Step 3

If we compute $Y^2(i, j)$, $Y^2(i, j+1)$, $Y^2(i, j+2)$ at the same time, we can merge 2 architecture.



(b) Show the scheduling of our hardware to prove its function is correct.

Cycle 1

$x(0, 0)$	$x(0, 0)$	$x(0, 0)$
$x(0, 1)$	$x(0, 1)$	$x(0, 1)$
$x(0, 2)$	$x(0, 2)$	$x(0, 2)$

Cycle 2

$x(1, 0)$	$x(0, 0) + x(1, 0)$	$x(0, 0) + x(1, 0)$
$x(1, 1)$	$x(0, 1) + x(1, 1)$	$x(0, 1) + x(1, 1)$
$x(1, 2)$	$x(0, 2) + x(1, 2)$	$x(0, 2) + x(1, 2)$

Cycle 3

$x(2, 0)$	$x(1, 0) + x(2, 0)$	$x(0, 0) + x(1, 0) + x(2, 0)$
$x(2, 1)$	$x(1, 1) + x(2, 0)$	$x(0, 1) + x(1, 1) + x(2, 0)$
$x(2, 2)$	$x(1, 2) + x(2, 0)$	$x(0, 2) + x(1, 2) + x(2, 0)$
		Sum: $y(1, 1)$

Cycle 4

$x(3, 0)$	$x(2, 0) + x(3, 0)$	$x(1, 0) + x(2, 0) + x(3, 0)$
$x(3, 1)$	$x(2, 1) + x(3, 1)$	$x(1, 1) + x(2, 0) + x(3, 1)$
$x(3, 2)$	$x(2, 2) + x(3, 2)$	$x(1, 2) + x(2, 0) + x(3, 2)$
		Sum: $y(2, 1)$

作業七常見錯誤如下：

- 1-(a) PE 架構（正解請參考附圖）
- 1. systolic 少標 x 或 y
- 1-(b)-(i)(ii) 會有 7 個 PE，若有做 folding 需註記

作業八常見錯誤如下：

- 1. x 或 y 都要在 $\{0\}$ 進出，要有相對應的 delay
- 1. y 少標接到 output 的時間點

另外作業七解答有些寫錯的地方，在這邊先跟大家說明：

- 1-(a) PE 的內部架構有誤（正解請參考附圖）。乘法器的輸入應為 x 和係數 a_{ji} ， D 中的 y 值應為加法器的輸入而非乘法器。
- 1-(b) 中 (i) 和 (ii) 的 systolic architecture x 和 y 的傳遞方向用原圖壓的話都反了，但因為不影響架構，只要 D 在正確的 edge 上就算對。
- 1-(b) 中 space-time representation 各有一些錯，像是 1-(b)-(i) 的 x 方向要從向上改成向下，以及 1-(b)-(ii) 中 a_{ji} 擺放位置錯誤等，這部分同學可參考 1-(a) 以及課程講義的圖比較不會搞混。（1-(b) 只須畫 systolic architecture 故這部分不計分）

