Calcolo Parallelo e Distribuito

a.a. 2021-2022

Prodotto Matrice-Vettore
approfondimenti
parte 5

Docente: Prof. L. Marcellino

PROBLEMA: Prodotto Matrice-Vettore

algoritmo

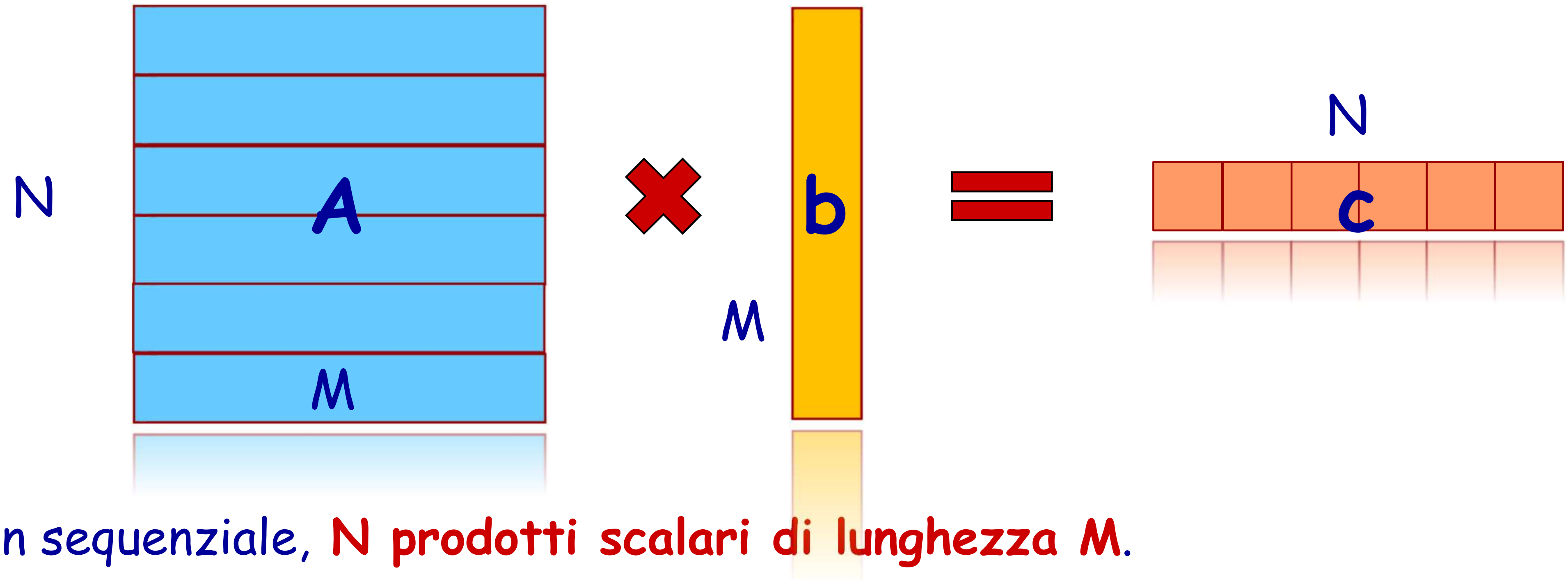
per il calcolo del prodotto

di una matrice *A* per un vettore b:

$$\boxed{A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}}$$

Complessità computazionale dell'algoritmo sequenziale

$$A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}$$



In sequenziale, **N prodotti scalari di lunghezza M**.

Per fare 1 prodotto scalare di lunghezza M, devo fare:

M molt + (M-1) add

Complessità computazionale dell'algoritmo sequenziale

$$A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}$$

In sequenziale, **N prodotti scalari di lunghezza M**, cioè:

N[M molt + (M-1) add]

molt ~ add

$$T_1(N \times M) = N[2M-1]\, t_{calc}$$

PROBLEMA: Prodotto Matrice-Vettore

Progettazione
di un algoritmo parallelo
per architettura MIMD

per il calcolo del prodotto
di una matrice A pr un vettore b:

$$\boxed{A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}}$$

III STRATEGIA

Decomposizione 1: BLOCCHI di RIGHE

+

Decomposizione 2: BLOCCHI di COLONNE

=

Decomposizione 3: BLOCCHI QUADRATI

Calcolo **di speedup ed efficienza** (**def classica**)

in ambiente **MIMD-DM**

e

in ambiente **MIMD-SM**

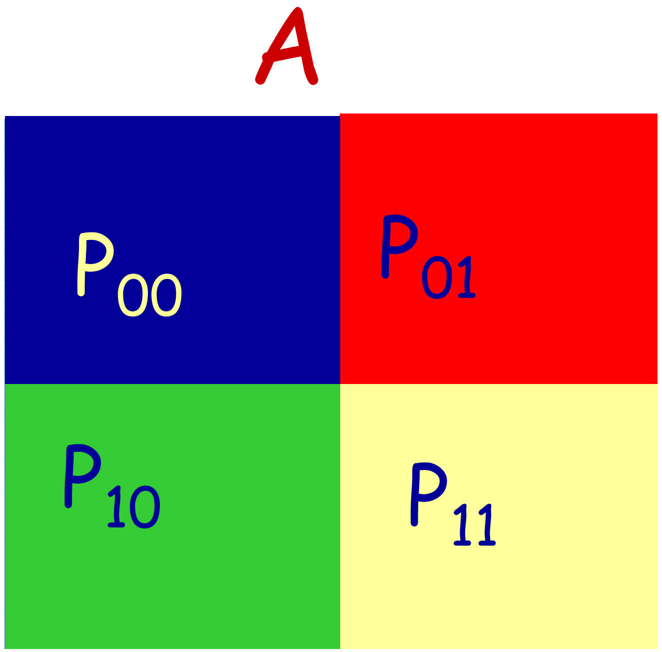
GRIGLIA DI PROCESSORI

P_0 (0,0)	P_1 (0,1)	P_2 (0,2)
P_3 (1,0)	P_4 (1,1)	P_5 (1,2)
P_6 (2,0)	P_7 (2,1)	P_8 (2,2)

III Strategia: speed-up/efficienza (**def classica**)

$$A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}$$

$$T_1(N \times M) = N[2M-1]\, t_{calc}$$



GRIGLIA

qxp=4=2x2 processori

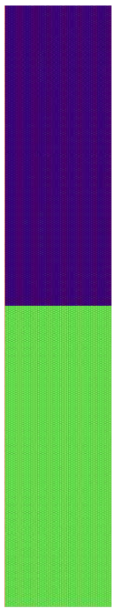
MIMD-DM

III Strategia: speed-up/efficienza (**def classica**)

$$A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}$$

$$T_1(N \times M) = N[2M-1]\, t_{calc}$$

b



P_{00}

Stesso blocco di b

P_{10}

P_{01}

Stesso blocco di b

P_{11}

GRIGLIA

qxp=4=2x2 processori

MIMD-DM

III Strategia: speed-up/efficienza (**def classica**)

$$A \in R^{N \times M}, \quad b \in R^{M}, c \in R^{N}$$

$\dim[A_{loc}] = (N/q)x(M/p)$

$\dim[b_{loc}] = M/p$

GRIGLIA

qxp=4=2x2 processori

MIMD-DM

Tutti contemporaneamente, **N/q prodotti scalari di lunghezza M/p**, cioè:

N/q [2M/p-1] tcalc

N.B.: non è detto che la **GRIGLIA** sia **quadrata**, potrebbe essere anche **rettangolare qxp=8=2x4**

III Strategia: speed-up/efficienza (**def classica**)

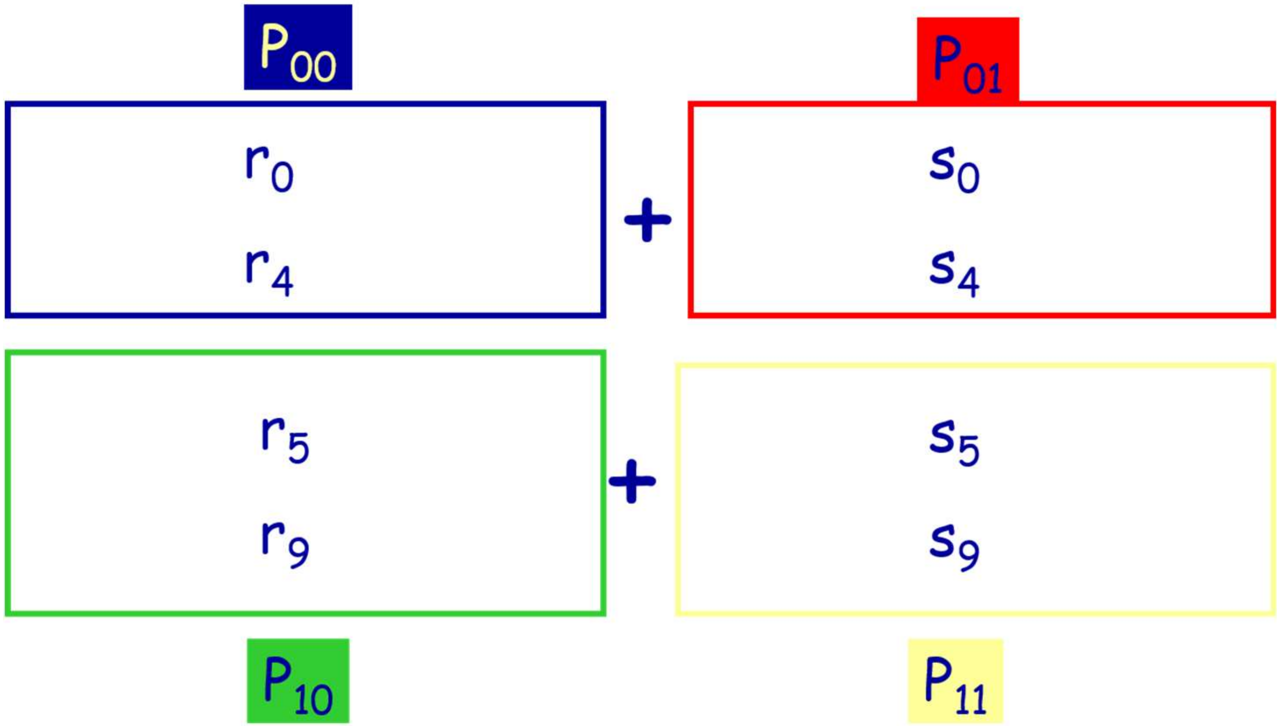
Collezione dei risultati:

comunicazione e somma in parallelo

GRIGLIA

$q \times p = 4 = 2 \times 2$ processori

MIMD-DM



III Strategia: speed-up/efficienza (**def classica**)

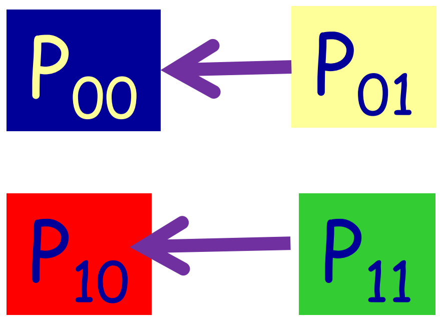
Collezione dei risultati

La scelta della strategia dipende da quanto vale p.

Se p è potenza di 2 allora conviene la II-III strategia, altrimenti (ed in particolare se è dispari) conviene usare la I strategia!!!

4=qxp=2x2 p=2

I strategia



I passo (e unico)

P_{01} P_{11} spediscono un vettore di N/q elementi a P_{00} P_{10} che, quindi, aggiornano il proprio vettore effettuando N/q somme, cioè:

$$N/q \text{ tcom} + N/q \text{ tcalc} = N/2 \text{ tcom} + N/2 \text{ tcalc}$$

III Strategia: speed-up/efficienza (**def classica**)

Collezione dei risultati

$4 = q \times p = 2 \times 2 \quad p=2$

I strategia

A questo punto mi posso anche fermare, sapendo che il vettore soluzione è diviso a metà tra P_{00} e P_{10}

tcom = c tcalc

2<= c <= 3

In generale:

(p-1)(N/q tcom + N/q tcalc) =

= (p-1)(c N/q tcalc + N/q tcalc)

III Strategia: speed-up/efficienza (**def classica**)

$$A \in R^{N\times M}, \quad b \in R^{M}, c \in R^{N}$$

MIMD-DM

GRIGLIA qxp

I strategia per collezione vettori

$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$

$= N[2M-1] /(N/q [2M/p-1] + cN/q (p-1) + N/q (p-1))$

$Oh = qxp\ T_{qxp}(NxM) - T_1(NxM) =$

$= pxq (N/q [2M/p-1] + cN/q (p-1) + N/q (p-1)\ tcalc +$

$- N[2M-1]\ tcalc$

$E_{qxp}(NxM) = S_{qxp}(NxM) / (qxp)=$

$= N[2M-1] / [(qxp)(N/q [2M/p-1] + cN/q (p-1) + N/q (p-1)]$

III Strategia: speed-up/efficienza (**def classica**)

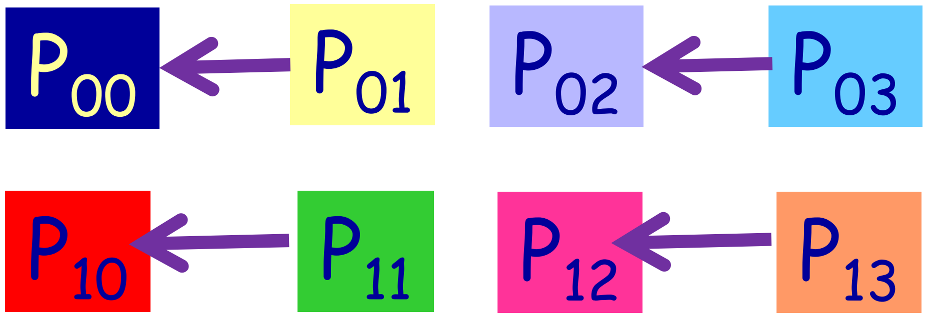
Collezione dei risultati

La scelta della strategia dipende da quanto vale p.

Se p è potenza di 2 allora conviene la II-III strategia, altrimenti (ed in particolare se è dispari) conviene usare la I strategia!!!

8=qxp=2x4 p=4

II strategia



I passo

$P_{01}\,P_{11}$ spediscono un vettore di N/q elementi a $P_{00}\,P_{10}$ e, contemporaneamente, $P_{03}\,P_{13}$ spediscono un vettore di N/q elementi a $P_{02}\,P_{12}$. I processori riceventi aggiornano con il proprio vettore effettuando N/q somme, cioè:

$$N/q \; tcom + N/q \; tcalc = N/2 \; tcom + N/2 \; tcalc$$

III Strategia: speed-up/efficienza (**def classica**)

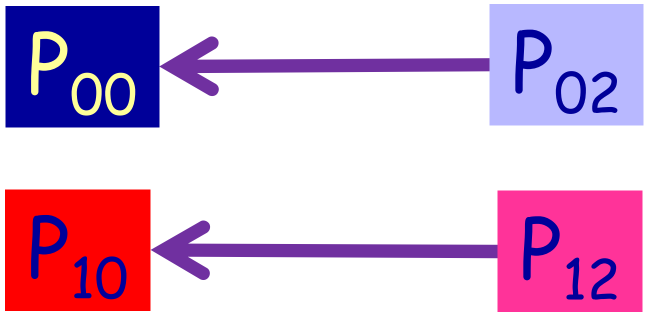
Collezione dei risultati

La scelta della strategia dipende da quanto vale p.

Se p è potenza di 2 allora conviene la II-III strategia, altrimenti (ed in particolare se è dispari) conviene usare la I strategia!!!

8=qxp=2x4   p=4

II strategia



II passo

$P_{02}\,P_{12}$ spediscono un vettore di N/q elementi a $P_{00}\,P_{10}$. I processori riceventi aggiornano con il proprio vettore effettuando N/q somme, cioè:

N/q tcom + N/q tcalc = N/2 tcom + N/2 tcalc

III Strategia: speed-up/efficienza (**def classica**)

Collezione dei risultati

matrice A: N righe, M colonne
Vettore b: M elementi

4=qxp=2x2 **p=4**

II strategia

A questo punto mi posso anche fermare, sapendo che il vettore soluzione è diviso a metà tra P_{00} e P_{10}

tcom = c tcalc

2<= c <= 3

In generale:

$$N/q \cdot \log_2(p)\ tcom + N/q \cdot \log_2(p)\ tcalc =$$

$$= c\ N/q \cdot \log_2(p)\ tcalc + N/q \cdot \log_2(p)\ tcalc$$

III Strategia: speed-up/efficienza (**def classica**)

MIMD-DM

GRIGLIA qxp

II strategia per collezione vettori

matrice A: N righe, M colonne

Vettore b: M elementi

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

$$= N[2M-1] /(N/q\ [\ 2M/p-1\ ] + cN/q\ \log_2(p) + N/q\ \log_2(p))$$

$$Oh = qxp\ T_{qxp}(NxM) - T_1(NxM) =$$

$$= pxq\ (N/q\ [\ 2M/p-1\ ] + cN/q\ \log_2(p) + N/q\ \log_2(p))\ t_{calc}$$

$$- N[2M-1]\ t_{calc}$$

$$E_{qxp}(NxM) = S_{qxp}(NxM) / (qxp)=$$

$$= N[2M-1] / [(qxp)(N/q\ [\ 2M/p-1\ ] + cN/q\ \log_2(p) + N/q\ \log_2(p))]$$

III Strategia: speed-up/efficienza (**def classica**)

Osservazioni:

- I conti fatti possono essere ripetuti in maniera del tutto simile per la III strategia della somma per collezionare i risultati
- I conti fatti hanno senso solo se si lascia il vettore “spezzato” lungo le righe

III STRATEGIA

Che cosa si può dire sull'implementazione

in ambiente **MIMD-SM** ?

È possibile pensare di lavorare con

BLOCCHI di RIGHE + COLONNE

della matrice *A*

e BLOCCHI di b.

Il tutto funziona come in ambiente distribuito sincronizzando gli accessi in memoria piuttosto che gestendo le comunicazioni

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

MIMD-SM

GRIGLIA qxp core

I strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

$$= N[2M-1] /(N/q\ [\ 2M/p-1\ ] + \cancel{cN/q\ (p-1)} + N/q\ (p-1))$$

$$Oh = qxp\ T_{qxp}(NxM) - T_1(NxM) =$$

$$= pxq\ (N/q\ [\ 2M/p-1\ ] + \cancel{cN/q\ (p-1)} + N/q\ (p-1)\ \ t_{calc} +$$

$$-\ N[2M-1]\ t_{calc}$$

$$E_{qxp}(NxM) = S_{qxp}(NxM)\ /\ (qxp) =$$

$$= N[2M-1]\ /\ [(qxp)(N/q\ [\ 2M/p-1\ ] + \cancel{cN/q\ (p-1)} + N/q\ (p-1)]$$

III Strategia: speed-up/efficienza (**def classica**)

MIMD-SM

GRIGLIA qxp core

II strategia per collezione vettori

matrice A: N righe, M colonne
Vettore b: M elementi

$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$

$= N[2M-1] \,/ (N/q\, [\, 2M/p-1\,] + \cancel{cN/q\, \log_2(p)} + N/q\, \log_2(p))$

$Oh = qxp\; T_{qxp}(NxM) - T_{qxp}(NxM) =$

$= pxq\; (N/q\, [\, 2M/p-1\,] + \cancel{cN/q\, \log_2(p)} + N/q\, \log_2(p))\;\; t_{calc}$

$- N[2M-1]\; t_{calc}$

$E_{qxp}(NxM) = S_{qxp}(NxM) \,/\, (qxp) =$

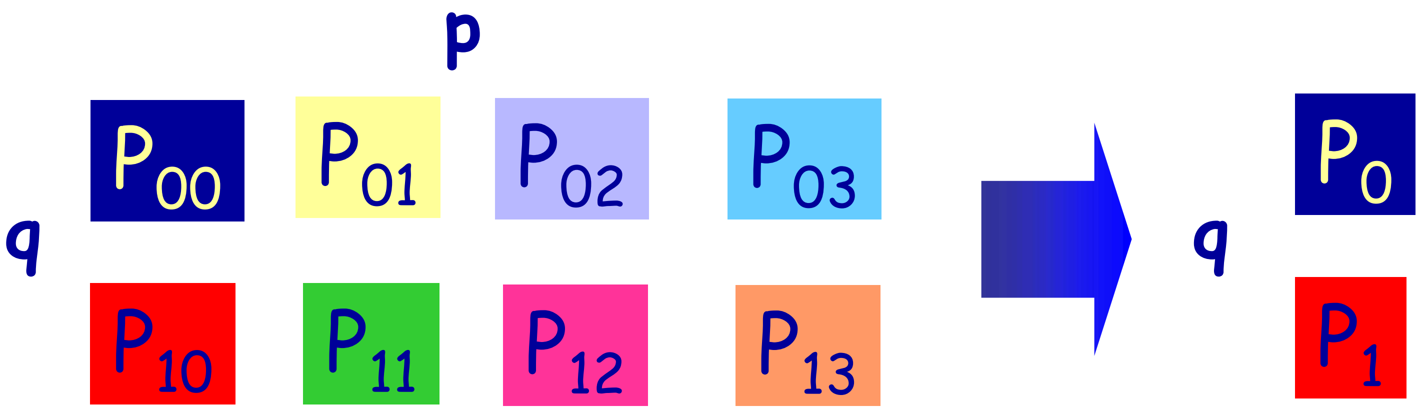
$= N[2M-1] \,/\, [(qxp)(N/q\, [\, 2M/p-1\,] + \cancel{cN/q\, \log_2(p)} + N/q\, \log_2(p))]$

Speed-up/efficienza (**def classica**)

MIMD-SM
MIMD-DM
GRIGLIA qxp

Attenzione:

- è possibile riconoscere la I strategia dalla III quando p=1

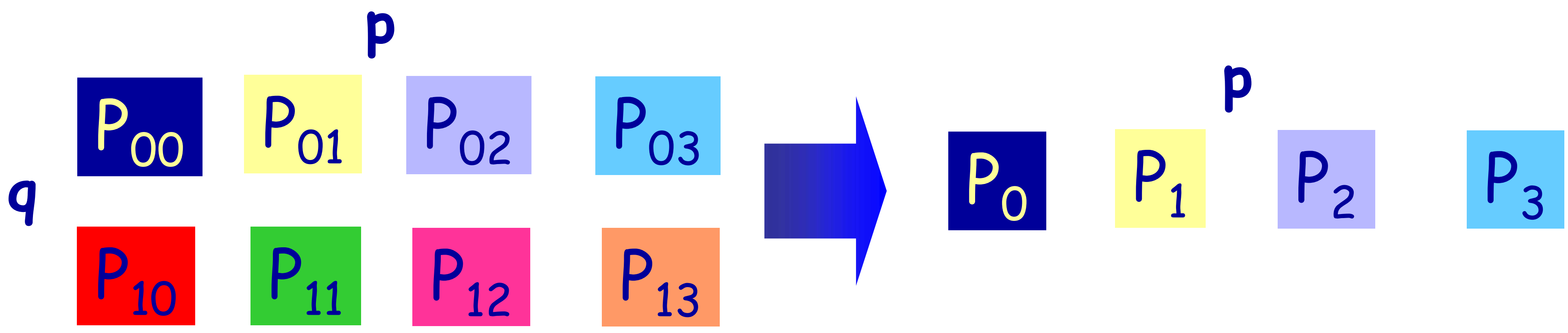


Speed-up/efficienza (**def classica**)

MIMD-SM
MIMD-DM
GRIGLIA qxp

Attenzione:

- è possibile riconoscere la I strategia dalla III quando p=1
- è possibile riconoscere la II strategia dalla III quando q=1



Speed-up/efficienza (**def classica**)

MIMD-SM
MIMD-DM

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

Attenzione:

- è possibile riconoscere la I strategia dalla III quando p=1
- è possibile riconoscere la II strategia dalla III quando q=1
- cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni processori avranno dei blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: RIGHE **mod(N,q)≠0**

Il numero di righe che avanza
(cioè il resto della divisione)
viene ridistribuito a tutti i processori riga che hanno prima coordinata strettamente minore del resto

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

GRIGLIA qxp

- **Attenzione:**

  cosa succede se **mod(N,q) ≠0** e/o **mod(M,p)≠0** ?

Alcuni processori avranno dei blocchi di matrice (e se serve del vettore) di dimensione maggiore

Es: RIGHE **mod(N,q)≠0**

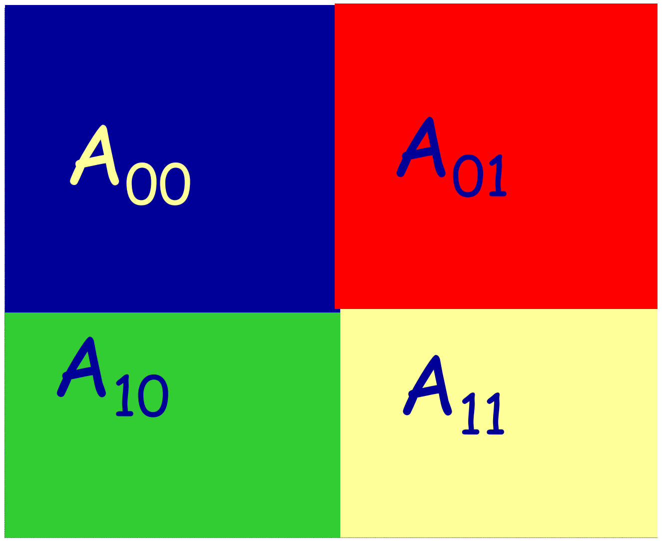
I processori che hanno coordinata riga strettamente minore del resto hanno una riga in più della matrice.
Nessuna variazione invece per il blocco relativo al vettore!

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(N,q) ≠0**

A



GRIGLIA
qxp=4=2x2

$\dim[b_{loc}] = M/p$

$\dim[A_{10}] = (N/q)\times(M/p)$

$\dim[A_{11}] = (N/q)\times(M/p)$

$\dim[A_{00}] = (N/q + 1)\times(M/p)$

$\dim[A_{01}] = (N/q + 1)\times(M/p)$

III Strategia: speed-up/efficienza (**def classica**)

matrice A: N righe, M colonne
Vettore b: M elementi

Es: se **mod(N,q) ≠0**

GRIGLIA
qxp=4=2x2

$\dim[A_{00}]=(N/q + 1)\times(M/p)$

$\dim[A_{01}]=(N/q + 1)\times(M/p)$

$\dim[b_{loc}]= M/p$

$\dim[A_{10}]=(N/q)\times(M/p)$

$\dim[A_{11}]=(N/q)\times(M/p)$

P_{00} P_{01}

P_{10} P_{11}

P_{00}: $\dim[r_{00}]=(N/q + 1)$, P_{01}: $\dim[s_{01}]=(N/q + 1)$

P_{10}: $\dim[r_{10}]=(N/q)$, P_{11}: $\dim[s_{11}]=(N/q)$

III Strategia: speed-up/efficienza (**def classica**)

MIMD-DM

GRIGLIA qxp

matrice A: N righe, M colonne
Vettore b: M elementi

I strategia per collezione vettori

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$
$$= N[2M-1] \,/\, (\boxed{[N/q+1]}\, [\, 2M/p-1\,] + \boxed{c(N/q+1)}\,(p-1) + \boxed{(N/q+1)}\,(p-1))$$

$$S_{qxp}(NxM) = T_1(NxM)/T_{qxp}(NxM) =$$

II strategia per collezione vettori

$$= N[2M-1] \,/\, (\boxed{[N/q+1]}\, [\, 2M/p-1\,] + c\,\boxed{(N/q+1)}\,\log_2(p) + \boxed{(N/q+1)}\,\log_2(p))$$

Osservazioni analoghe valgono per l'ambiente MIMD-SM,

i risultati sono gli stessi chiaramente <u>senza</u> i contributi relativi alle spedizioni