

# Parallelismo

## parallelismo spaziale

replicazione dell'hardware per processare più dati contemporaneamente;

## parallelismo temporale

divisione del processo in più stadi eseguibili contemporaneamente;

## grado di parallelismo $n$

numero di unità replicate nel parallelismo spaziale, di stadi in quello temporale;

## tempo di completamento $T_C^{(n)}$

tempo necessario per elaborare  $m$  input con grado di parallelismo  $n$ ;

## latenza

tempo necessario per produrre l'output corrispondente ad un input;

## tempo di servizio $T_S^{(n)}$

intervallo tra la generazione di due output consecutivi;

## tempo di servizio ideale $T_{|S|}^{(n)}$

$$T_{|S|}^{(n)} = \frac{T_S^{(1)}}{n}$$

## banda di elaborazione (throughput)

output generati per unità di tempo (inverso del tempo di servizio).

## scalabilità

$$S^{(n)} = \frac{T_S^{(1)}}{T_S^{(n)}} \text{ o } S_C^{(n)} = \frac{T_C^{(1)}}{T_C^{(n)}}, \text{ idealmente vale } n;$$

## speedup

$\frac{T_S^{\text{top}}}{T_S^{(n)}}$ , dove  $T_S^{\text{top}}$  è il tempo di servizio del miglior algoritmo sequenziale conosciuto (anziché utilizzare lo stesso algoritmo con grado di parallelismo 1 come nella scalabilità);

## tempo di interarrivo $T_A$

tempo che trascorre tra l'arrivo degli input; se è sufficientemente alto non conviene parallelizzare.