

# Medidas de Desempenho em Computação Paralela

Heitor Silvério Lopes

*hslopes@utfpr.edu.br*

# Motivação para medir desempenho

---

- Quantificar o ganho de tempo de processamento de um sistema paralelo
  - OBS. Medidas de complexidade computacional (de tempo e de memória) não são adequadas para algoritmos paralelos
  
- Avaliar tanto a ociosidade quanto o excesso de carga de processadores
  
- Comparar algoritmos diferentes executados sobre o mesmo problema

# Speedup

- ❑ Também conhecido como Fator de Aceleração
- ❑ É definido para um sistema paralelo homogêneo
- ❑ É definido como a razão entre a velocidade de processamento do sistema com um processador ( $T_1$ ) e do sistema paralelo com  $m$  processadores ( $T_m$ )
- ❑ Para algoritmos não-determinísticos, deve-se utilizar os tempos médios obtidos em  $n$  execuções independentes

$$s_m = \frac{T_1}{T_m}$$

Normalmente:  $1 \leq s_m \leq m$

## Speedup

---

- $s_m = 1$  significa que não houve nenhum ganho de tempo na versão paralela.  
 $s_m = p$  significa aceleração linear (ideal).
- Em alguns casos particulares é possível  $s_m > p$  (aceleração superlinear)
- Se  $s_m < 1$  a paralelização é inútil (aceleração sublinear)

# Taxonomia de medidas de *speedup*

- Proposto por Alba\*
- Strong Speedup (tipo I):
  - compara o tempo de processamento paralelo contra o algoritmo sequencial mais eficiente conhecido
- Weak Speedup (tipo II):
  - Compara o tempo do algoritmo paralelo desenvolvido por um pesquisador contra seu próprio algoritmo sequencial (supondo que ambos forneçam a mesma acurácia). Há duas variantes:
    - Versus Panmixia: quando se compara o tempo de um algoritmo paralelo contra o algoritmo sequencial canônico
    - Orthodox: é quando se compara o tempo de um algoritmo paralelo rodando em um processador, com o mesmo algoritmo em  $m$  processadores.
- Muito importante: só se pode comparar os tempos de processamento de algoritmos sequencial e paralelo que tenham a mesma qualidade (média) de soluções.

# Eficiência de execução

---

- ❑ Mede a fração de tempo que um processador é, de fato, utilizado.
- ❑ É definida como o *speedup* dividido pelo número de processadores

$$e_m = \frac{s_m}{m} = \frac{T_1}{T_m \cdot m}$$

- ❑ Idealmente deveria ser 1, mas os processadores também utilizam tempo para comunicação, alocação de memória, sistema operacional, etc, logo, na prática nunca chega a 1

## Fração serial

- Definida por Karp & Flatt\* e mede a proporção do algoritmo que é inerentemente sequencial

$$f_m = \frac{1/s_m - 1/m}{1 - 1/m}$$

- Idealmente deve ser constante para um algoritmo para diferentes valores de  $m$
- Um aumento de  $f_m$  à medida que  $m$  aumenta indica que a granularidade do paralelismo é muito fina
- Se  $f_m$  diminuir à medida que  $m$  aumentar indica que o *speedup* está se tornando superlinear

## Custo

---

- Soma do tempo que cada processador utilizou
- É o equivalente a um processador sequencial hipotético com o desempenho equivalente à soma dos processadores em paralelo

$$C_m = m.T_m$$



# Taxa de ocupação

- Inclui todo o tempo gasto por um processador para executar o algoritmo, levando em consideração não só o processamento, mas também a comunicação
- É calculado individualmente para cada processador  $i$

$$T_{ocup}[\%] = \frac{T_i \cdot 100}{T_m}$$

- Se  $T_{ocup} = 0$  o processador não participou do processamento.  $T_{ocup} = 1$  somente se houver um único processador (sistema sequencial). Quando o balanço de carga é perfeito,  $T_{ocup}$  se aproxima de 100%, portanto é uma boa medida de balanceamento de carga de um sistema de processamento paralelo

## Outras métricas disponíveis

---

- Fator de Utilização
- Taxa de execução
- Taxa de participação
- Taxa de comunicação