APPROXIMATION ALGORITHM FOR SCHEDULING A CHAIN OF TASKS ON HETEROGENEOUS SYSTEMS

Massinissa Ait aba

CONTEXT AND MOTIVATIONS

- **Increasing massive calculations**
  - HPC : Physical calculations, medical data, weather
  - Data center : Google search, Facebook

- **Significant increase in energy consumption**
  - Optimization of the ratio performance / watt

- **Heterogeneous micro-server**
  - Several execution resources : CPU, GPU, FPGA,…
  - Different properties : execution frequencies, energy consumption

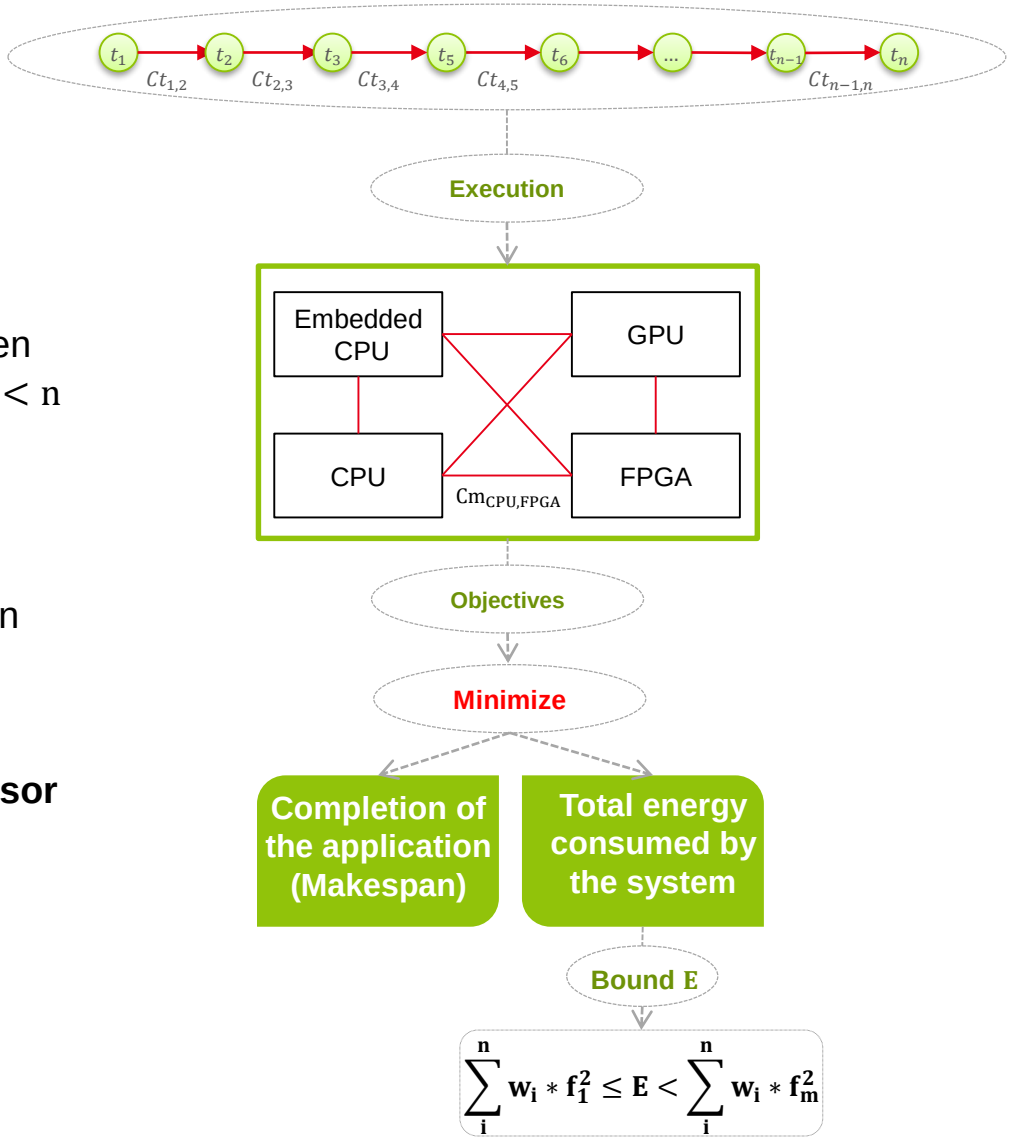
- **Platform : RECS@lBox Compute Unit 3.0 (Antares)**

HETEROGENEOUS PLATFORM SCHEDULING

1. **Notation**
2. **State of the art**
3. **Problem modeling**
4. **Methods of resolution**
   - Preemptive scheduling
   - Non preemptive scheduling
   - Approximation ratio
5. **Numerical results**
6. **Conclusion and perspectives**

NOTATION

- **Data**
  - **Tasks**
    - $T = \{t_1, t_2, \ldots, t_n\}$, $|T| = n$
    - $W = \{w_1, w_2, \ldots, w_n\}$
    - $Ct_{i,i+1}$: cost of communication between each pairs of tasks t_i and t_{i+1}, $i < n$
  - **Platform**
    - $P = \{p_1, p_2, \ldots, p_m\}$, $|P| = m$
    - $F = \{f_1, f_2, \ldots, f_m\}$, $f_1 \leq f_2 \leq \ldots \leq f_m$
    - $Cm_{k,l}$: cost of communication between each pair of processors p_k and p_l
  - **Execution cost of task t_i on p_j processor**
    - Execution time: $execut_{i,j} = w_i / f_j$
    - Energy: $e_{i,j} = w_i * f_j^2$



STATE OF THE ART

Articles	Machines environment			criteria		Approximation Ratio	Methods of Resolution
	HCS	PREC	DVFS	Finish execution Makespan	Energy		
Zhong, Xiliang, Cheng (2007)			√	constraint	objective		Model + Heuristics
Lee, Young, Zomaya (2009)	√	√	√	objective	objective		Heuristic
Young, Pasricha, Maciejewski (2013)	√	√	√	constraint	constraint		Heuristics
Longxin, Kenli, Li (2016)	√	√		constraint	objective		Genetic algorithm
Perez Vasquez (2014)	√	√	√	objective	objective		Pareto heuristic/ Game Theory
Tarplee, Friese, Maciejewski (2016)	√			objective	objective		
Xie, Xiao, Li (2016)	√	√	√	objective	constraint		Heuristics
Aupy, Benoit,Dufossé, Robert (2013)	√	√	√	constraint	objective	√	Heuristics and exact algorithms for particular cases
Our work	√	√		objective	constraint	√	Approximation algorithm

- HCS : Heterogeneous Computer Systems
- PREC : Constraints of precedence
- DVFS : Dynamic Voltage Frequency Scaling

PROBLEM MODELING

- **Variables**
  - $start_i$ = starting time of the task t_i
  - $x_{i,j} = \begin{cases} 1 & \text{if task } t_i \text{ is placed on the processing element } p_j \\ 0 & \text{otherwise} \end{cases}$

- **Constraints**

Each task must be executed once

Constraint on total energy consumed

Constraint of precedence

$$(P) \begin{cases} \sum_{j=1}^{m} x_{i,j} = 1 \quad \forall i = \overline{1..n} \\ \sum_{i=1}^{n}\sum_{j=1}^{m} x_{i,j} * e_{i,j} \leq E \text{ , with } \sum_{i}^{n} w_i * f_1^2 \leq E < \sum_{i}^{n} w_i * f_m^2 \\ start_i + x_{i,j_1} * execut_{i,j_1} + x_{i,j_1} * x_{i+1,j_2}\left(Ct_{i,i+1} + Cm_{j_1,j_2}\right) \leq start_{i+1} \\ \forall j_1 = \overline{1..m}, \forall j_2 = \overline{1..m}, j_1 \neq j_2 \end{cases}$$

- **Objective**
  - Minimize (Makespan): $Z(min) = start_n + \sum_{j=1}^{m} x_{n,j} * execut_{n,j}$

METHODS OF RESOLUTION: SOLVER

- **Mixed Integer Quadratic Constrained Program (MIQCP)**

$$
(P)\begin{cases}
Z(\min) = \mathrm{start}_n + \sum_{j=1}^{m} x_{n,j} * \mathrm{execut}_{n,j} \\
\sum_{j=1}^{m} x_{i,j} = 1 \quad \forall i = \overline{1..n} \\
\sum_{i=1}^{n}\sum_{j=1}^{m} x_{i,j} * e_{i,j} \leq E \quad , \text{with} \sum_{i}^{n} w_i * f_1^2 \leq E < \sum_{i}^{n} w_i * f_m^2 \\
\mathrm{start}_i + x_{i,j_1} * \mathrm{execut}_{i,j_1} + x_{i,j_1} * x_{i+1,j_2}\left(\mathrm{Ct}_{i,i+1} + \mathrm{Cm}_{j_1,j_2}\right) \leq \mathrm{start}_{i+1} \\
\forall j_1 = \overline{1..m}, \forall j_2 = \overline{1..m}, j_1 \neq j_2 \\
x_{i,j} \in \{0,1\} \quad , \mathrm{start}_i \geq 0 \ , i = \overline{1..n} \ , j = \overline{1..m}
\end{cases}
$$

- **Solver**
  - Implementation in C++ of the model (P)
  - Data modeling (application and platform)
  - Cplex


Exact method

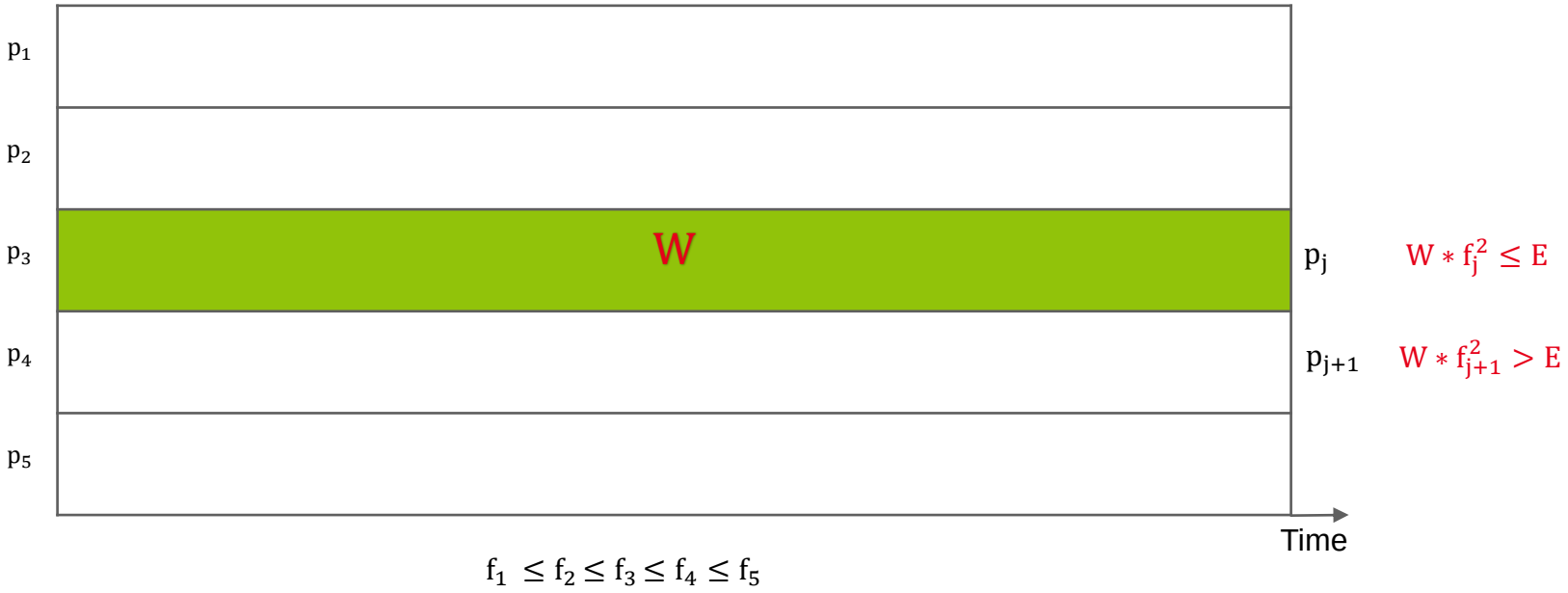

Quadratic constraints


Scaling up difficult

METHODS OF RESOLUTION: 1) PREEMPTIVE SCHEDULING (PS)

- **Find the processing element p_j with $j = \max\{l \in \{1..m\}, \sum_{i=1}^{n} w_i * f_l \leq E\}$**

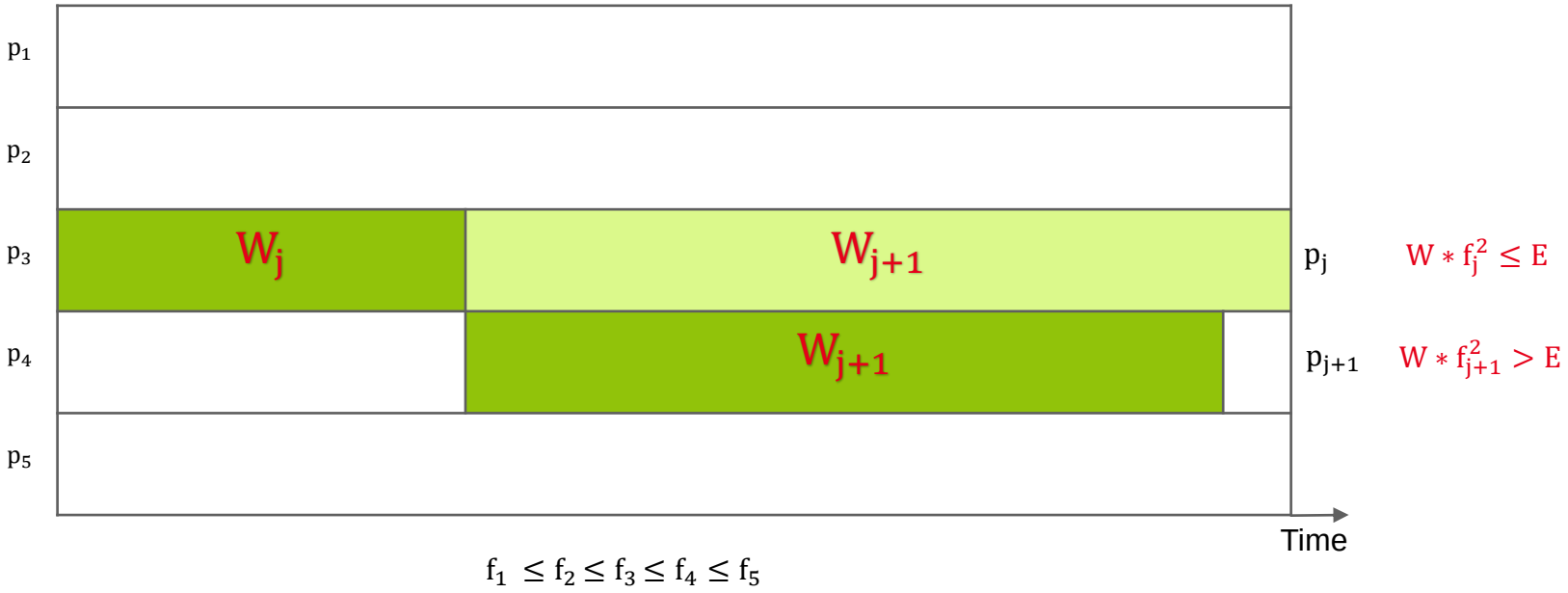
$$W = \sum_{i=1}^{n} w_i$$



METHODS OF RESOLUTION: 1) PREEMPTIVE SCHEDULING (PS)

- **Saturate the energy constraint**

$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$



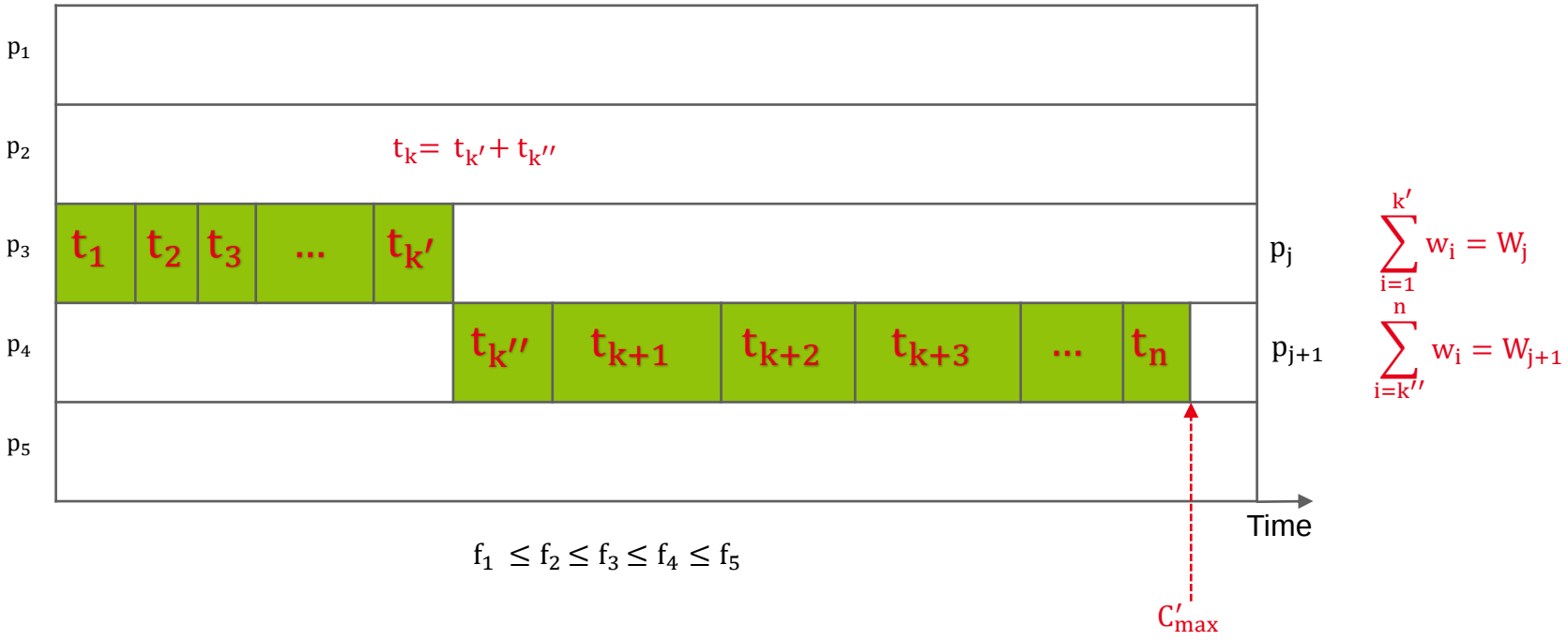
$$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$$

$$Energy = W_j * f_j^2 + W_{j+1} * f_{j+1}^2 = E$$

METHODS OF RESOLUTION: 1) PREEMPTIVE SCHEDULING (PS)

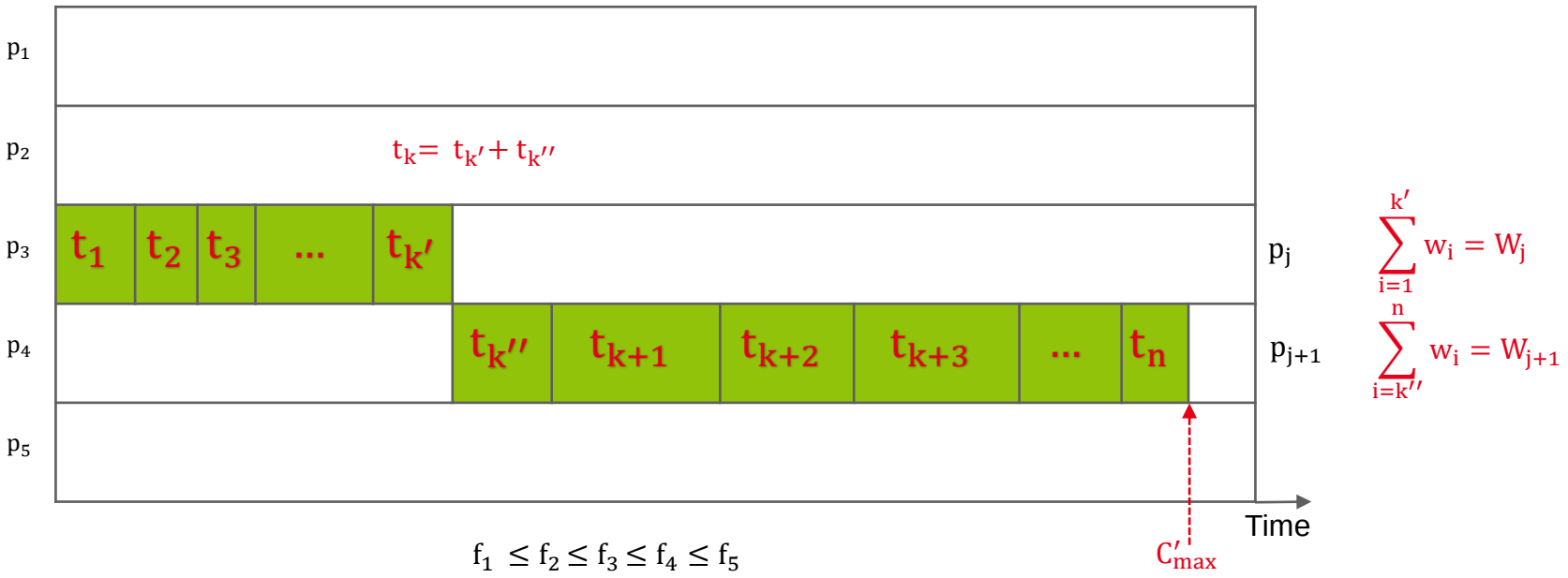
- **Distribute the tasks on both processing elements p_j and p_{j+1}**

$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$



$$C'_{max} = \frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}$$

METHODS OF RESOLUTION: 1) PREEMPTIVE SCHEDULING (PS)



Lemma1: The set S of schedules that saturate energy constraint is dominant

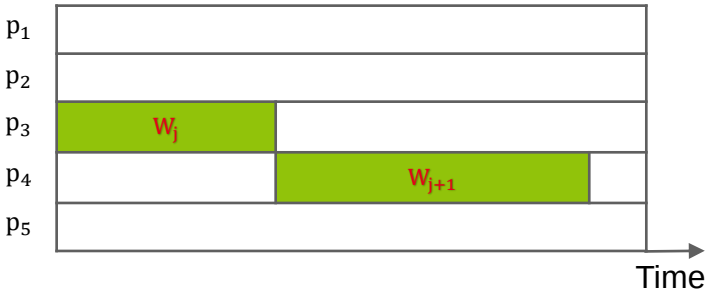
Lemma2: For any schedule s ∈ S, $C^s_{max} \geq C'_{max}$

Theorem: *Algorithm PS provides the optimal solution for non−preemptive scheduling*

$$C'_{max} = \frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}$$

METHODS OF RESOLUTION: 2) NON-PREEMPTIVE SCHEDULING

- **Transform the solution of the preemptive scheduling to obtain a solution realizable for the non-preemptive scheduling**



$$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$$

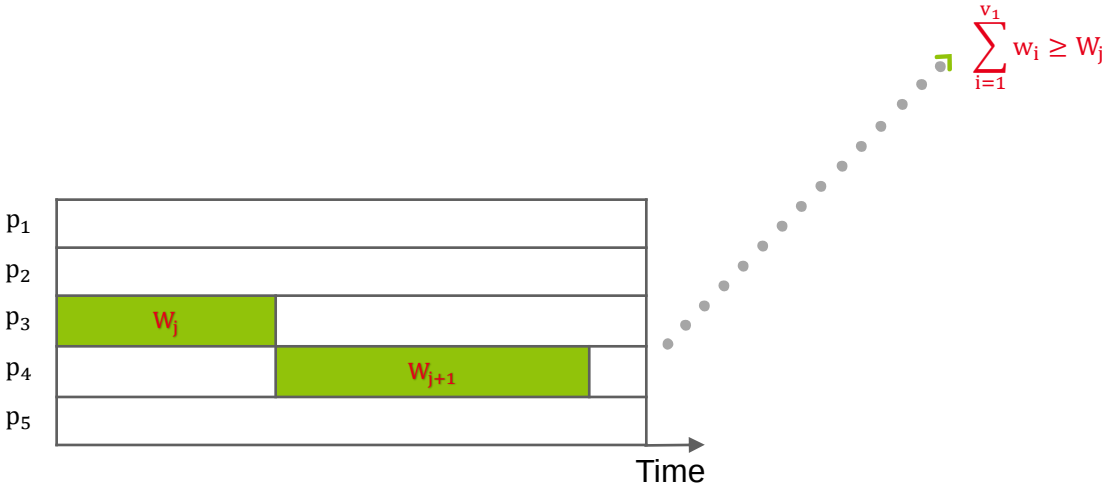
$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$

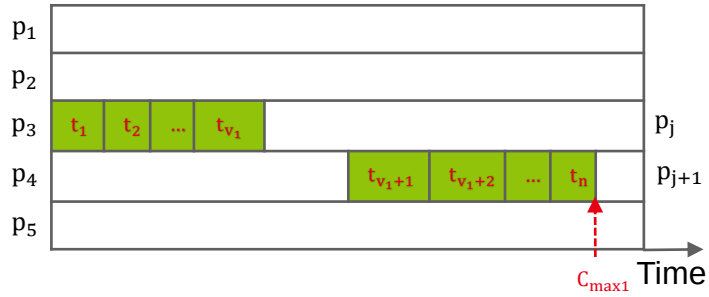
$$\text{Energy} = W_j * f_j^2 + W_{j+1} * f_{j+1}^2 = E$$

METHODS OF RESOLUTION: 2) NON-PREEMPTIVE SCHEDULING

- **Transform the solution of the preemptive scheduling to obtain a solution realizable for the non-preemptive scheduling**
  - Case 1: execute tasks t_1 to t_{v_1} on p_j then the rest on p_{j+1}

$$v_1 \in \{1..n\} \text{ ie } \widehat{C_{max1}} = \min\{\frac{\sum_{i=1}^{v_1} w_i}{f_j} + \frac{\sum_{i=v_{1+1}}^{n} w_i}{f_{j+1}} + Ct_{v_1,v_{1+1}} + Cm_{j,j+1}, \sum_{i=1}^{v_1} w_i \geq W_j\}$$





$$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$$

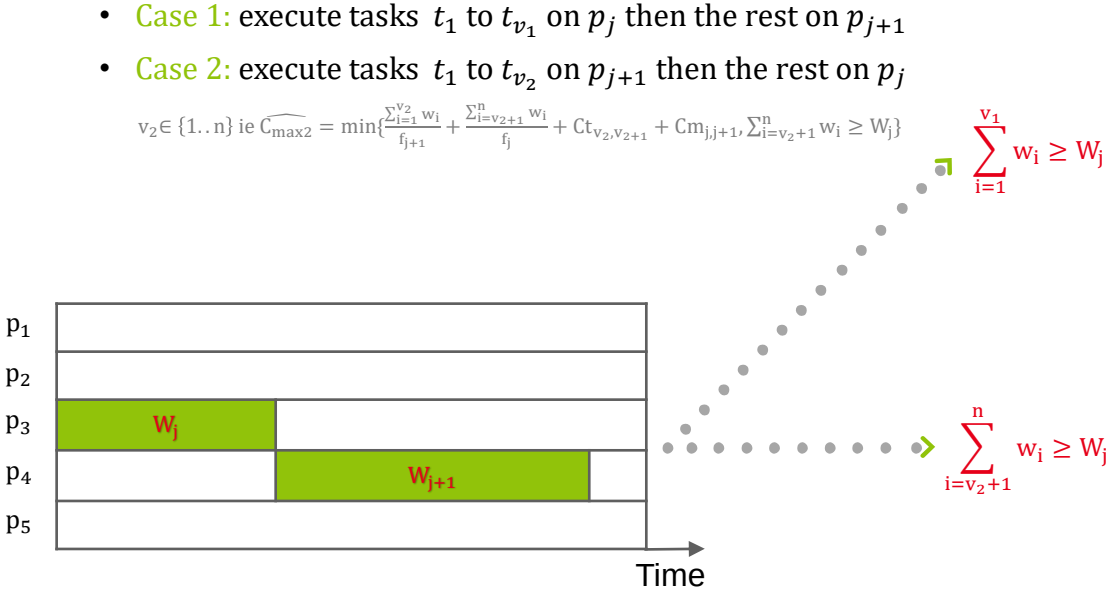
$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$

$$\text{Energy} = W_j * f_j^2 + W_{j+1} * f_{j+1}^2 = E$$

METHODS OF RESOLUTION: 2) NON-PREEMPTIVE SCHEDULING

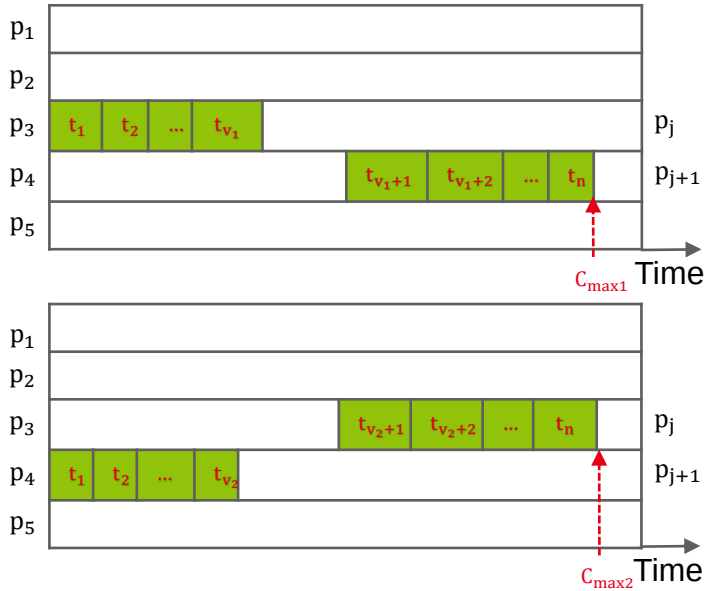
- **Transform the solution of the preemptive scheduling to obtain a solution realizable for the non-preemptive scheduling**
  - Case 1: execute tasks t_1 to t_{v_1} on p_j then the rest on p_{j+1}
  - Case 2: execute tasks t_1 to t_{v_2} on p_{j+1} then the rest on p_j

$$v_2 \in \{1..n\} \text{ ie } \widehat{C_{max2}} = \min\{\frac{\sum_{i=1}^{v_2} w_i}{f_{j+1}} + \frac{\sum_{i=v_2+1}^{n} w_i}{f_j} + Ct_{v_2,v_2+1} + Cm_{j,j+1}, \sum_{i=v_2+1}^{n} w_i \geq W_j\}$$



$$\sum_{i=1}^{v_1} w_i \geq W_j$$

$$\sum_{i=v_2+1}^{n} w_i \geq W_j$$



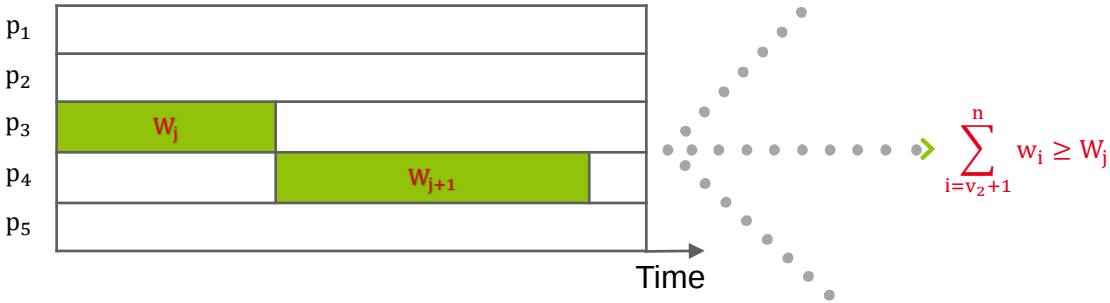
$$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$$

$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$

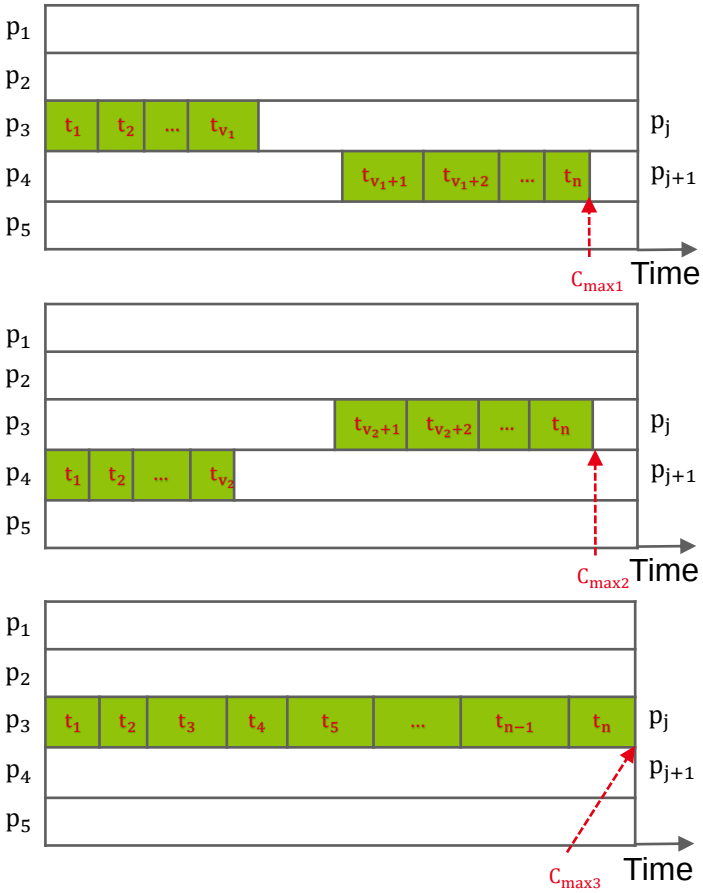
$$\text{Energy} = W_j * f_j^2 + W_{j+1} * f_{j+1}^2 = E$$

METHODS OF RESOLUTION: 2) NON-PREEMPTIVE SCHEDULING

- **Transform the solution of the preemptive scheduling to obtain a solution realizable for the non-preemptive scheduling**
  - Case 1: execute tasks t_1 to t_{v_1} on p_j then the rest on p_{j+1}
  - Case 2: execute tasks t_1 to t_{v_2} on p_{j+1} then the rest on p_j
  - Case 3: execute all tasks on p_j



$$\sum_{i=1}^{v_1} w_i \geq W_j \qquad \sum_{i=v_2+1}^{n} w_i \geq W_j$$



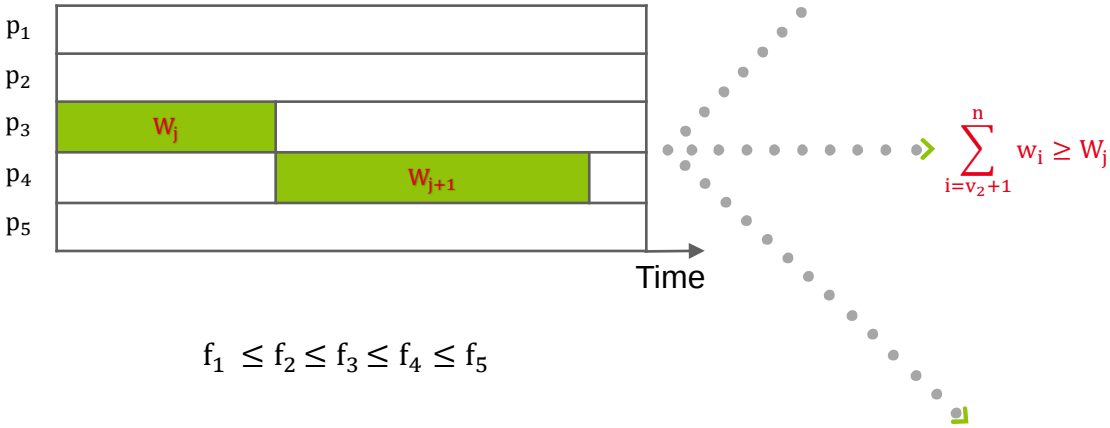
$$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$$

$$W = \sum_{i=1}^{n} w_i = W_j + W_{j+1}$$

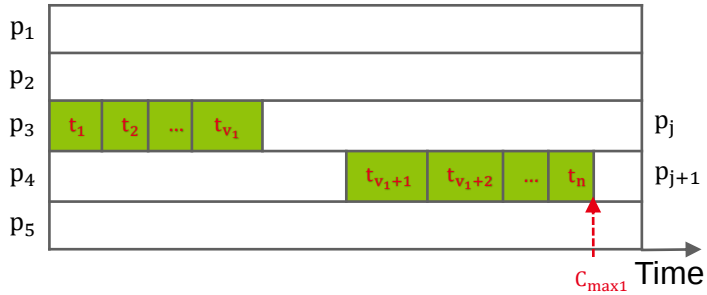
$$\text{Energy} = W_j * f_j^2 + W_{j+1} * f_{j+1}^2 = E$$

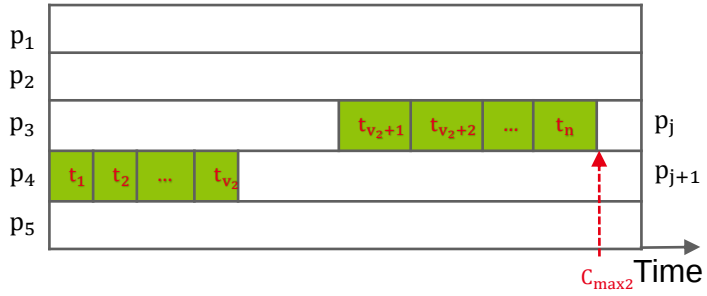
METHODS OF RESOLUTION: 2) NON-PREEMPTIVE SCHEDULING

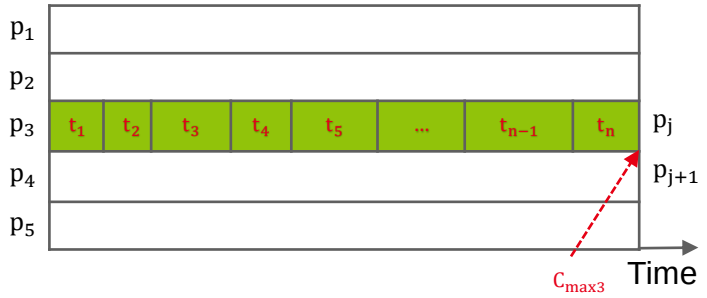
- **Transform the solution of the preemptive scheduling to obtain a solution realizable for the non-preemptive scheduling**
  - Case 1: execute tasks t_1 to t_{v_1} on p_j then the rest on p_{j+1}
  - Case 2: execute tasks t_1 to t_{v_2} on p_{j+1} then the rest on p_j
  - Case 3: execute all tasks on p_j



$f_1 \leq f_2 \leq f_3 \leq f_4 \leq f_5$







$$C_{max} = \min\{C_{max1}, C_{max2}, C_{max3}\}$$

METHODS OF RESOLUTION: 2) APPROXIMATION RATIO

- $C^{\star}_{max}$: optimal solution for non-preemptive scheduling
- C_{max}: solution obtained by the approximate method (NPS)
- C'_{max}: optimal solution for preemptive scheduling (PS), $C'_{max} = \frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}$

Theorem: the approximation ratio in the worst case is given by $\frac{C_{max}}{C^{\star}_{max}} \leq \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}} < \frac{f_{j+1}}{f_j}$

METHODS OF RESOLUTION: 2) APPROXIMATION RATIO

- $C^{\star}_{max}$: optimal solution for non-preemptive scheduling
- C_{max}: solution obtained by the approximate method (NPS)
- C'_{max}: optimal solution for preemptive scheduling (PS), $C'_{max} = \frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}$

Theorem: the approximation ratio in the worst case is given by $\frac{C_{max}}{C^{\star}_{max}} \leq \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}} < \frac{f_{j+1}}{f_j}$

In the worse case, we execute all tasks on the processing element p_j : $C_{max} \leq \frac{W}{f_j}$ with $W = \sum_{i=1}^{n} w_i$

$$\frac{C_{max}}{C'_{max}} \leq \frac{\frac{W}{f_j}}{\frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}} = \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}}$$

$$C'_{max} \leq C^{\star}_{max} \longrightarrow \frac{C_{max}}{C^{\star}_{max}} \leq \frac{C_{max}}{C'_{max}} \longrightarrow \frac{C_{max}}{C^{\star}_{max}} \leq \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}}$$

METHODS OF RESOLUTION: 2) APPROXIMATION RATIO

- $C^{\star}_{max}$: optimal solution for non-preemptive scheduling
- C_{max}: solution obtained by the approximate method (NPS)
- C'_{max}: optimal solution for preemptive scheduling (PS), $C'_{max} = \frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}$

Theorem: the approximation ratio in the worst case is given by $\frac{C_{max}}{C^{\star}_{max}} \leq \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}} < \frac{f_{j+1}}{f_j}$

In the worse case, we execute all tasks on the processing element p_j : $C_{max} \leq \frac{W}{f_j}$ with $W = \sum_{i=1}^{n} w_i$

$$\frac{C_{max}}{C'_{max}} \leq \frac{\frac{W}{f_j}}{\frac{W_j}{f_j} + \frac{W_{j+1}}{f_{j+1}}} = \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}}$$

$$C'_{max} \leq C^{\star}_{max} \longrightarrow \frac{C_{max}}{C^{\star}_{max}} \leq \frac{C_{max}}{C'_{max}} \longrightarrow \frac{C_{max}}{C^{\star}_{max}} \leq \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}}$$

$$\frac{f_j}{f_{j+1}} < 1 \longrightarrow \frac{W}{W_j + \frac{f_j * W_{j+1}}{f_{j+1}}} < \frac{W}{\frac{f_j}{f_{j+1}} * (W_j + W_{j+1})} = \frac{f_{j+1}}{f_j}$$

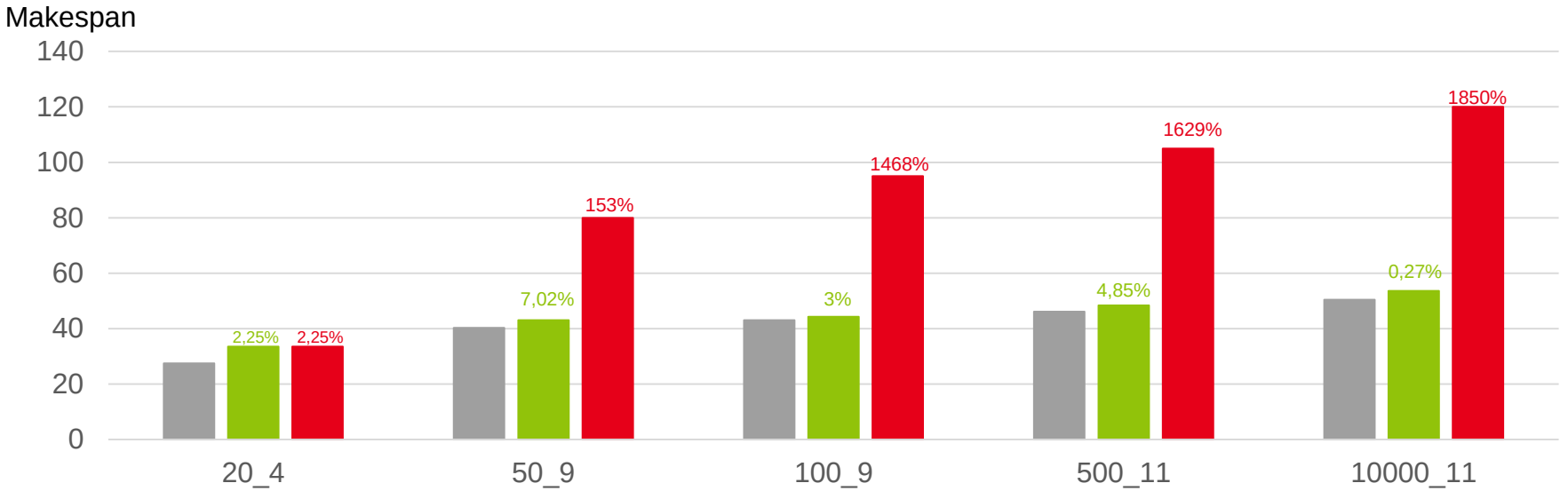
NUMERICAL RESULTS

- Implementation in C++ of PS and NPS algorithm
- Solver Cplex to solve the model (P)
- **30 instances generated randomly for 3 different sizes, small, medium and large**

- preemptive scheduling solution
- Non-preemptive scheduling solution
- Cplex solution (time limit = 1h)

Number of tasks	Number of machines
20	4
50	9
100	9
500	11
10000	11

- **Results**



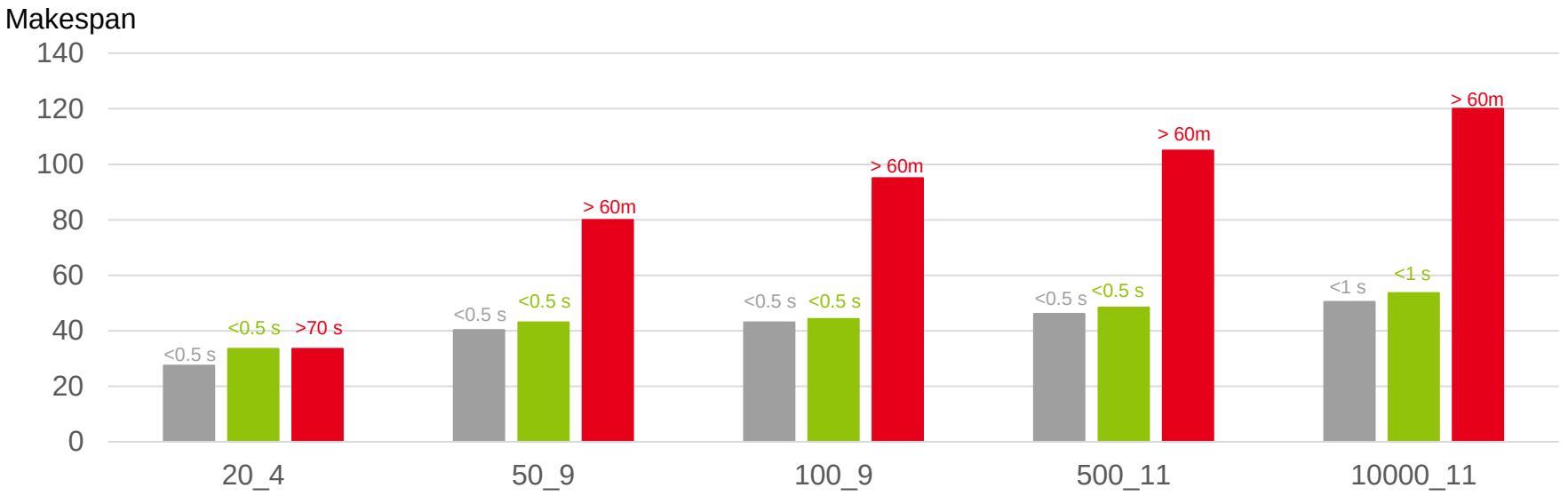
NUMERICAL RESULTS

- Implementation in C++ of PS and NPS algorithm
- Solver Cplex to solve the model (P)
- **30 instances generated randomly for 3 different sizes, small, medium and large**

- preemptive scheduling solution
- Non-preemptive scheduling solution
- Cplex solution (time limit = 1h)

Number of tasks	Number of machines
20	4
50	9
100	9
500	11
10000	11

- **Run time**



CONCLUSION AND PERSPECTIVES

- **Scheduling problem on heterogeneous platform**
  - Optimization of the performance/watt ratio

- **Scheduling a chain of tasks**
  - Preemptive Scheduling: optimal solution
  - Non Preemptive Scheduling: approximate solution with a guaranteed performance

- **Perspectives**
  - Extension to more general classes of graphs (DAG)
  - Tests on real applications and real platform (RECS)

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA SACLAY NANO-INNOV | BAT. 861 – PC142
91191 Gif-sur-Yvette Cedex - FRANCE
www-list.cea.fr

Établissement public à caractère industriel | RCS Paris B 775 685 019