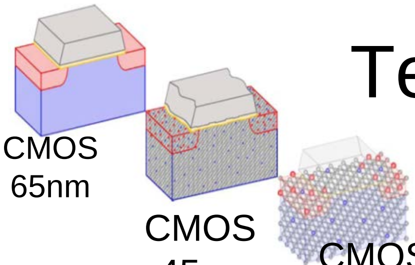
Many-core thermal management and design

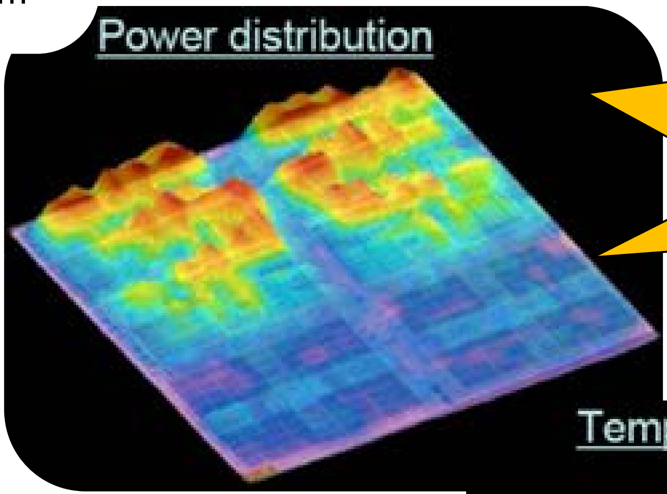
Andrea Bartolini, Matteo Cacciari, Michele Lombardi, Mohammad Sadegh Sadri, Francesco Beneventi

Background



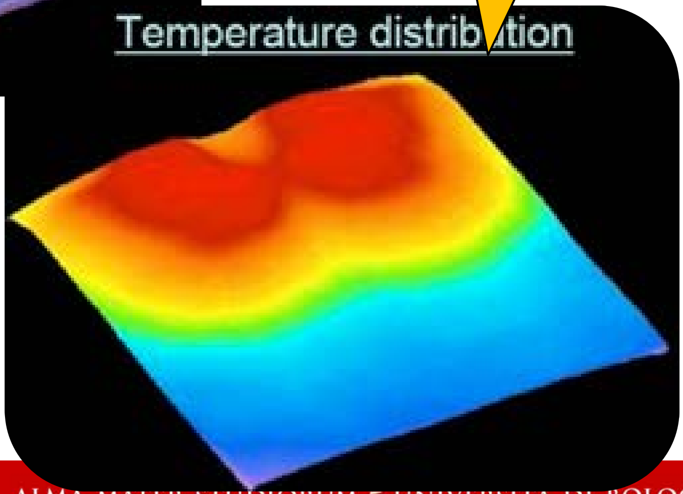
Technology scaling

High power densities

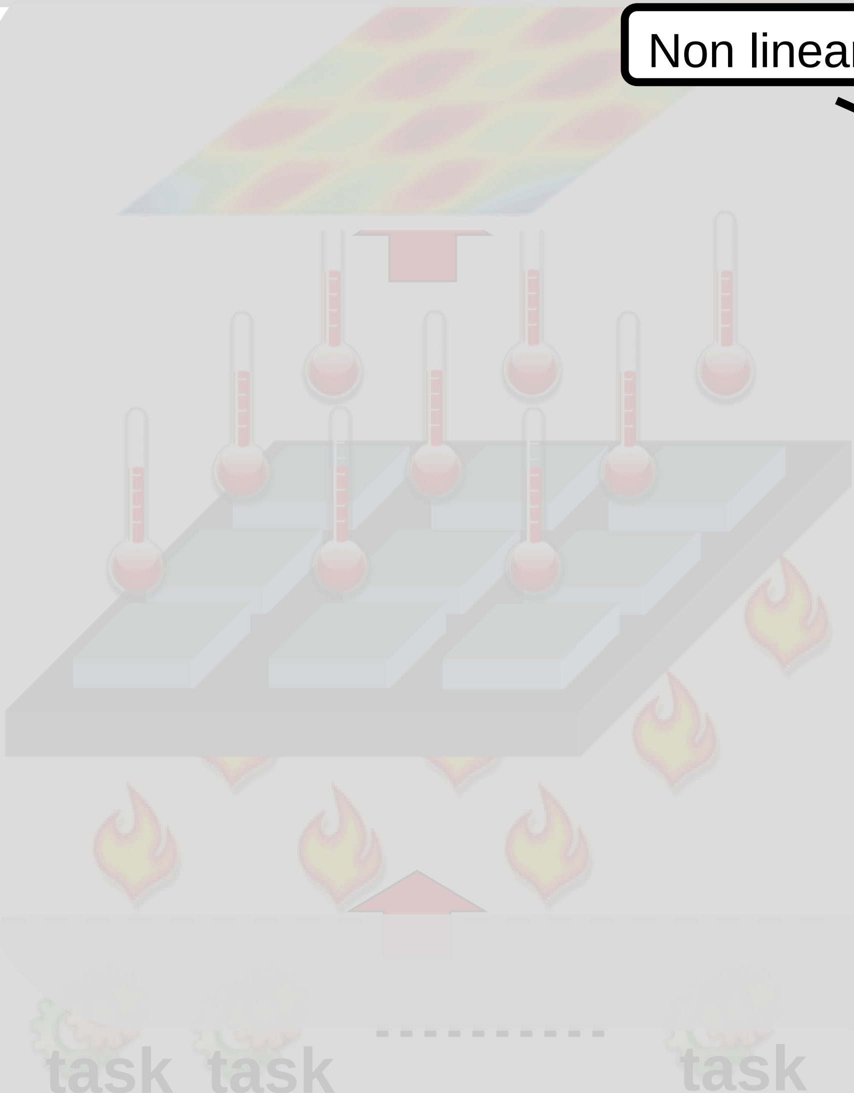


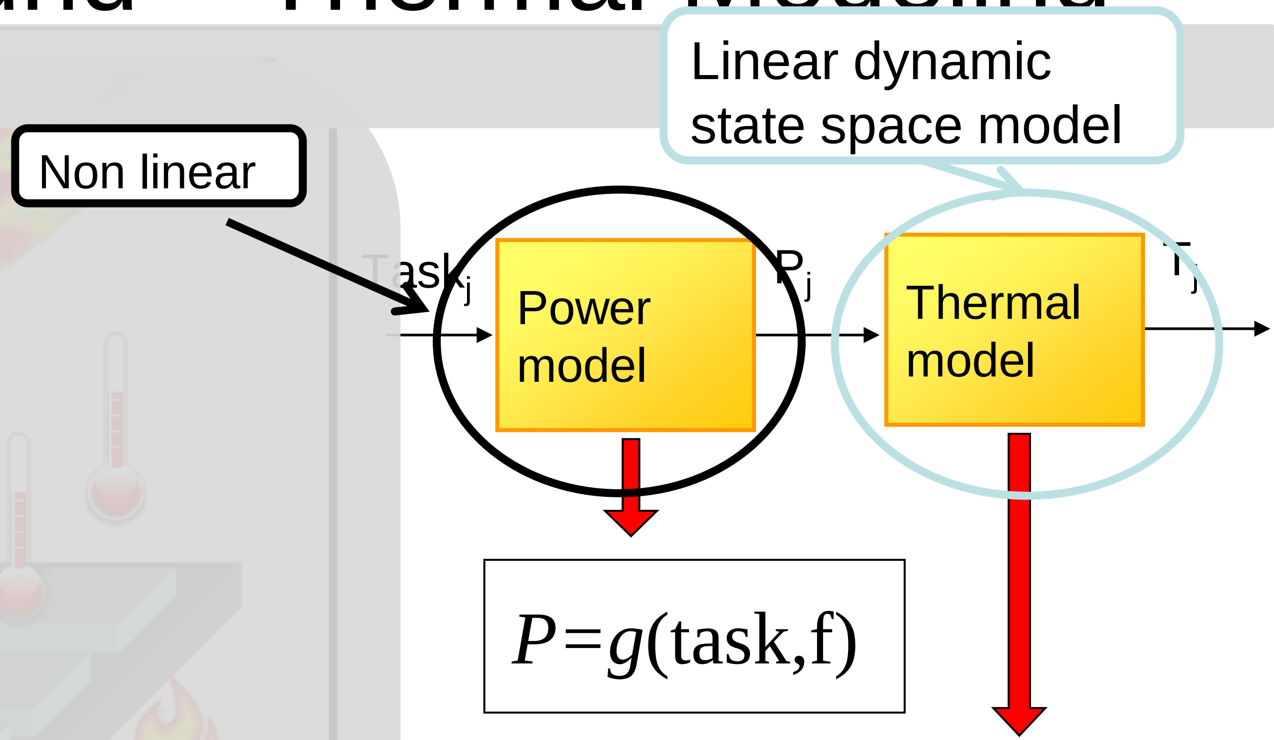
Thermal issues:
hot-spots, thermal gradients…

Non uniform thermal map



Background – Thermal Modeling





$$P = g(\text{task}, f)$$

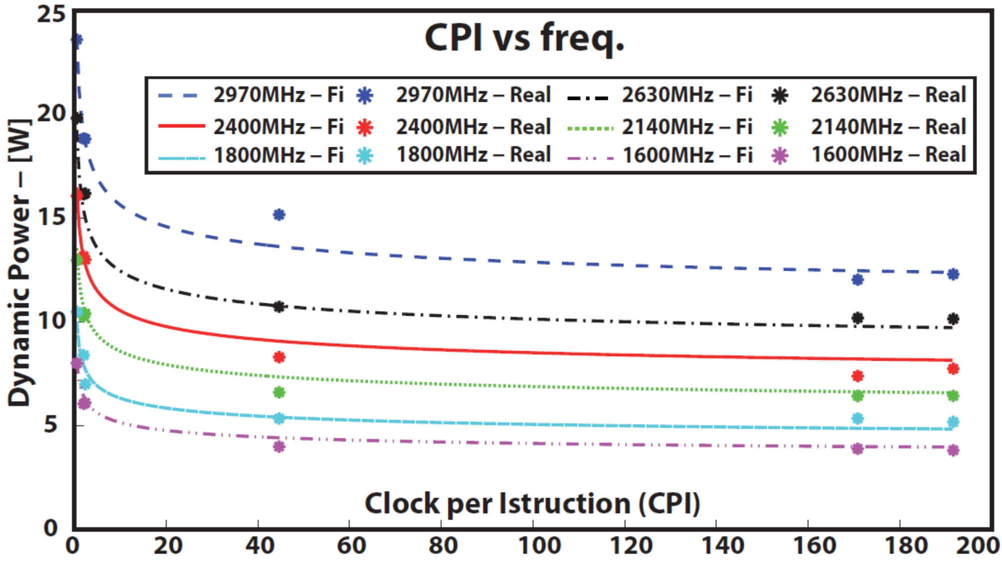
$$x_{n+1} = Ax_n + BP_{n,j}$$

$$T_{n,j} = Cx_n$$

Background Power

Power

- Platform:
  - Intel server system S7000FC4UR
  - 16 cores - 4 quad cores Intel® Xeon® X7350, 2.93GHz
- *At the wall Power consumption*
  - test:
    - set of synthetic benchmarks with different memory pattern accesses
    - forcing all the cores to run at different performance levels
    - for each benchmark we extract the clocks per instruction metrics (CPI) and correlate it with the power



Power is function of frequency and workload properties

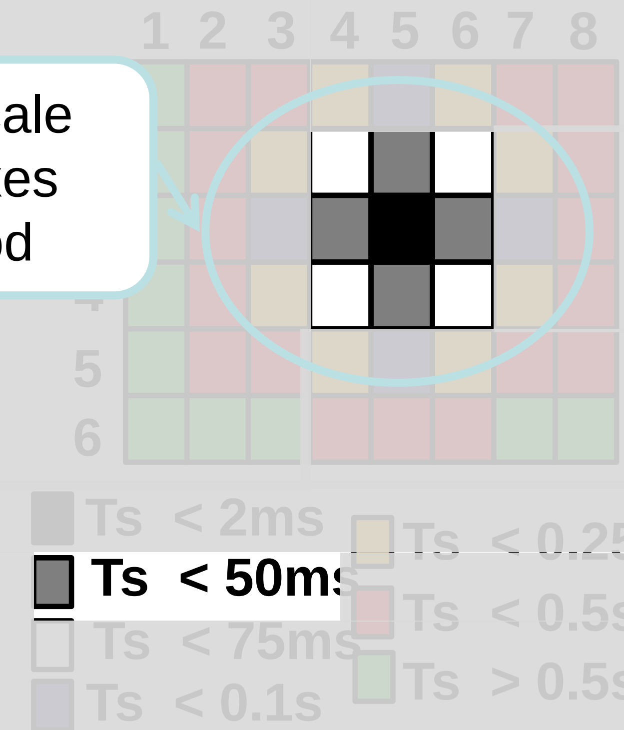
$$P_D = k_A \cdot V_{DD}^2 \cdot f_{CK} + k_B + (k_C + k_D \cdot f_{CK}) \cdot CPI^{k_E}$$

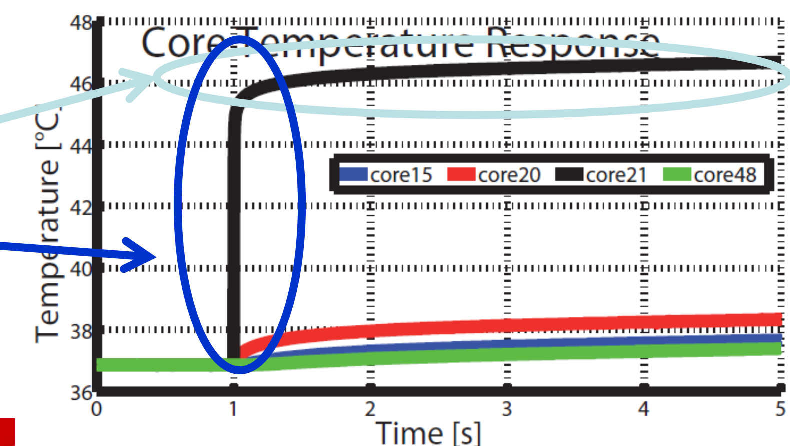
Background – Thermal transient

Thermal locality (Direct Fourier law implication):

- Continuous model:
  - Thermal neighborhood = Physical
- Discrete model:
  - Thermal neighborhood depends on sample time
- Hotspot simulation of 'Intel SCC like' 48core
  - Each core : Area = 11.82mm2, Pmax = 2.6W
  - We powered on only Core(5,3)
  - T neighborhood > +0.1°C
- Thermal transient – Model Order
  - Different building materials reflects in different time constants [1]
    - Silicon die, heat spreader, heat sink
    - Second order model

In O.S time scale
=> cardinal axes
Neighbourhood

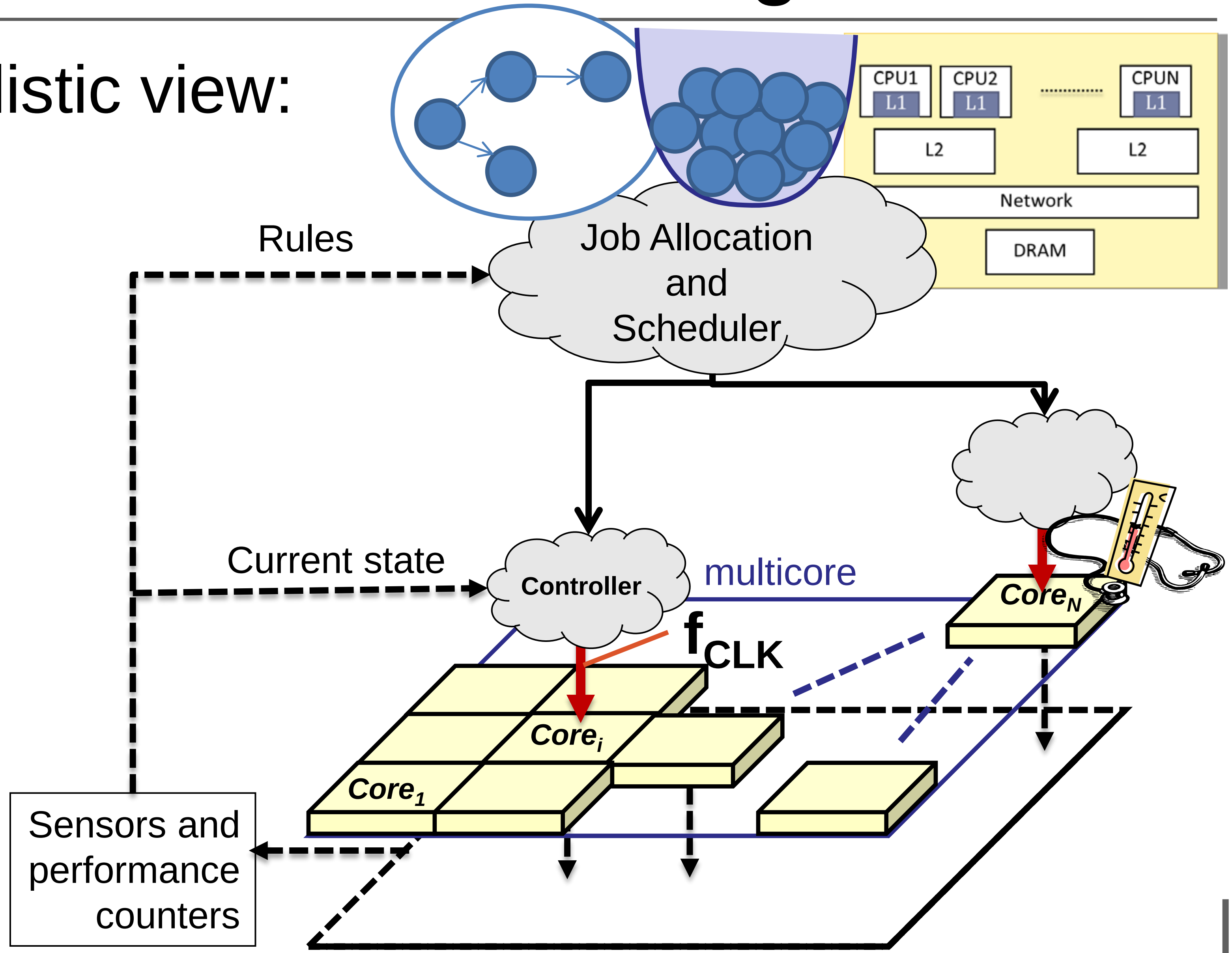




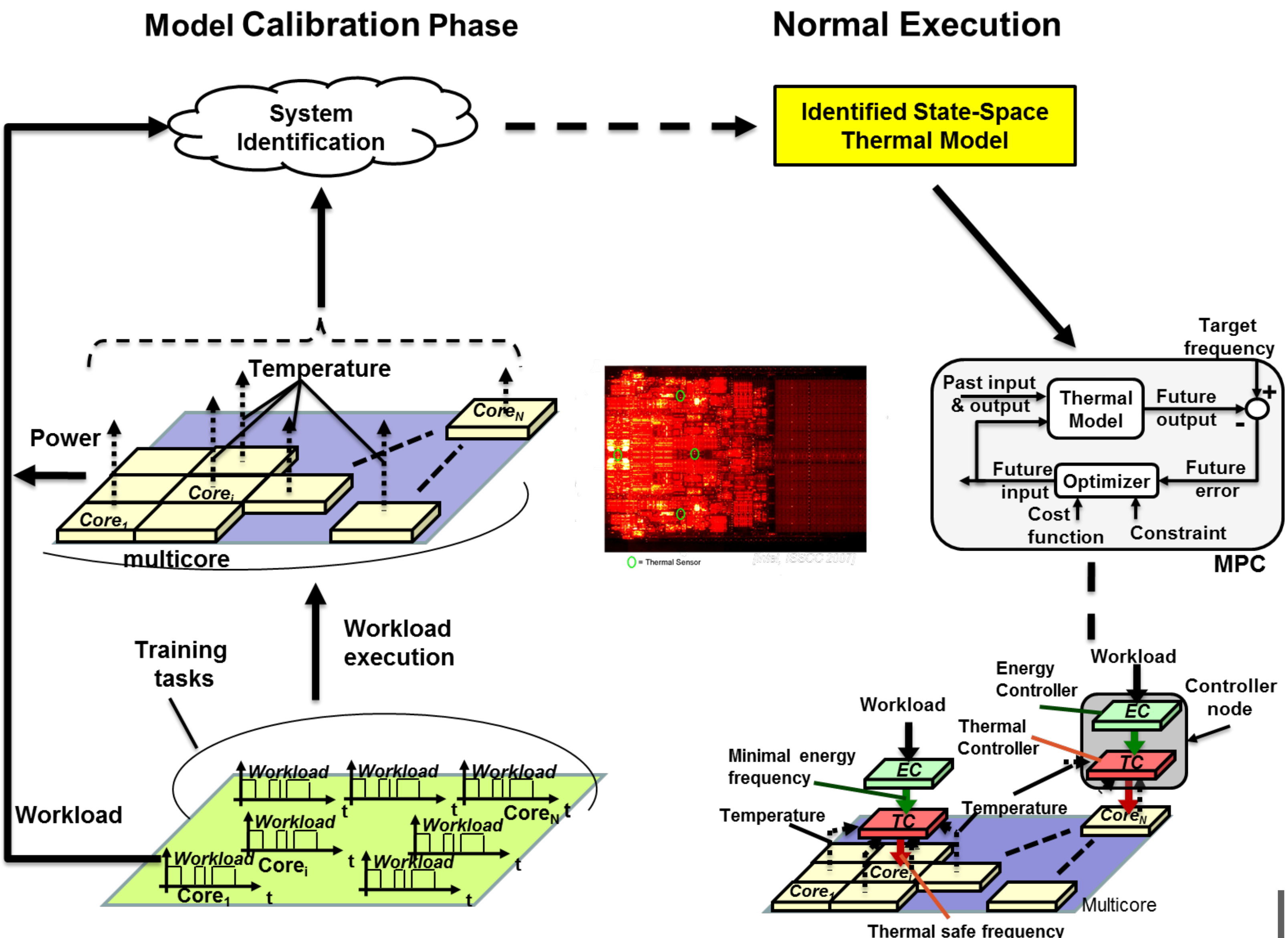
[1] W. Huang Differentiating the roles of IR measurement and simulation for power and temperature-aware design 2009.

Thermal management

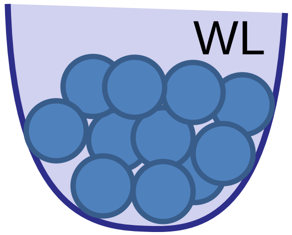
Holistic view:

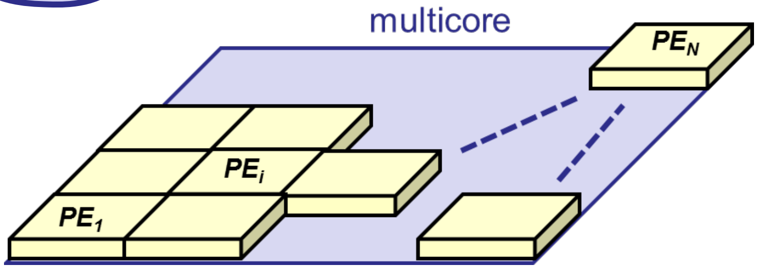


Distributed model predictive control



Thermal-aware task allocation





Problem:

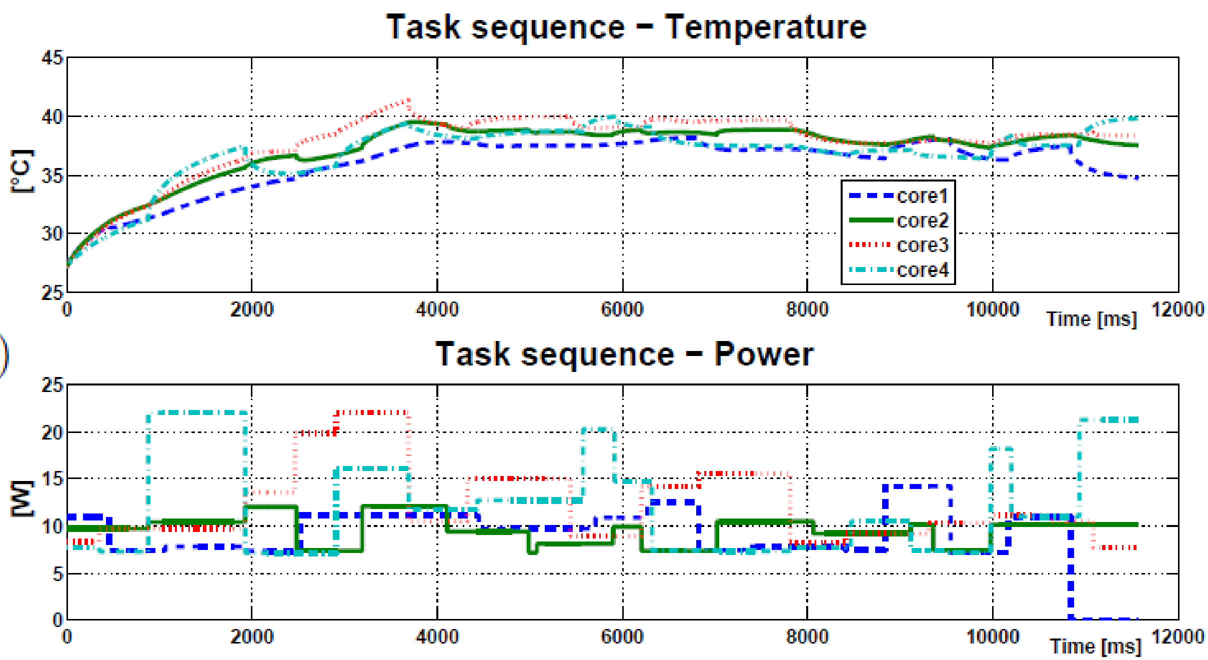
- PE_i processing elements
  WL tasks (wl_i)
  $f_{MIN} < f_{PE} < f_{MAX}$
- Given wl_i choose $\{PE_i , f_{PE}\}$
  - Global Deadline is respected
  - Minimize final T_{PEAK}

Our solution:

- Off-line learn the relation

$$T_j(\tau_0 + \Delta) = f(\overline{T}(\tau_0), \overline{P}, \Delta, T_{env})$$

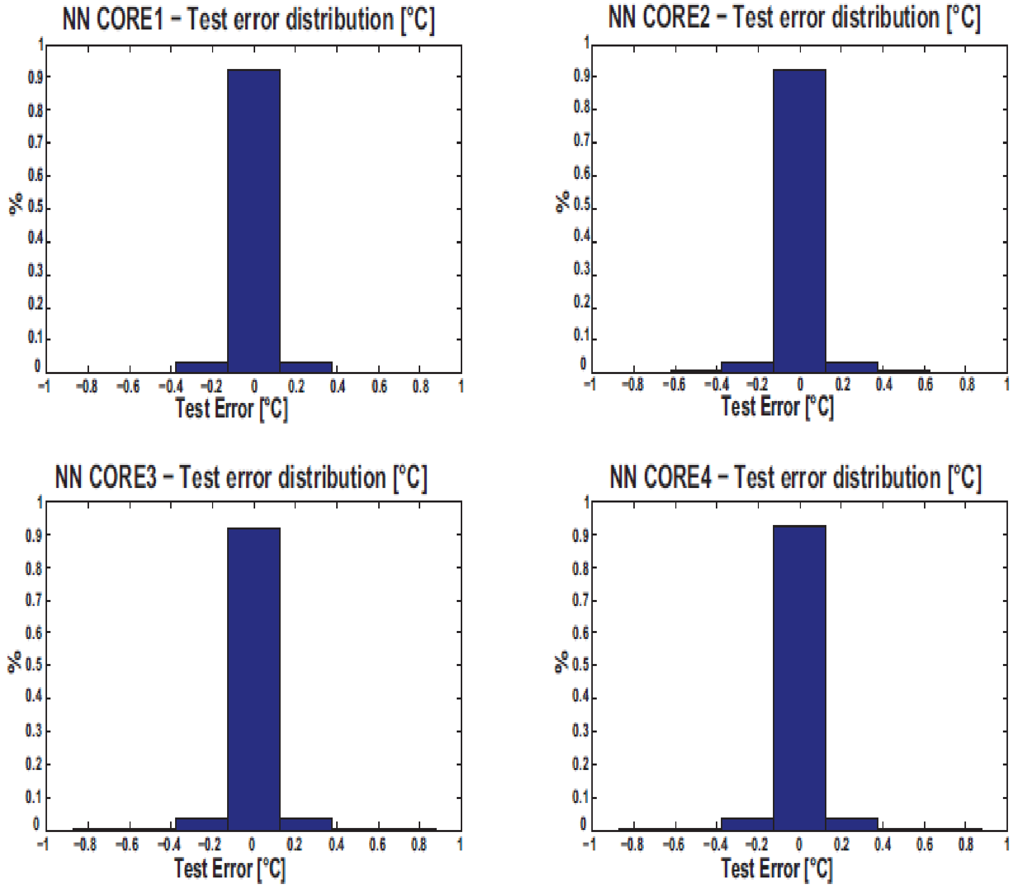
- Solver:
  - Use it as additional constraint in the search tree



Thermal-aware task allocation

Neural network:

- 2 layers – dimensions:
  - 13 input
  - 10 hidden layer size
  - 1 output



Thermal-aware task allocation

Results:

- Our NN approach vs:
  - minimize power (PP)
  - minimize cumulative duration (HH)
  - At different starting temperature

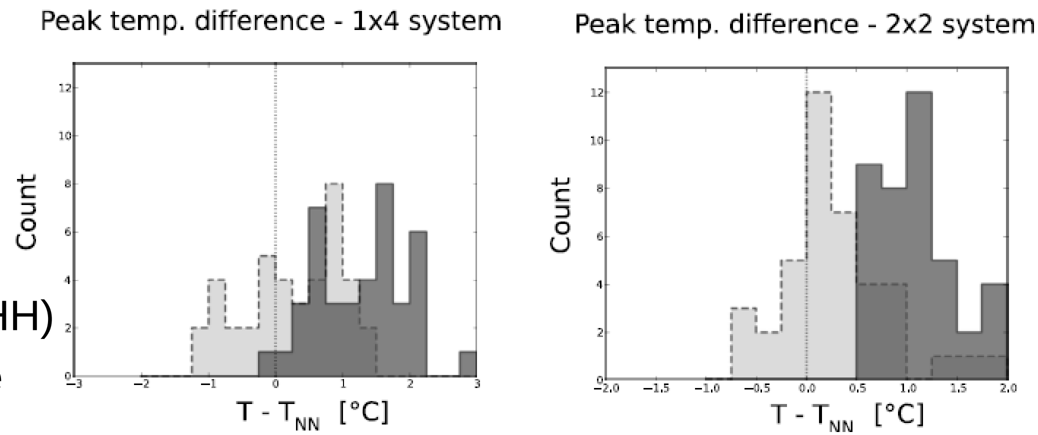


Fig. 3. Difference from NN in final peak temperature for the HH (dark grey) and the PP (light grey) approach

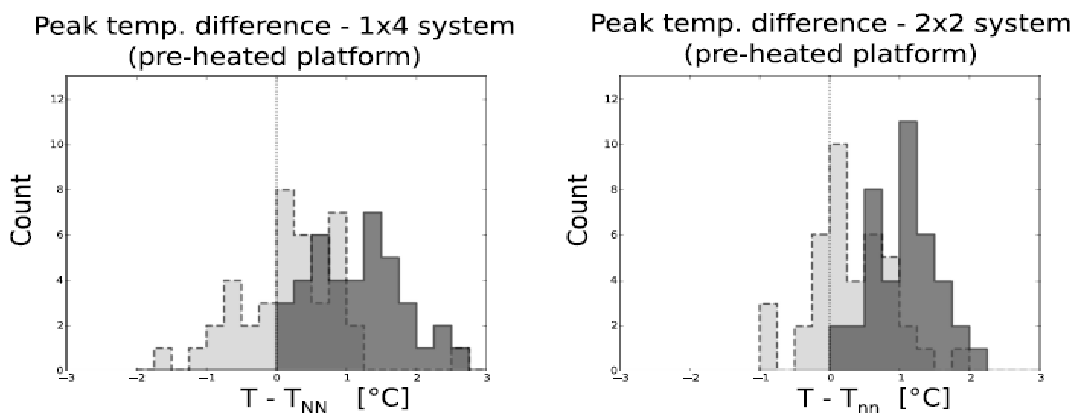


Fig. 4. Difference from NN in final peak temperature for the HH (dark grey) and the PP (light grey) approach – pre-heated platform

Future Work

- Communication aware MPC
  - Implement in SCC
- Distributed MPC
  - Implementing in SCC
- Thermal aware scheduling:
  - Multi-stage allocation
  - Distributed NN
  - On-line thermal aware scheduling