Solving Stochastic Unit Commitment at Industrial Scale using Parallel Computing: A Case Study of Central Western Europe

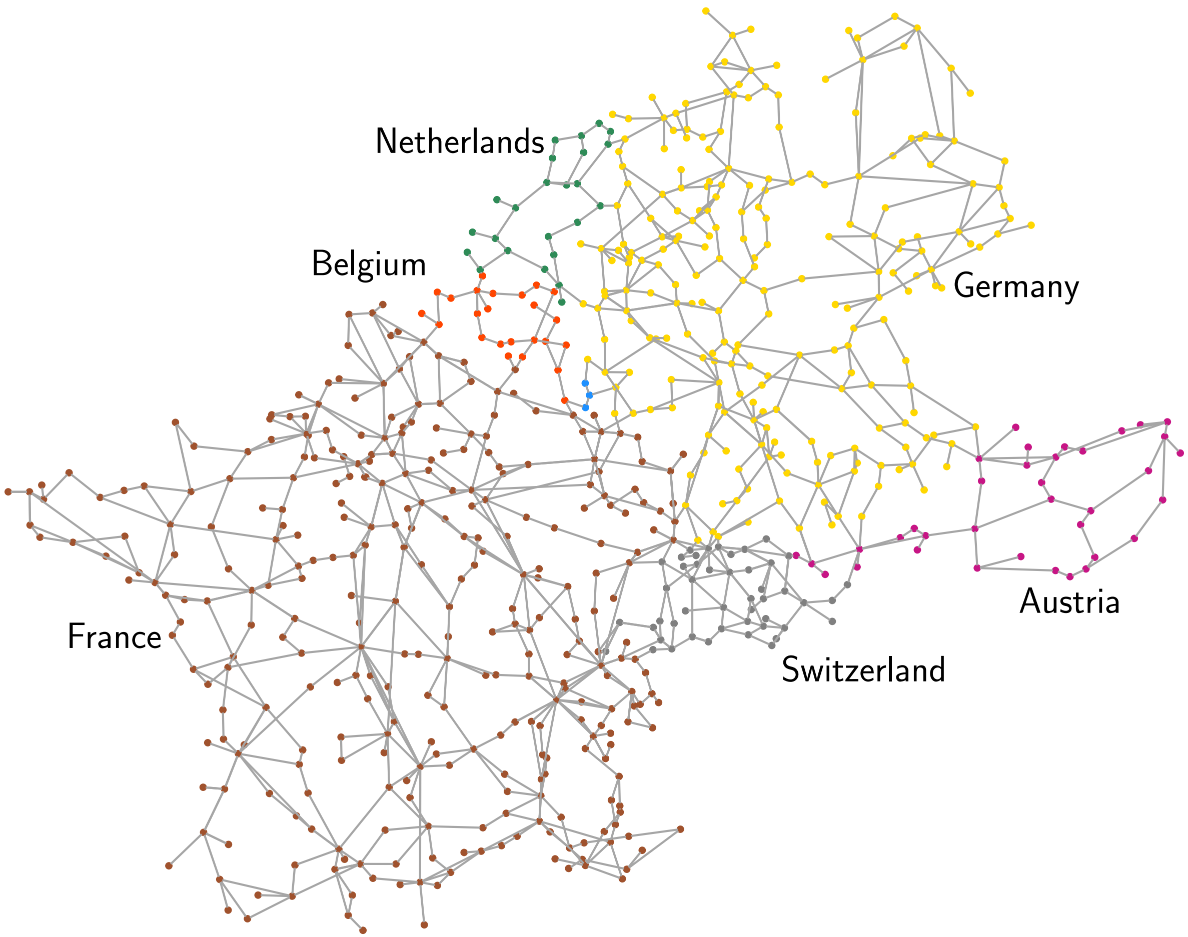
Ignacio Aravena & Anthony Papavasiliou

Prepared for presentation at CORE@50 Conference,
Louvain-la-Neuve, Belgium. May 23–27, 2016.

Motivation

- European electricity markets are structured as zonal markets (in contrast to nodal pricing)
- Zonal market design can affect power systems operations, Ehrenmann and Smeers (2005)
- Continental Europe (Germany) leading renewable energy integration, 82 GW of solar PV power and 108 GW of wind power
- Questions:
  - What are the effects of uncertainty, stemming from renewable resources, on operations under zonal markets?
  - How the performance of the zonal market design and a centralized nodal design compare to each other under current integration levels?

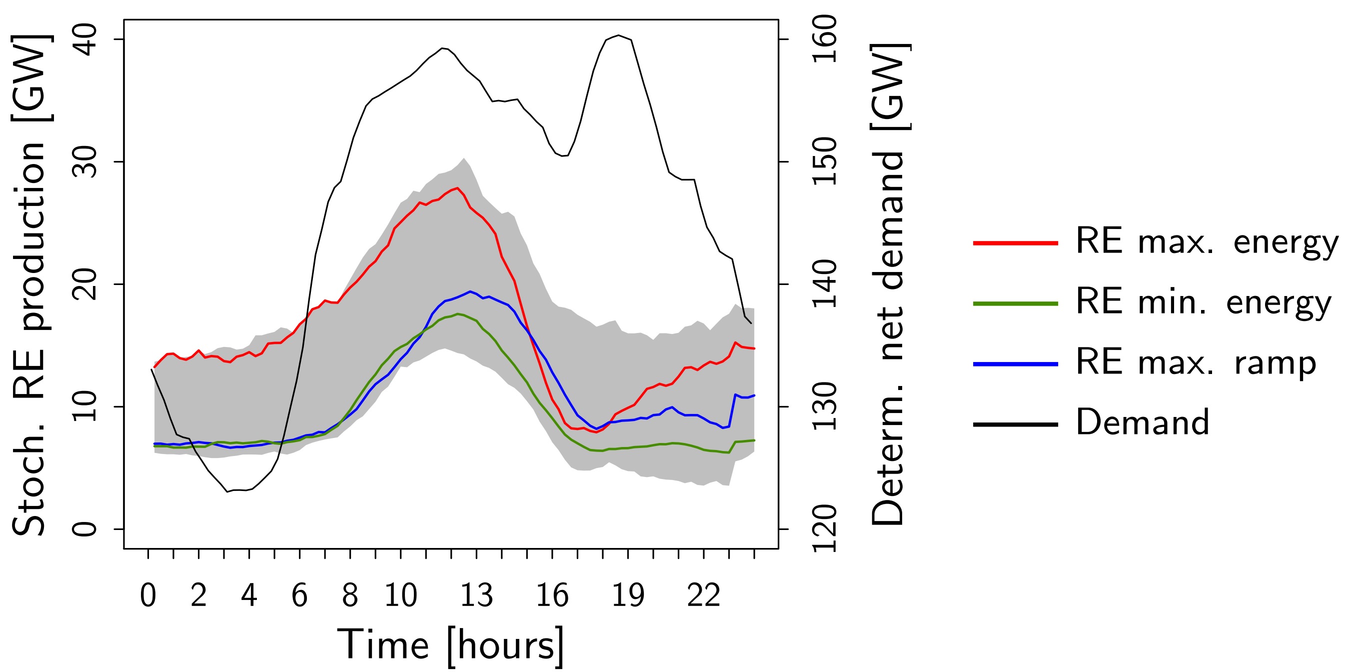
Central Western European (CWE) network



CWE grid model of Hutcheon and Bialek (2013): 7 countries, 679 nodes, 1073 lines

Supply and demand

- 656 thermal generators: 85 GW NUCLEAR, 40 GW CHP, 99 GW SLOW, 14 GW FAST and 10 GW AGGREGATED (small)
- 47.3 GW of solar PV power and 51.2 GW of wind power
- Multi-area renewable production and demand with 15' time resolution

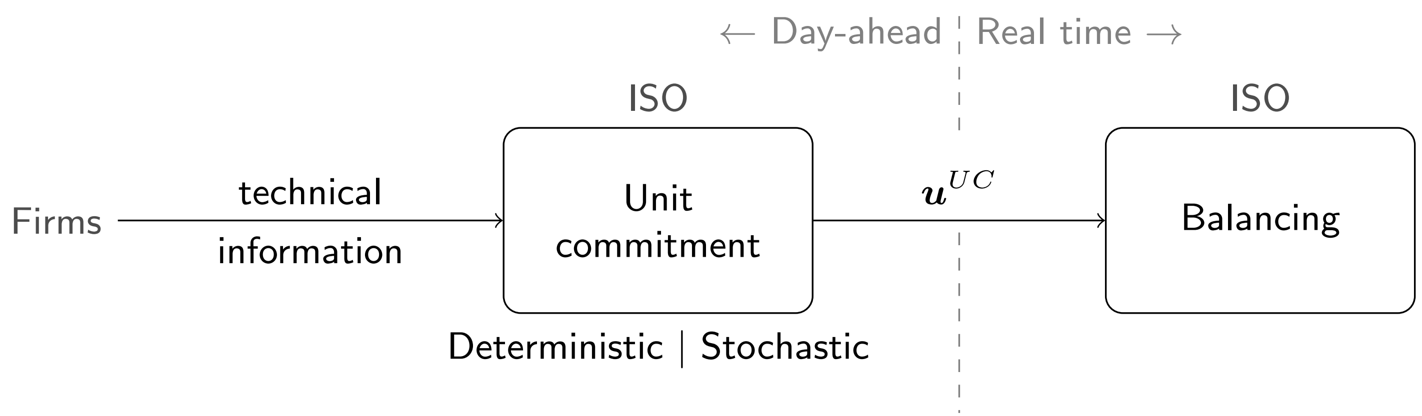


Outline

- Benchmark models: deterministic and stochastic UC
  - Asynchronous algorithm for stochastic UC
- European electricity market model
- Policy comparison results and analysis
- Conclusions

Benchmark models: deterministic and stochastic UC

Centralized unit commitment



- Standard in power systems literature
- Deterministic UC models currently used for day-ahead scheduling in MISO, PJM, CAISO and other systems worldwide
- Stochastic UC model useful for systems with significant renewable integration, Papavasiliou and Oren (2013)

Solving stochastic UC for the CWE system

- One scenario subproblem has 444 thousand variables, 539 thousand constraints and 9552 integers
- Certain scenario subproblems can take up to 75 times more running time than others
  - 10' for easy subproblems – 12 hours for hard subproblems
  - Synchronous decomposition schemes not effective
- **Idea**: use simpler algorithms for which each iteration requires to evaluate only a subset of subproblems
- Relevant literature: Bertsekas & Tsitsiklis (1989), Tseng, (2001), Nedić *et al.*, (2001), Kiwiel, (2004), Fercoq & Richtárik, (2013), Liu *et al.*, (2014)

Dual decomposition

$$
\begin{aligned}
SUC: \quad \min_{\substack{p,u,v\\ w,z}} \quad & \sum_{s\in S} \pi_s \sum_{g\in G} \Big(NLC_g\, \mathbf{1}^T \boldsymbol{u}_{g,s} + SUC_g\, \mathbf{1}^T \boldsymbol{v}_{g,s} + c_g\big(\boldsymbol{p}_{g,s}\big) \Big) \\
\text{s.t.} \quad & \big(\boldsymbol{p}_s, \boldsymbol{u}_s, \boldsymbol{v}_s\big) \in \mathcal{D}_s \\
& \big(\boldsymbol{w}, \boldsymbol{z}\big) \in \mathcal{D}^{wz} \\
& \boldsymbol{w}_g = \boldsymbol{u}_{g,s} \quad (\pi_s \boldsymbol{\mu}_{g,s}), \quad \boldsymbol{z}_g = \boldsymbol{v}_{g,s} \quad (\pi_s \boldsymbol{\nu}_{g,s}) \quad \forall g \in G_{\mathsf{SLOW}}, s \in S
\end{aligned}
$$

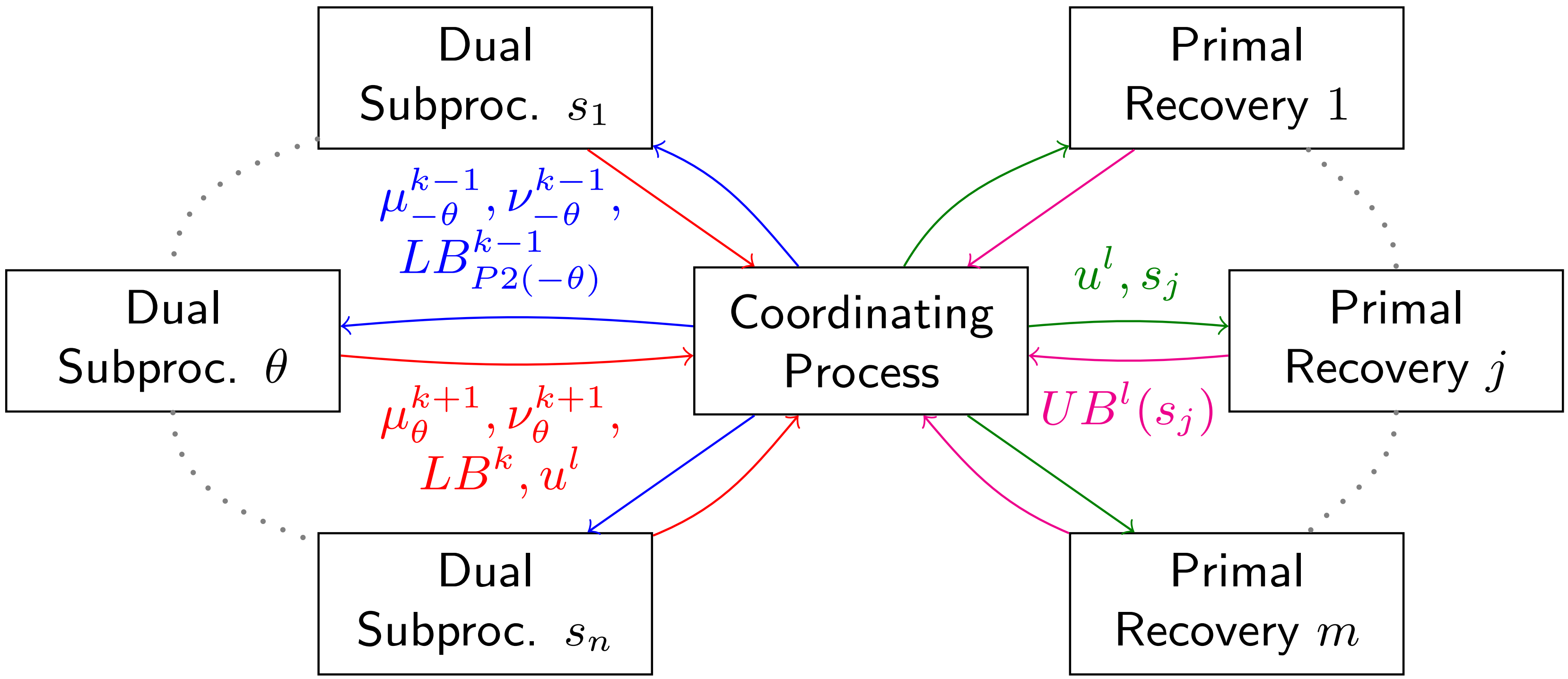
Dual decomposition

$$\begin{aligned} SUC: \quad \min_{\substack{p,u,v \\ w,z}} \quad & \sum_{s \in S} \pi_s \sum_{g \in G} \Big(NLC_g \, \mathbf{1}^T \boldsymbol{u}_{g,s} + SUC_g \, \mathbf{1}^T \boldsymbol{v}_{g,s} + c_g\big(\boldsymbol{p}_{g,s}\big)\Big) \\ \text{s.t.} \quad & \big(\boldsymbol{p}_s, \boldsymbol{u}_s, \boldsymbol{v}_s\big) \in \mathcal{D}_s \\ & \big(\boldsymbol{w}, \boldsymbol{z}\big) \in \mathcal{D}^{wz} \\ & \boldsymbol{w}_g = \boldsymbol{u}_{g,s} \quad (\pi_s \boldsymbol{\mu}_{g,s}), \quad \boldsymbol{z}_g = \boldsymbol{v}_{g,s} \quad (\pi_s \boldsymbol{\nu}_{g,s}) \quad \forall g \in G_{\mathsf{SLOW}}, s \in S \end{aligned}$$

$$\begin{aligned} P2(s): \quad \min_{p,u,v} \quad & \pi_s \sum_{g \in G} \Big((NLC_g \mathbf{1} + \boldsymbol{\mu}_{g,s})^T \boldsymbol{u}_{g,s} + (SUC_g \mathbf{1} + \boldsymbol{\nu}_{g,s})^T \boldsymbol{v}_{g,s} + c_g\big(\boldsymbol{p}_{g,s}\big)\Big) \\ \text{s.t.} \quad & \big(\boldsymbol{p}_s, \boldsymbol{u}_s, \boldsymbol{v}_s\big) \in \mathcal{D}_s \end{aligned}$$

$$\begin{aligned} P1: \quad \min_{w,z} \quad & - \sum_{g \in G_{\mathsf{SLOW}}} \left(\left(\sum_{s \in S} \pi_s \boldsymbol{\mu}_{g,s} \right)^T \boldsymbol{w}_g + \left(\sum_{s \in S} \pi_s \boldsymbol{\nu}_{g,s} \right)^T \boldsymbol{z}_g \right) \\ \text{s.t.} \quad & \big(\boldsymbol{w}, \boldsymbol{z}\big) \in \mathcal{D}^{wz} \end{aligned}$$

Proposed scheme



Note: $\mu_\theta^k, \nu_\theta^k$ are maintained within Dual Sub-process θ

Standard block-coordinate descent iteration

- $k(\theta)$: current iteration in sub-process θ
- Dual Sub-process θ:
  - Evaluates subproblem P2 for scenario θ with current multipliers $\boldsymbol{\mu}_\theta^{k(\theta)}, \boldsymbol{\nu}_\theta^{k(\theta)}$
  - Evaluates P1 with current full multipliers

$$\boldsymbol{\mu} := \left(\boldsymbol{\mu}_{s_1}^{k(s_1)}, \ldots, \boldsymbol{\mu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\mu}_{s_n}^{k(s_n)}\right)$$
$$\boldsymbol{\nu} := \left(\boldsymbol{\nu}_{s_1}^{k(s_1)}, \ldots, \boldsymbol{\nu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\nu}_{s_n}^{k(s_n)}\right)$$

  - Computes block-coordinate subgradient update on $\boldsymbol{\mu}_\theta, \boldsymbol{\nu}_\theta$
- **Problem:** dual function is never fully evaluated $\rightarrow$ impossibility to compute lower bounds

Modified dual iterations

- Dual Sub-process θ:
  - Evaluates subproblem P2 for scenario θ with the current multipliers $\boldsymbol{\mu}_\theta^{k(\theta)}, \boldsymbol{\nu}_\theta^{k(\theta)} \rightarrow LB_{P2(\theta)}^{k(\theta)}$
  - Evaluates P1 with **delayed** multipliers $\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\nu}} \rightarrow LB_{P1}^{k(\theta)}$

$$\bar{\boldsymbol{\mu}} := \left(\boldsymbol{\mu}_{s_1}^{k(s_1)-1}, \ldots, \boldsymbol{\mu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\mu}_{s_n}^{k(s_n)-1}\right)$$
$$\bar{\boldsymbol{\nu}} := \left(\boldsymbol{\nu}_{s_1}^{k(s_1)-1}, \ldots, \boldsymbol{\nu}_\theta^{k(\theta)}, \ldots, \boldsymbol{\nu}_{s_n}^{k(s_n)-1}\right)$$

  - Computes block-coordinate subgradient update on $\boldsymbol{\mu}_\theta, \boldsymbol{\nu}_\theta$
  - Computes **lower bound on objective using last evaluations of P2** subproblems for other scenarios,

$$\text{Objective} \geq LB_{P1}^{k(\theta)} + LB_{P2(\theta)}^{k(\theta)} + \sum_{s \neq \theta} LB_{P2(s)}^{k(s)-1}$$

Lower bound initialization

- Certain scenario subproblems can take up to 75 times more than others to be solved $\rightarrow$ **one scenario** can delay the computation of the first "full" lower bound

- Use a relaxation of P2 to obtain an initial lower bound (not useful for updating dual multipliers)

- Which relaxation?
  - Linear relaxation of P2
  - Sequence of OPF problems

Primal recovery

- Recovering primal candidates (1st stage) from P2 subproblems $\rightarrow$ good quality solutions from first iterations, Ahmed (2013)
- Accumulating large number of primal candidates: prune bad candidates if possible
  - Pruning candidates based on cuts from Angulo *et al.* (2014)
  - Second stage cost non-increasing function of $\boldsymbol{u}$: $\boldsymbol{u}^i \geq \boldsymbol{u}^j \Rightarrow C_2(\boldsymbol{u}^i) \leq C_2(\boldsymbol{u}^j)$, hence

$$LB(\boldsymbol{u}^{\text{new}}) = C_1(\boldsymbol{u}^{\text{new}}) + \max_{\substack{j \in J \\ u^j \geq u^{\text{new}}}} C_2(\boldsymbol{u}^j)$$

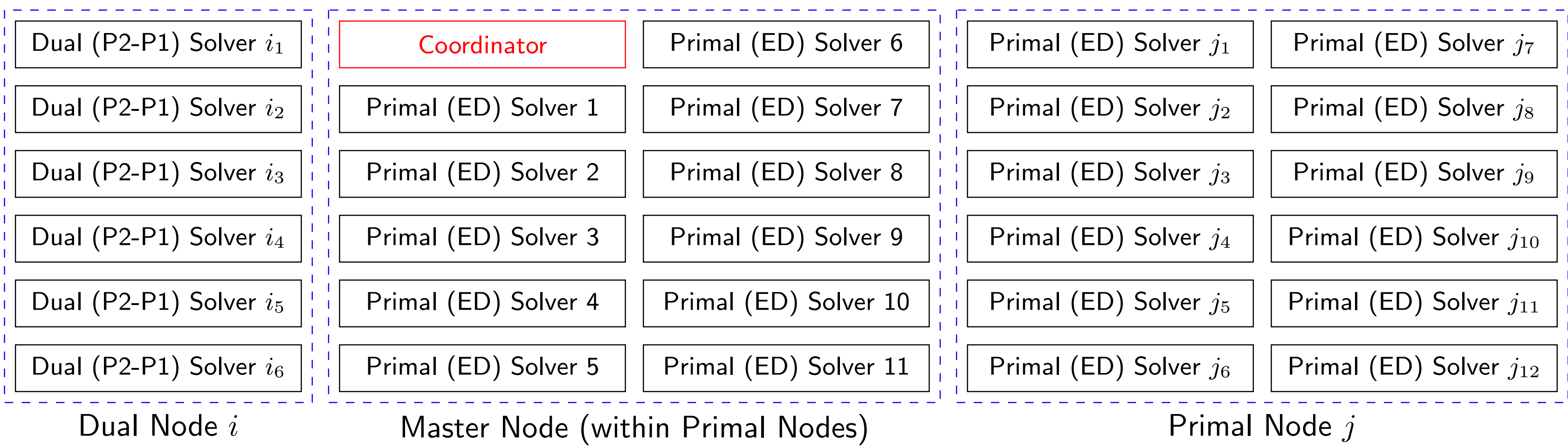
Running time comparison: CWE system instances

Model	Scenarios	Variables	Constraints	Integers
Determ2R	1	570,432	655,784	9,552
Determ3R	1	636,288	719,213	9,552
Stoch30	30	13,334,400	16,182,180	293,088
Stoch60	60	26,668,800	32,364,360	579,648
Stoch120	120	53,337,600	64,728,720	1,152,768

- Asynchronous SUC implemented in Mosel using the `mmjobs` module and the XPress solver
- Lawrence Livermore National Laboratory Sierra cluster: 23,328 cores on 1,944 nodes, 2.8 Ghz, 24 GB/node

Running times comparison: implementation details

- Using 10 nodes per SUC instance:
  - 5 nodes dedicated to dual iterations / 6 sub-processes per node (subproblem P2 memory requirements)
  - 5 nodes dedicated to primal recovery / 12 primal recovery scenario sub-problems per node



Running times comparison

Solution statistics over 8 instances (day types).

Model	Nodes	Running time [hours]	Worst final gap [%]
Determ2R	1	1.9 (0.6 – 4.2)	0.95
Determ3R	1	$\geq$ 9.4 (6.3 – 10.0)	4.91
Stoch30[1]	10	1.1 (0.7 – 2.2)	0.93
Stoch30i[2]	10	0.8 (0.3 – 1.8)	1.00
Stoch60[1]	10	3.2 (1.1 – 8.4)	1.00
Stoch60i[2]	10	1.5 (0.6 – 4.7)	0.97
Stoch120[1]	10	$\geq$ 6.1 (1.6 – 10.0)	1.68
Stoch120i[2]	10	$\geq$ 3.0 (1.4 – 10.0)	1.07

Termination criteria: 1% optimality or 10 hours wall-time.

[1] Dual initialization using linear relaxation of P2.

[2] Dual initialization using sequential OPF.

Running times comparison: optimality vs. wall-time

Solution statistics over 8 instances (day types).

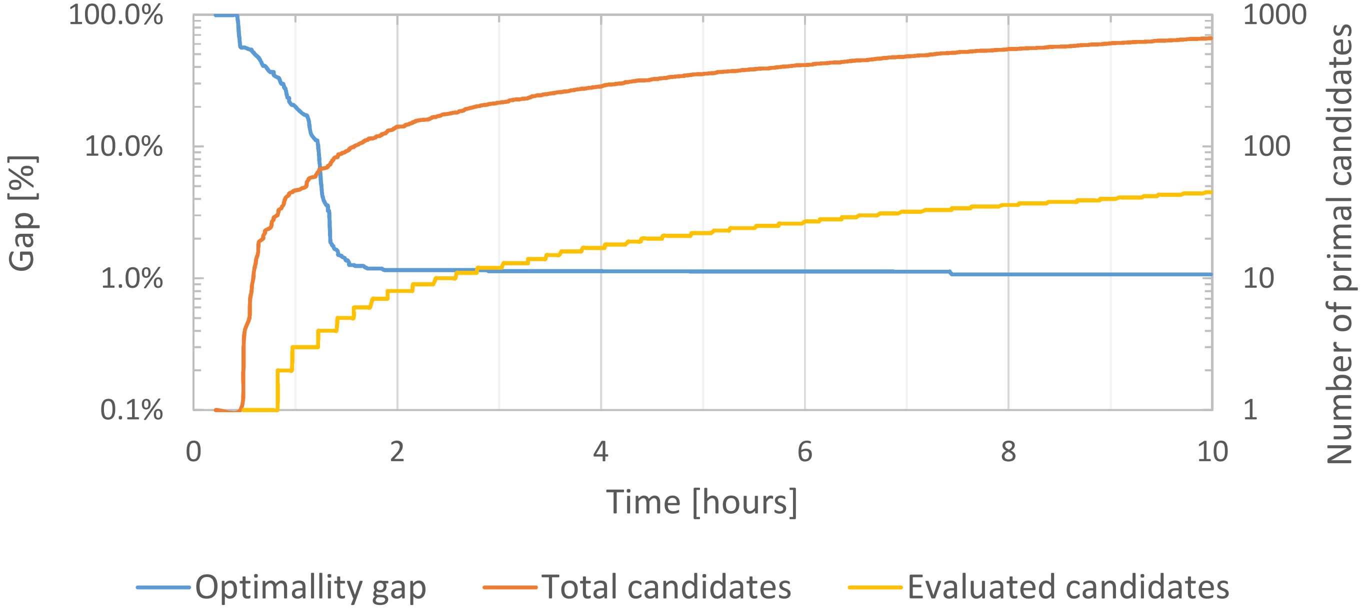
Model	Worst gap [%]			
	1 hour	2 hours	4 hours	8 hours
Stoch30	7.59	1.02	0.93	–
Stoch30i	1.90	1.00	–	–
Stoch60	23.00	5.32	5.22	4.50
Stoch60i	4.60	1.57	1.03	0.97
Stoch120	70.39	31.66	4.61	1.87
Stoch120i	46.69	27.00	1.42	1.07

- Lower bound initialization using sequential OPF observed to be very effective, sometimes avoiding to solve P2 for hard scenarios
- Asynchronous SUC algorithm capable of achieving acceptable optimality gaps within running time of deterministic UC

Room for improvement: evaluation of primal candidates

Bounds and primal candidates.

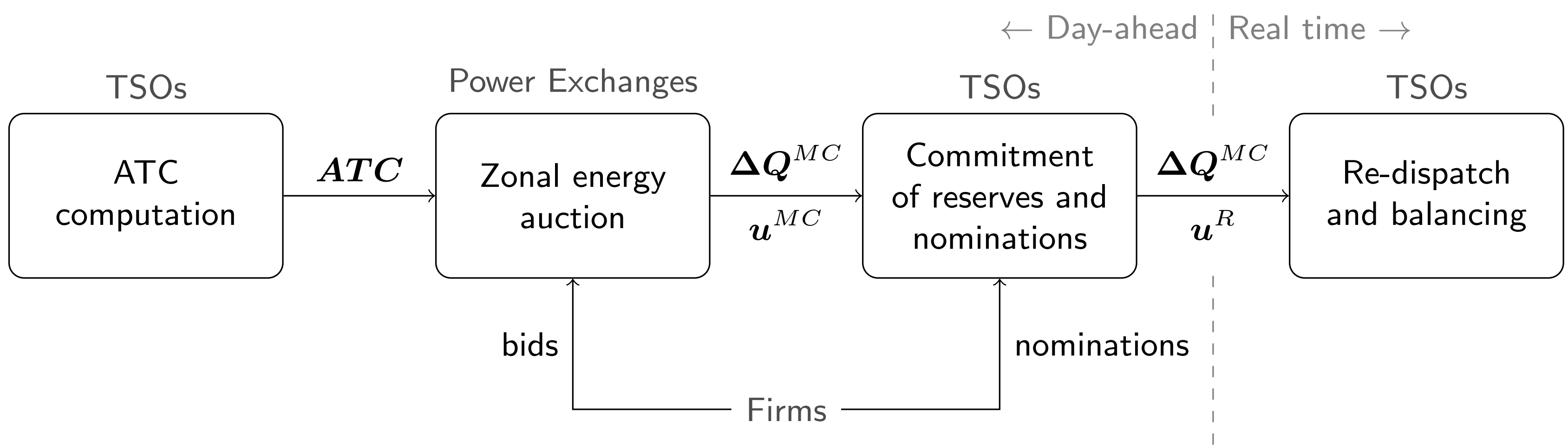
Stoch120i, 5 dual – 5 primal nodes, summer weekday (worst case).



- Pruning of primal candidates is not effective: discards less than 1% of candidates
- Valuable computational resources spent in **detailed** evaluation of sub-optimal candidates

European electricity market model

Market Coupling (MC)

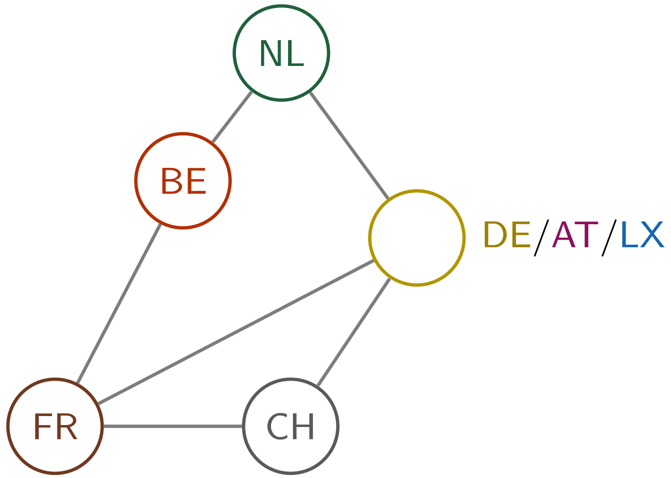


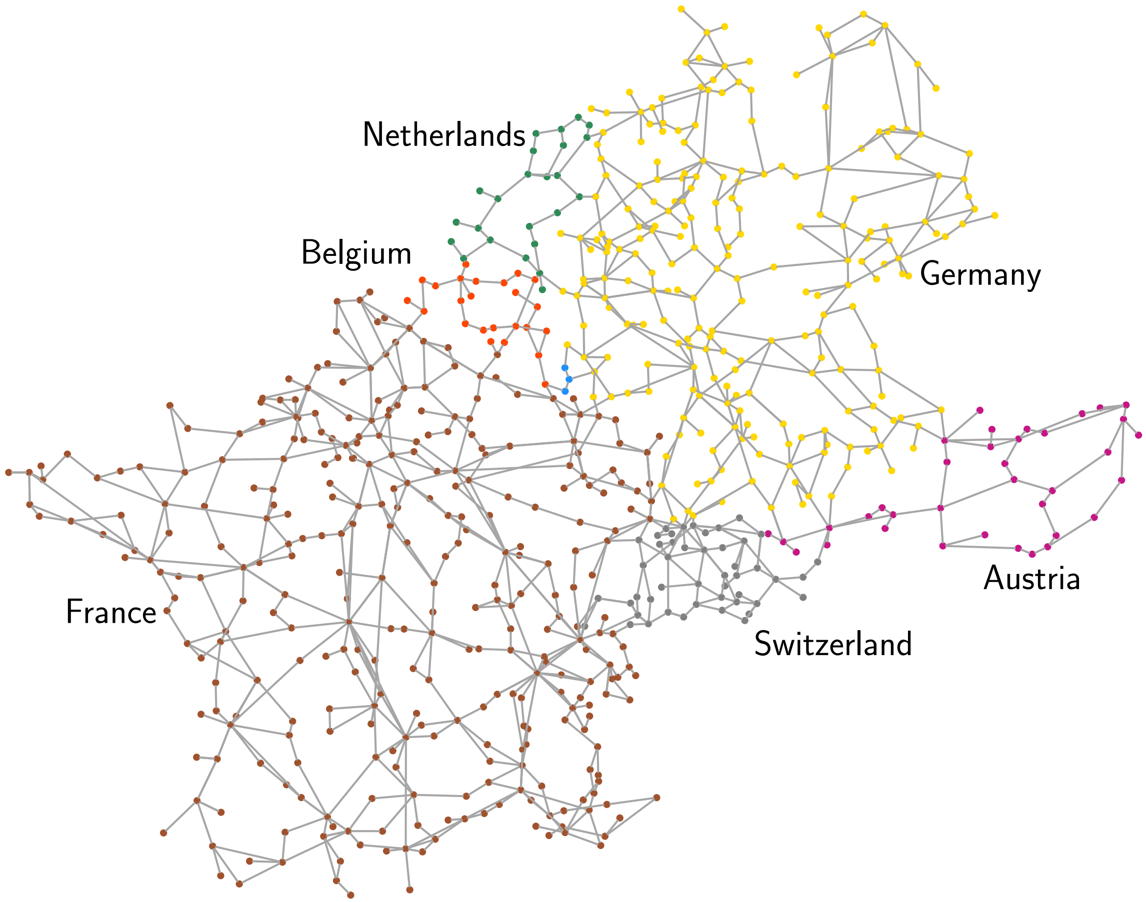
- Previous work: Ehrenmann and Smeers (2005), Leuthold *et al.* (2009), van der Weijde and Hobbs (2011), Oggioni and Smeers (2011), (2012), Kunz (2013)

Main differences between MC and UC

Market Coupling (MC)	Unit Commitment (UC)
PX(s), TSO(s) (partial system knowledge)	ISO (complete system knowledge)
Exchange	Power pool
Sequential market clearing	Simultaneous market clearing
Zonal energy clearing (one price per zone)	Nodal energy clearing (one price per node)
Respecting day-ahead zonal net positions on real time	Fully coordinated balancing

Zonal vs nodal pricing in the CWE network





Policy comparison results and analysis

Simulation setting

- Commitment of NUCLEAR and CHP decided prior to day-ahead
- Commitment of SLOW units decided in day-ahead, commitment of FAST units decided in real time. Production of all units decided on real time.

- 8 day types: 4 seasons $\times$ weekdays/weekends
- Real time operation cost estimated using 120 Monte Carlo samples

- Comparing performance of 4 policies
  - Market coupling respecting net positions, **MCNetPos**
  - Market coupling free international re-dispatch, **MCFree**
  - Deterministic unit commitment, **DetermUC**
  - Stochastic unit commitment, **StochUC**

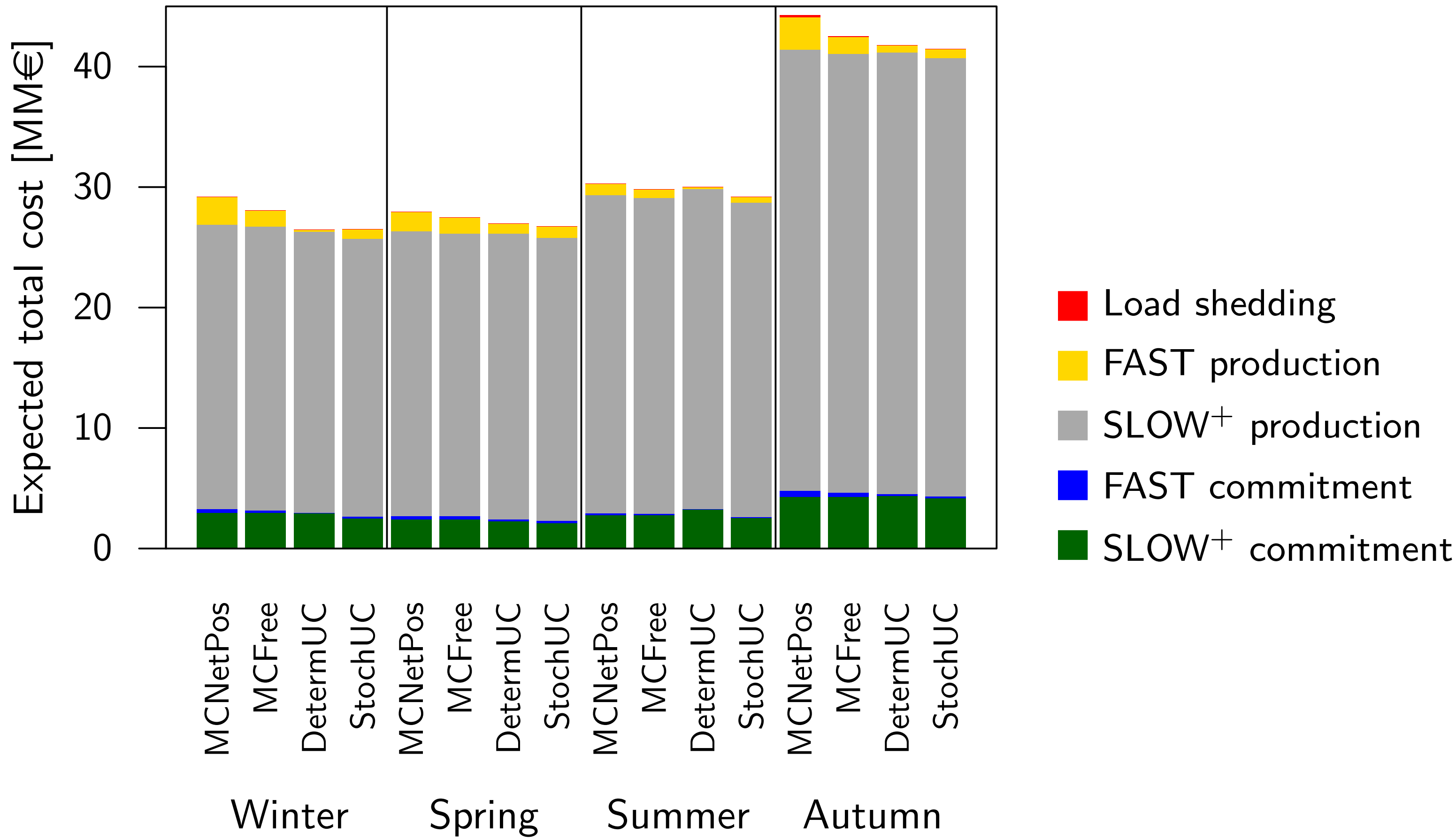
Policy comparison: expected operation costs

Expected policy costs and efficiency losses with respect to deterministic UC

Policy	Expected cost [MM€/d]	Efficiency losses [%]	Efficiency losses [MM€/year]
MCNetPos	30.42	6.2	650
MCFree	29.45	2.8	294
Deterministic UC	28.64	–	–
Stochastic UC	28.49	–0.5	–55
Perf. Foresight	28.32	–1.1	117

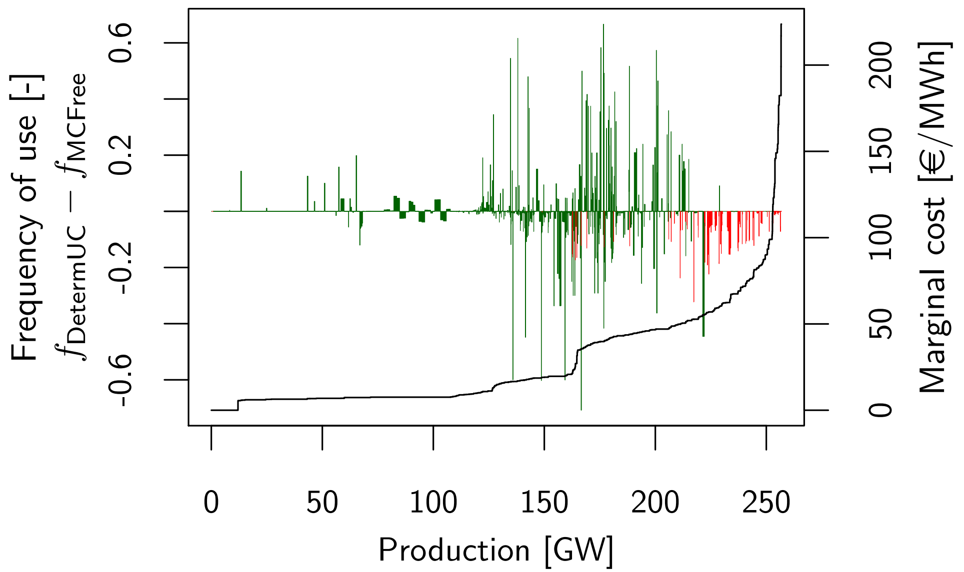
- Small efficiency gains of stochastic UC compared to efficiency losses due to market design
- Congestion management costs for Germany during 2015: 688MM€, ENTSO-E

Policy comparison: cost composition weekdays



Deterministic UC vs. MCFree

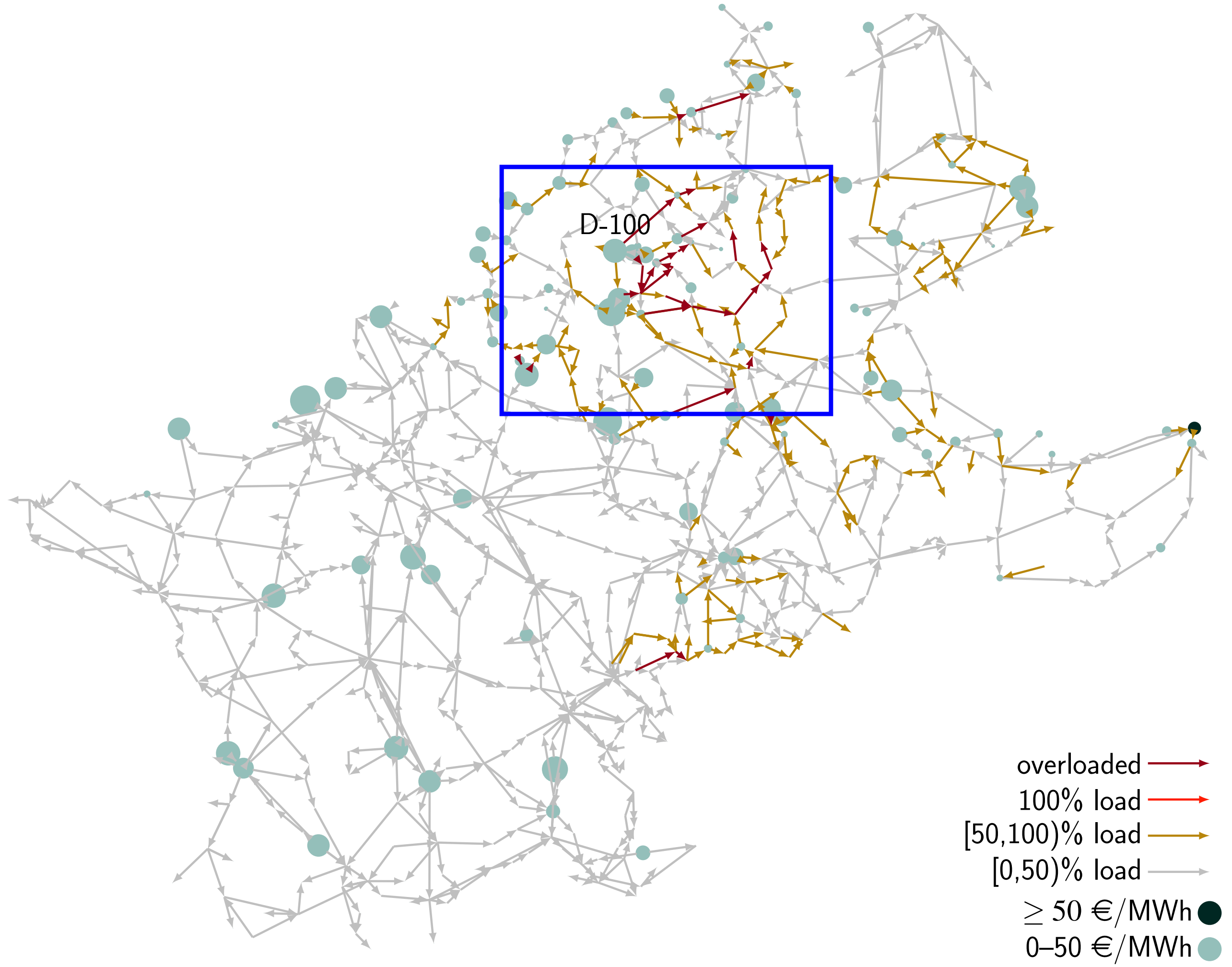
- In day-ahead, deterministic UC and MCFree commit similar amounts of SLOW capacity, but in different nodes
- In real time, MCFree resorts to more FAST generators, including very expensive units



Difference in frequency of use of supply function bins for autumn weekday

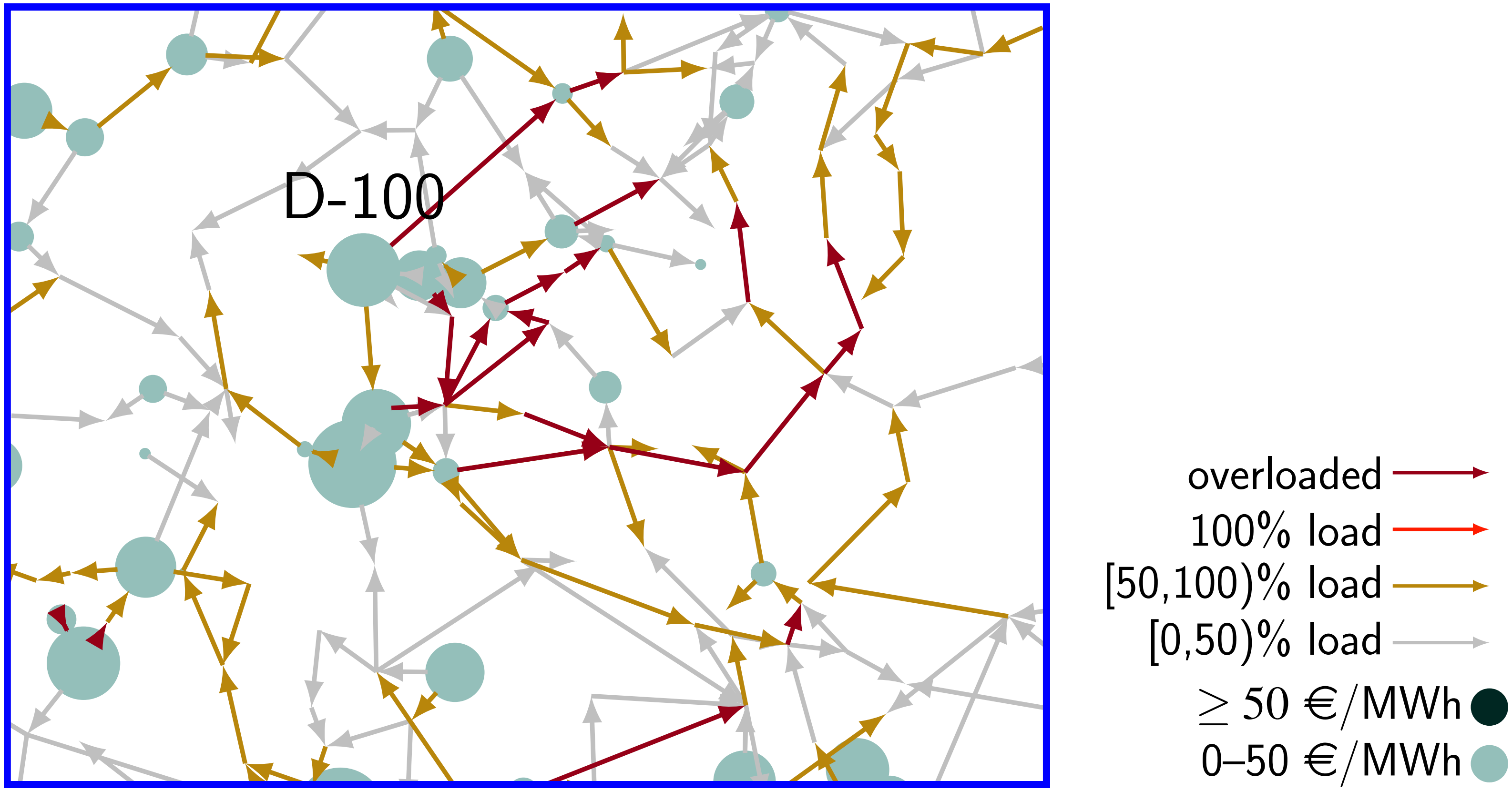
Deterministic UC vs. MCFree

MCFree day-ahead schedule for spring weekday at 17:00–18:00



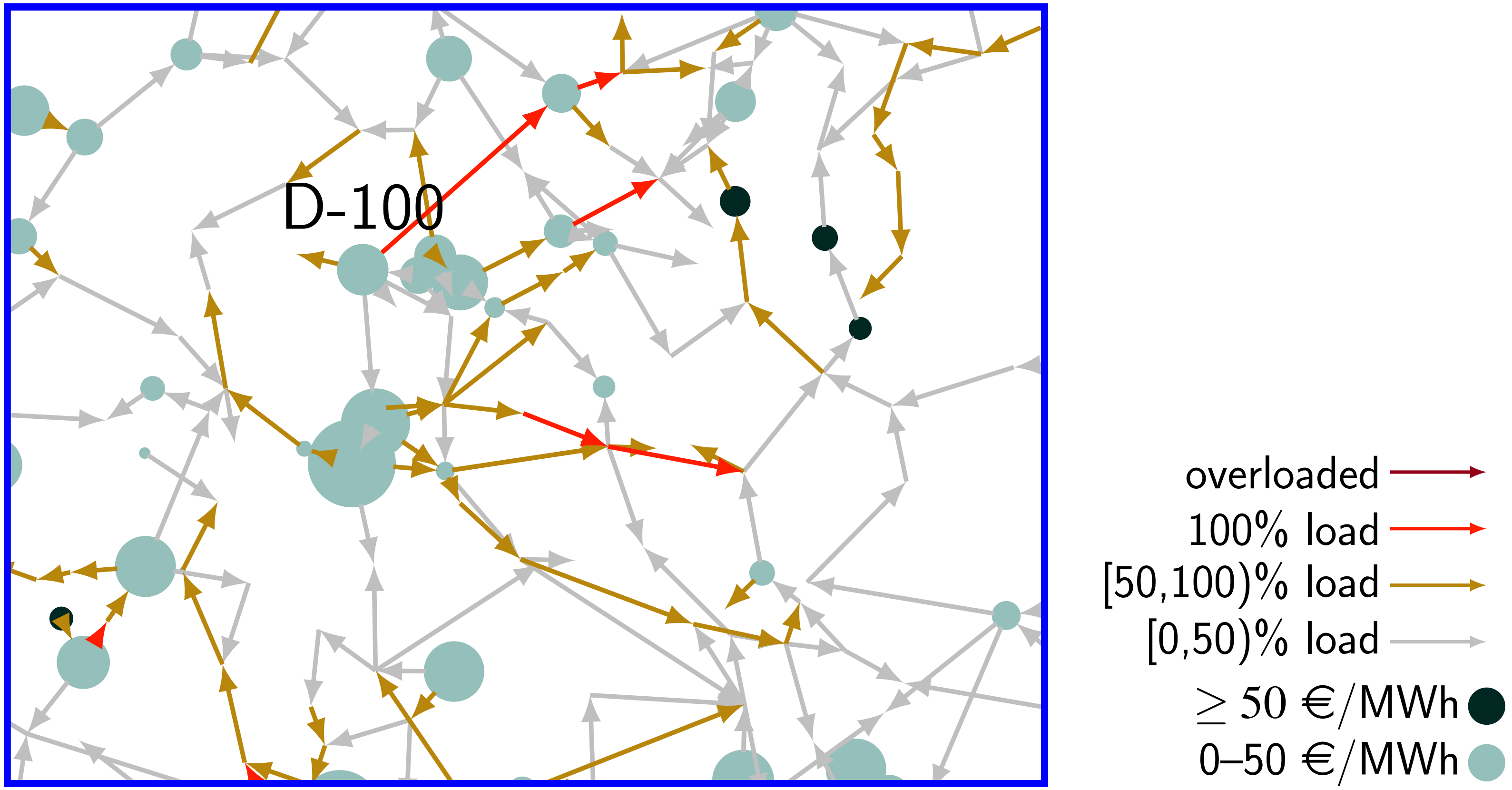
Deterministic UC vs. MCFree

MCFree day-ahead schedule for spring weekday at 17:00–18:00



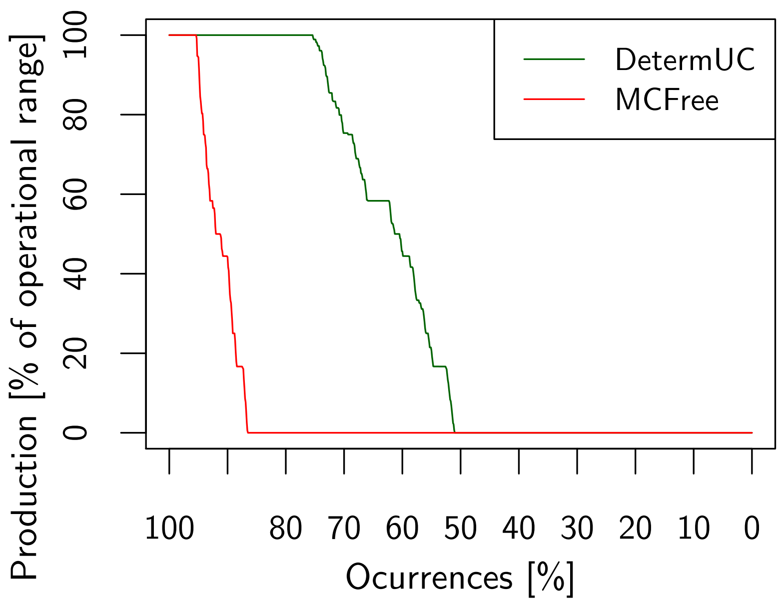
Deterministic UC vs. MCFree

MCFree real-time operation for a sample of spring weekday at 17:30–17:45



Deterministic UC vs. MCFree

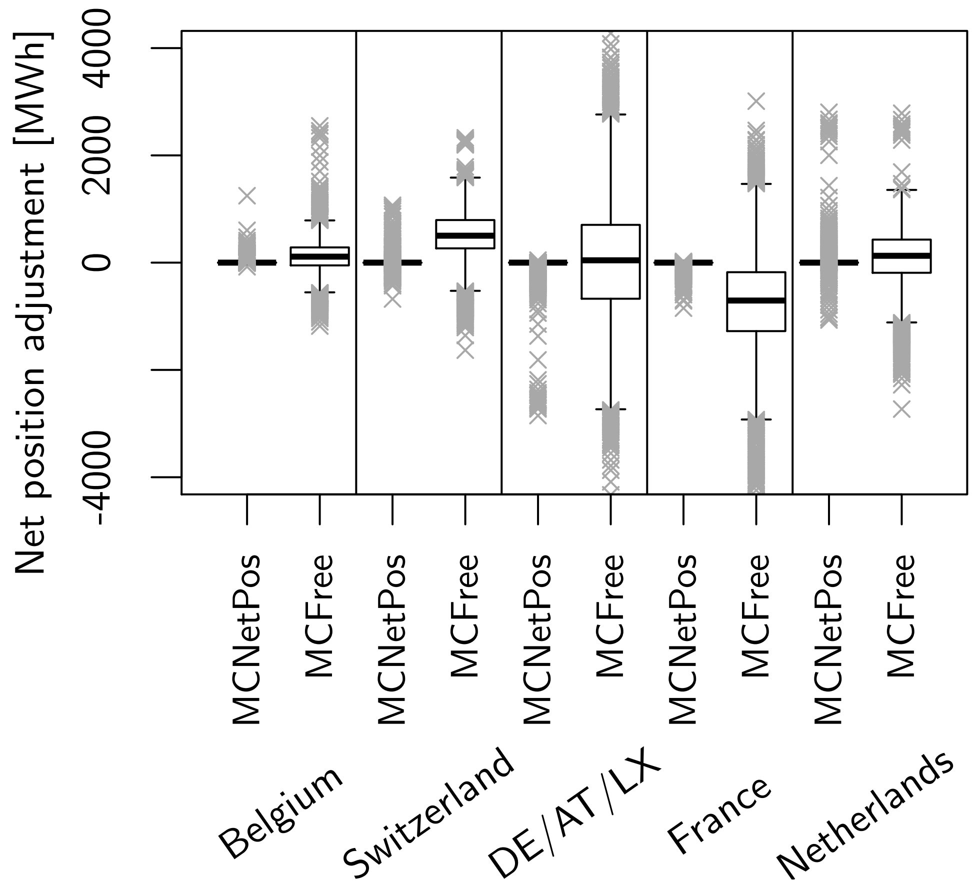
- Infeasibility of day-ahead schedules due to congestion is persistent across periods and day types
- Cheap SLOW generators are re-dispatched down to their technical minimum, while expensive FAST generators are re-dispatched up $\rightarrow$ increase in production cost of FAST units



Production duration curve of SLOW units committed exclusively either by DetermUC or MCFree

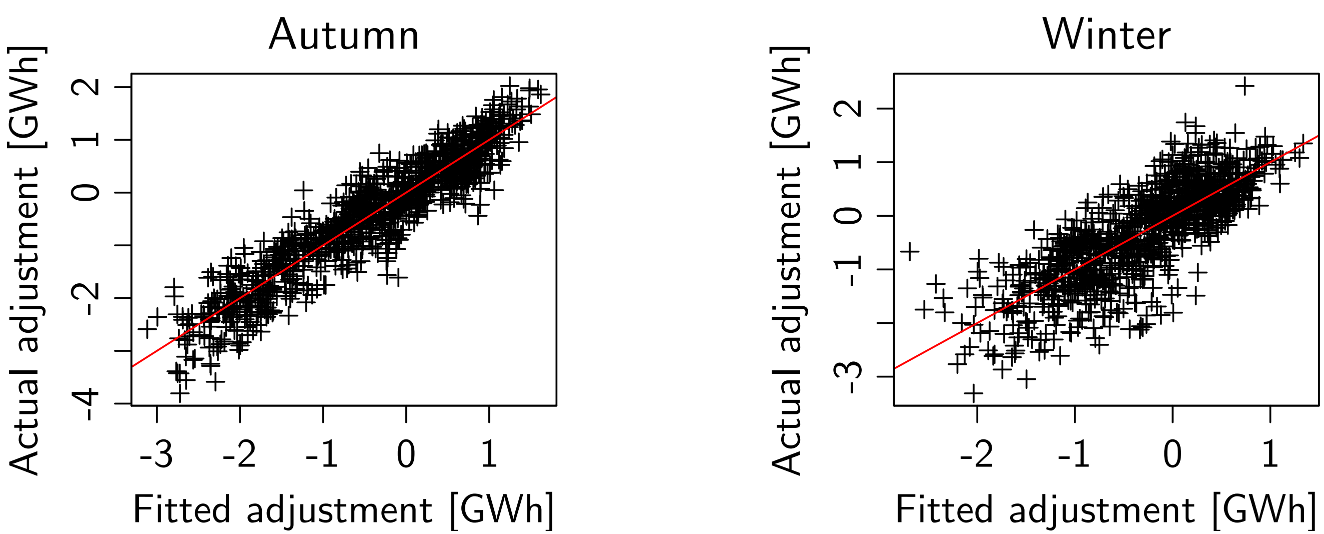
MCFree vs. MCNetPos

Zonal net position adjustments in real time with respect to day-ahead zonal net positions



MCFree vs. MCNetPos

- Adjustment of net positions is driven by renewable forecast error for DE/AT/LX, limited by zonal net demand and day-ahead net position
- Fully coordinated balancing (MCFree) performs better than zonal balancing (MCNetPos): sharing of shortage and excess of renewable supply across zones



Prediction of real time net position adjustment of DE/AT/LX using renewable forecast error for DE/AT/LX, zonal net demand and day-ahead net positions

Conclusions

Conclusions

- Efficiency losses in the European electricity market are due to
  1. suboptimal day-ahead commitment (*MCFree* – *DetermUC*, 2.8%)
  2. uncoordinated balancing (*MCNetPos* – *MCFree*, 3.4%)
- Efficiency losses of type 1 can be strengthened by changing patterns in power flows due to renewable integration
- Efficiency losses of type 2
  - Directly related to renewable forecast errors $\rightarrow$ higher integration levels would imply larger efficiency losses
  - Present in both European and in wide US interconnections
  - They can be corrected through coordinated balancing, 50 Hertz *et al.* (2014), Y. Makarov *et al.* (2010)

Perspectives

- Extensions of asynchronous algorithm
  - Pruning and scoring primal candidates
  - Dynamical queue management for dual and primal processes
  - Multi-stage stochastic UC
- Extensions of present European electricity market model
  - Flow-based MC model
  - Intraday market

Thank you

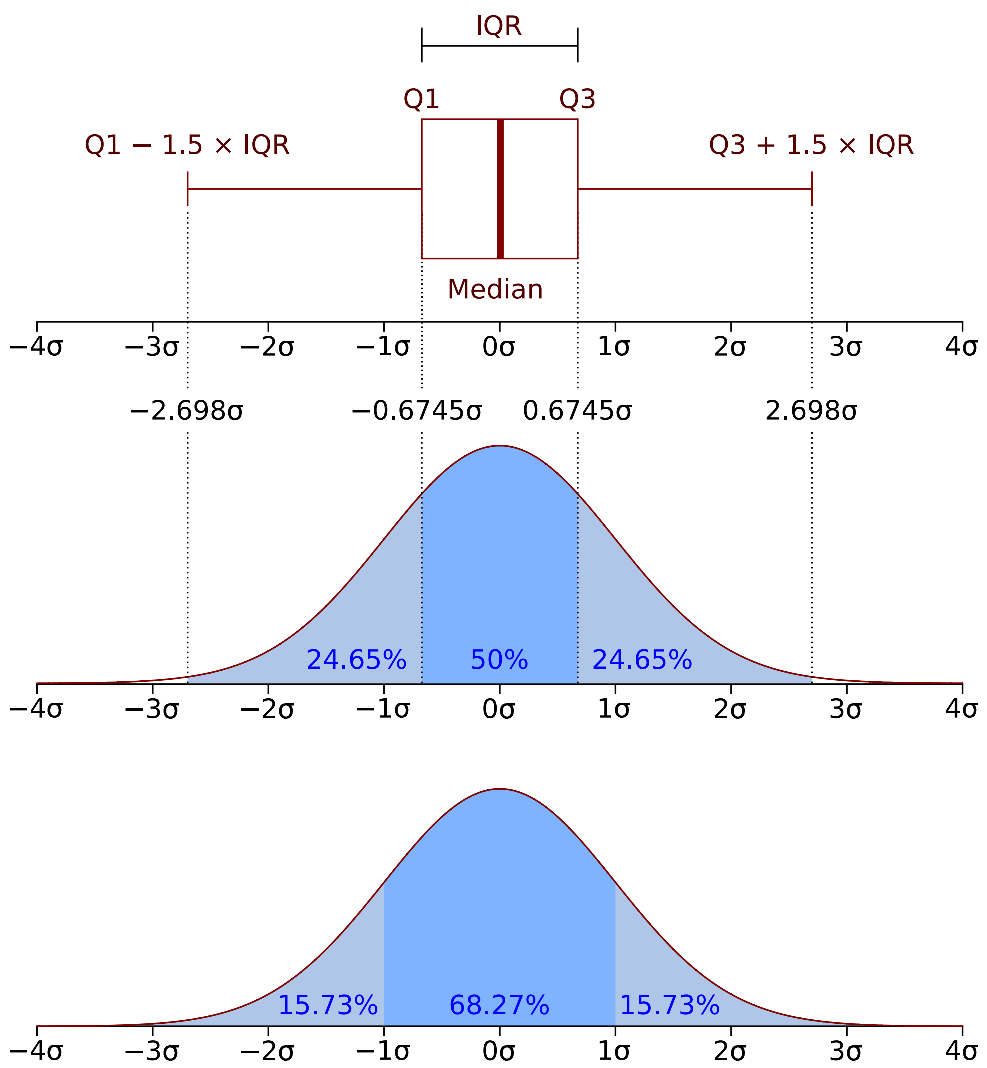
Contact:

Ignacio Aravena, ignacio.aravena@uclouvain.be
http://sites.google.com/site/iaravenasolis/

Anthony Papavasiliou, anthony.papavasiliou@uclouvain.be
http://perso.uclouvain.be/anthony.papavasiliou/

Appendix

Box plot, Gaussian standard distribution



Wikipedia. Box plot. http://en.wikipedia.org/wiki/Box_plot

Available Transfer Capacity computation

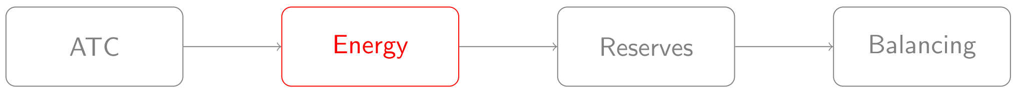
ATC → Energy → Reserves → Balancing

$$
\begin{aligned}
TTC^{+}_{(a,b),\tau} := \max \quad & \textit{Cross border flow } a \to b \\
\text{s.t.} \quad & \textit{Optimal power flow constraints, } \tau \\
& \textit{Base case exchange on other borders, } \tau \\
& \textit{Capacity margin for reserves for } a,\ b
\end{aligned}
$$

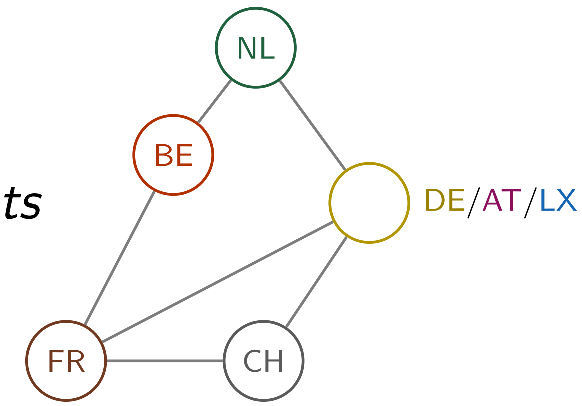
$$ATC^{+}_{(a,b),\tau} = TTC^{+}_{(a,b),\tau} - TRM$$

- $ATC^{+}_{(a,b),\tau}$: Preliminary ATC from zone a to zone b for hour τ
- Computed using the full network, internal and cross border lines thermal ratings, and security criteria
- Checked for simultaneous feasibility, TenneT (2014)
- Problem based on UCTE (2004)

Day-ahead energy market clearing



max *Welfare*
s.t. *Strong duality constraint*
Zonal (transportation) network constraints
Bid acceptance/rejection constraints
Surplus non-negativity constraints



- Hourly resolution
- Demand and renewable producers submitting continuous bids
- Thermal generators modeled as submitting block bids
- Model adapted from MILP model of Madani and Van Vyve (2014)

Commitment of reserves and nominations

ATC → Energy → **Reserves** → Balancing

$$\begin{aligned} \min \quad & \textit{Operation cost for zone } a \\ \text{s.t.} \quad & \textit{Unit commitment constraints} \\ & \textit{Reserves constraints} \\ & \textit{Zone } a \textit{ net position and minimum commitment for} \\ & \quad \textit{SLOW generators from day-ahead energy market} \end{aligned}$$

- Zonal reserves, no network, 15' resolution, hourly commitment
- Three types of reserves: FCR (available in 30", spinning), automatic FRR (available in 7.5', spinning) and manual FRR (available in 15')
- Based on 50 Hertz *et al.* (2014) and ENTSO-E (2014)

Re-dispatch and balancing in real time

ATC → Energy → Reserves → Balancing

min	*Total operation cost*
s.t.	*Unit commitment constraints for FAST units*
	Real (nodal) network constraints
	Renewable energy supply realization
	Fixed net positions on zones and SLOW commitment

- 15’ resolution dispatch, hourly commitment for FAST units
- Simulating over several samples of renewable supply
- Deviations from day-ahead net positions penalized at the maximum marginal cost of any generator in the system
- Based on Oggioni *et al.* (2012)