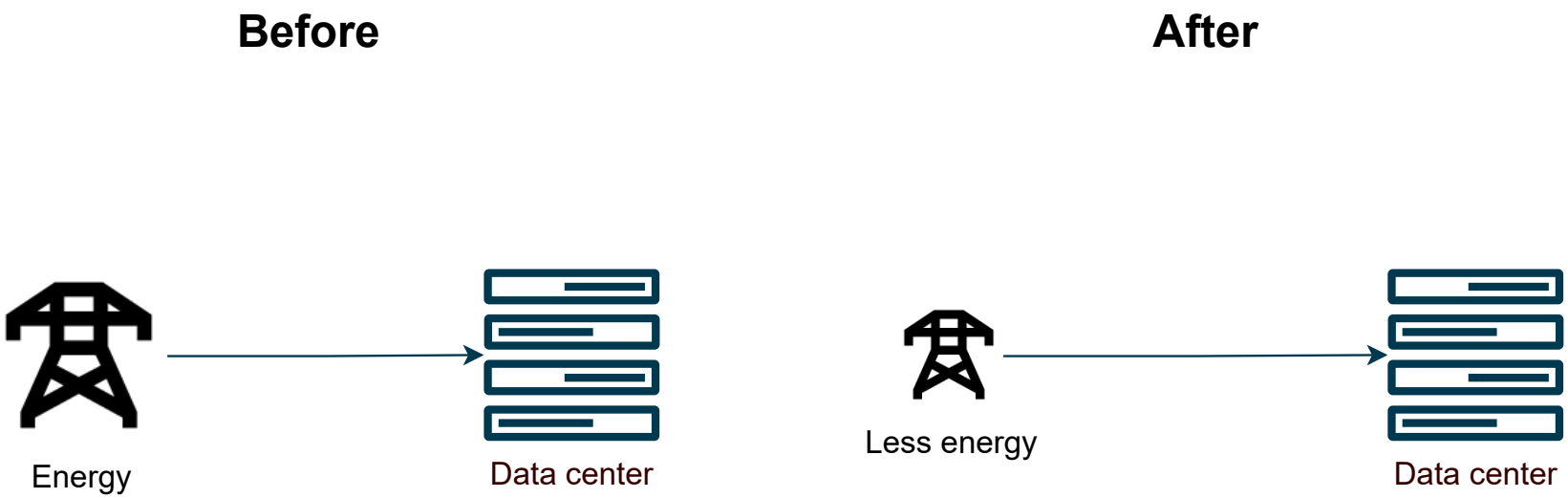
Characterization of user behaviors for demand response in data centers

Maël Madon

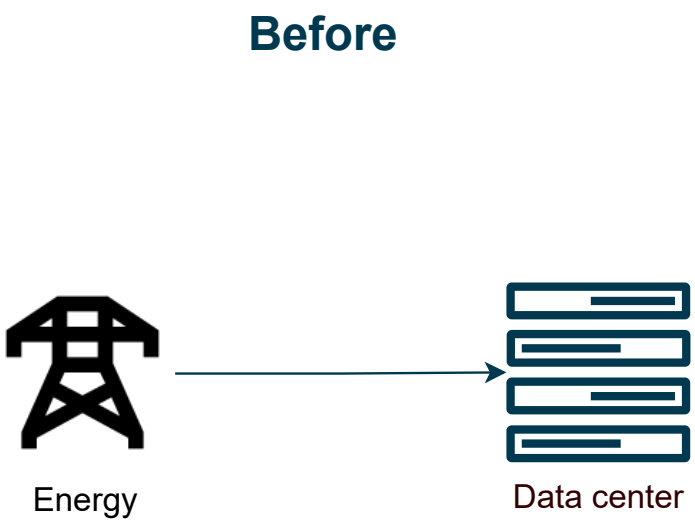
Euro-Par 2022 @ University of Glasgow

Introduction

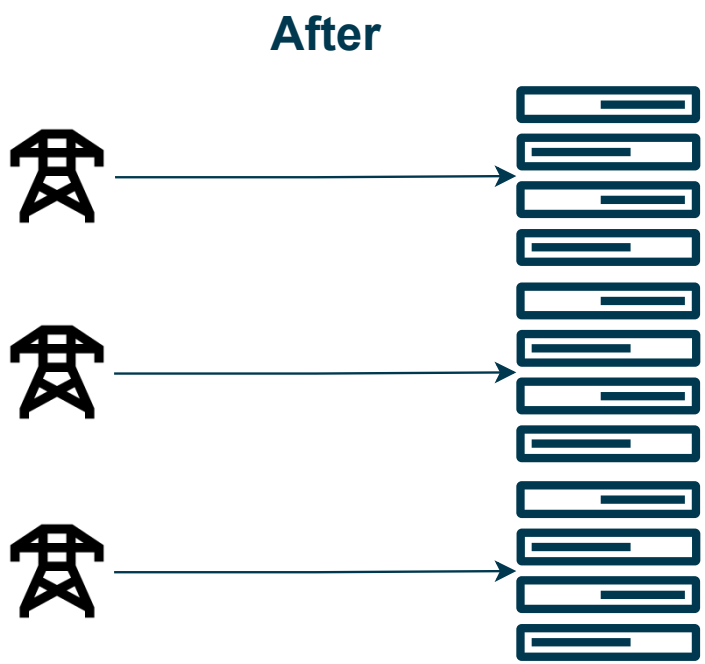
Rebound effect



Rebound effect



Data center = 1
Energy = 2



Data center = 3
Energy = 3

Definitions

Rebound effect

Unrealized savings due to a rebound in the demand caused by the efficiency improvements

- Systemic effect, from the user side => engineers typically unequipped to face it

Sufficiency policies (IPCC, 2022)

A set of measures and daily practices that avoid demand for energy, materials, land and water while delivering human well-being for all within planetary boundaries.

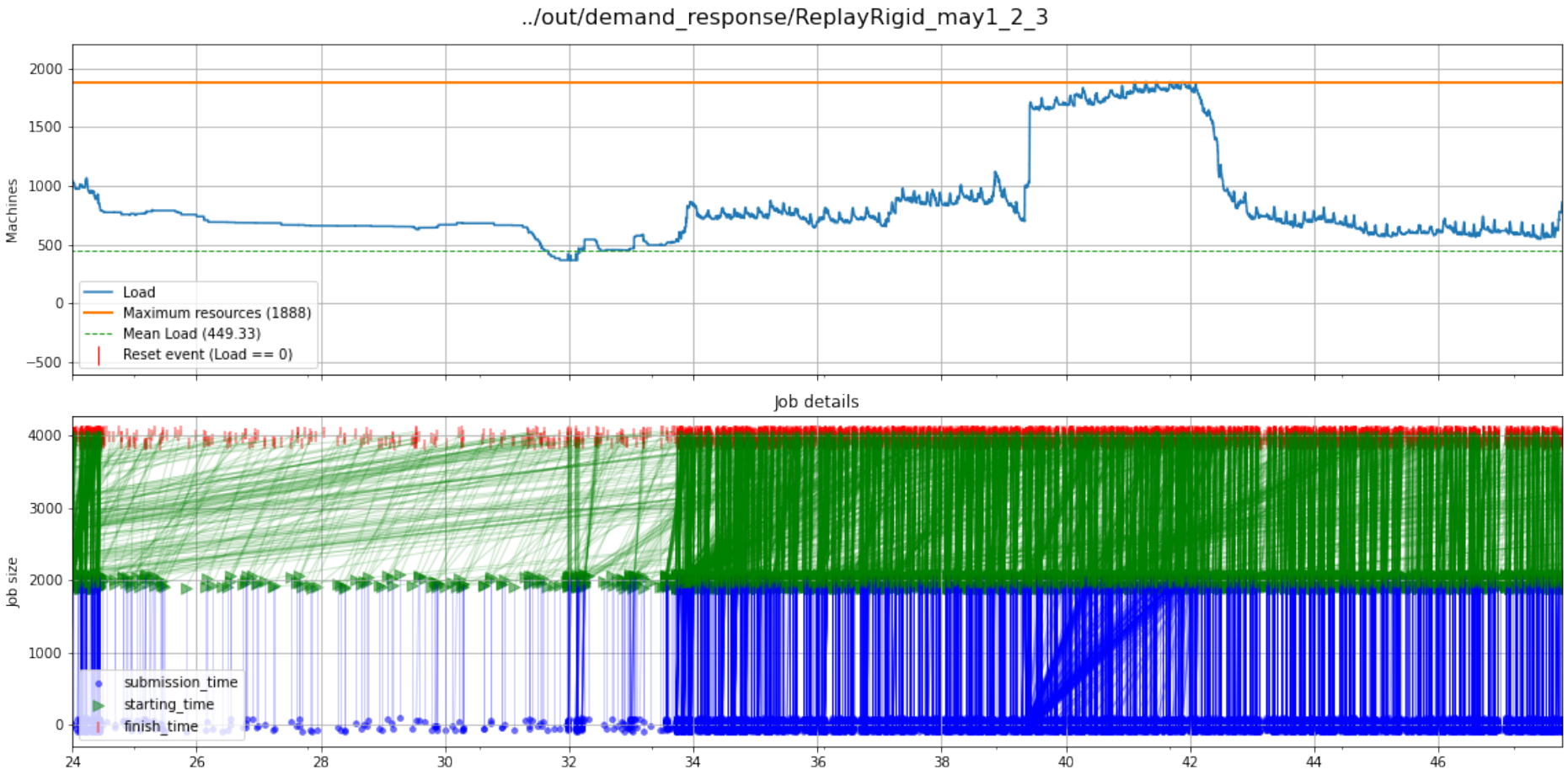
Research question

What would "sufficiency" mean for data centers?

- Auto-regulate ourselves, accept to make efforts
- Prerequisite: **understand how the user submission behavior affects the load in the data center**

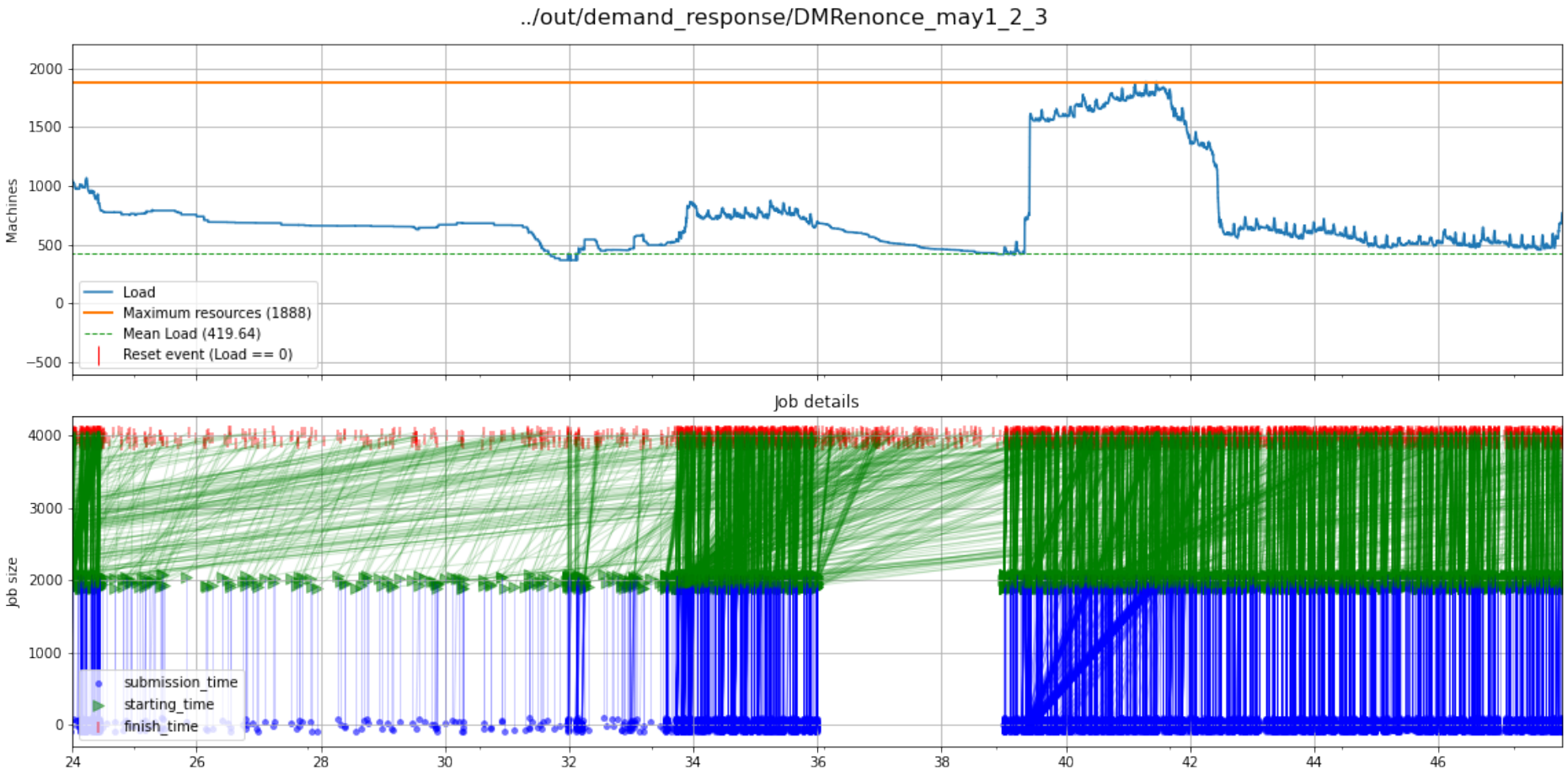
Illustration

- Behavior during demand response window: **rigid**



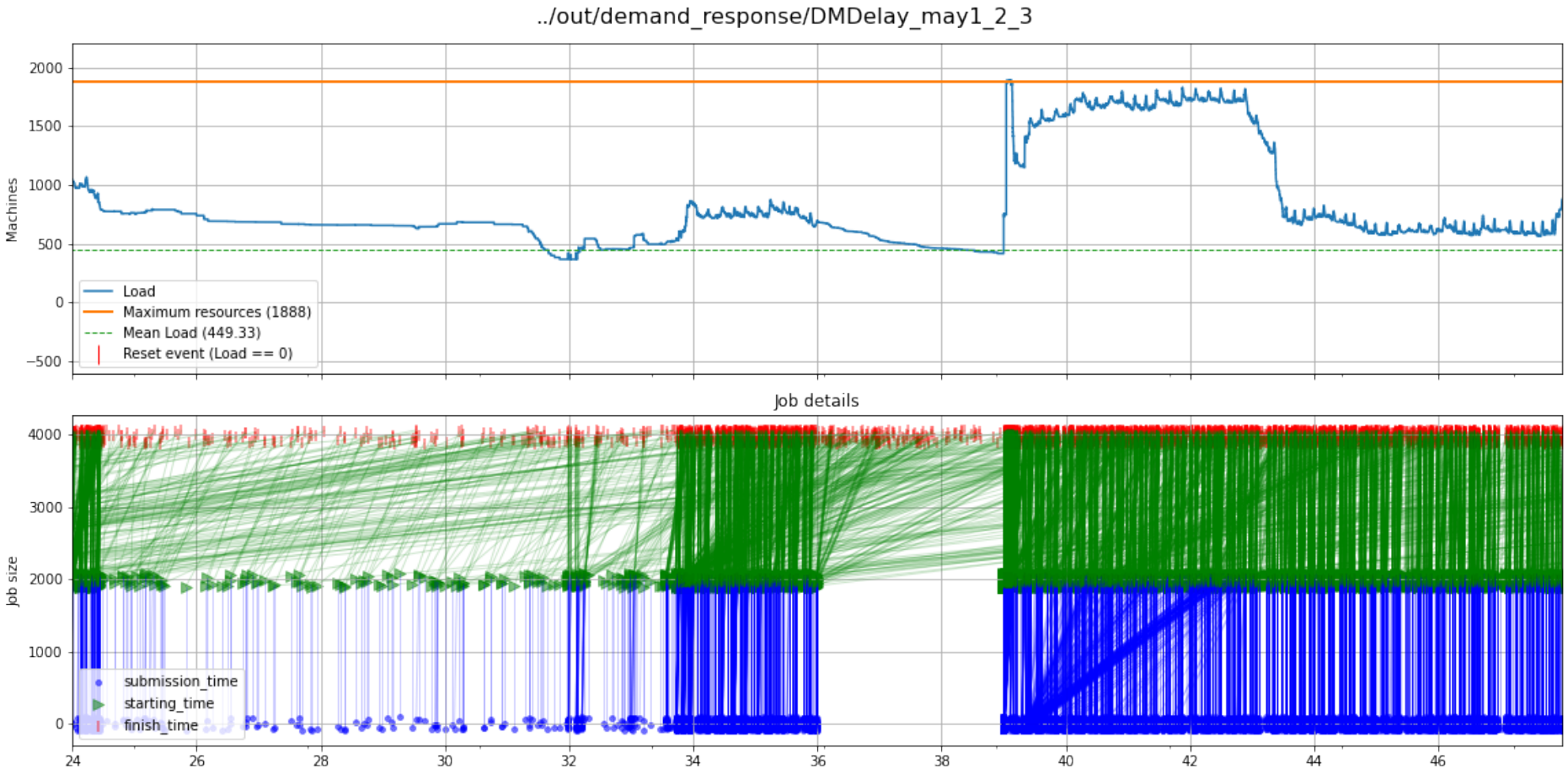
Illustration

- Behavior during demand response window: **renounce**



Illustration

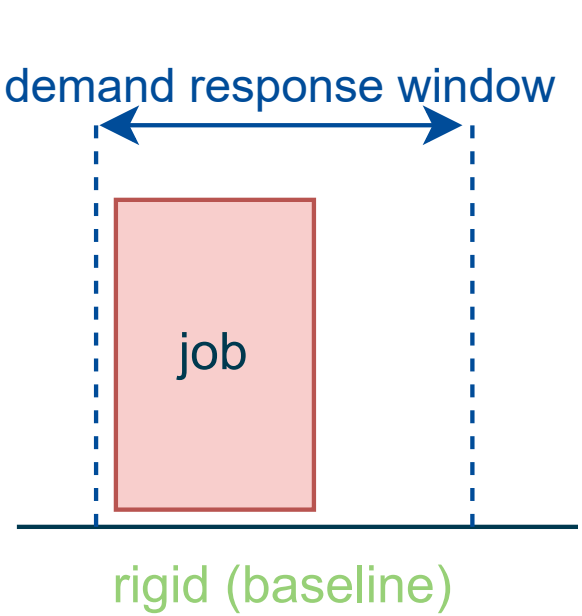
- Behavior during demand response window: **delay**

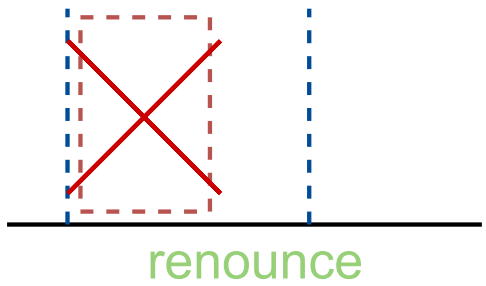


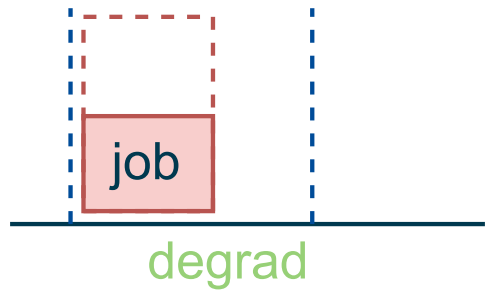
Simulating demand response behaviors

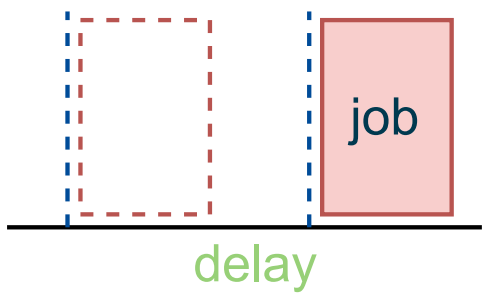
Behaviors

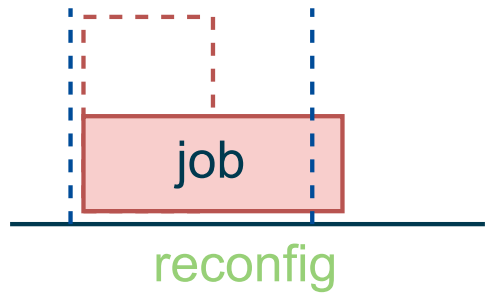
- Five behaviors studied:



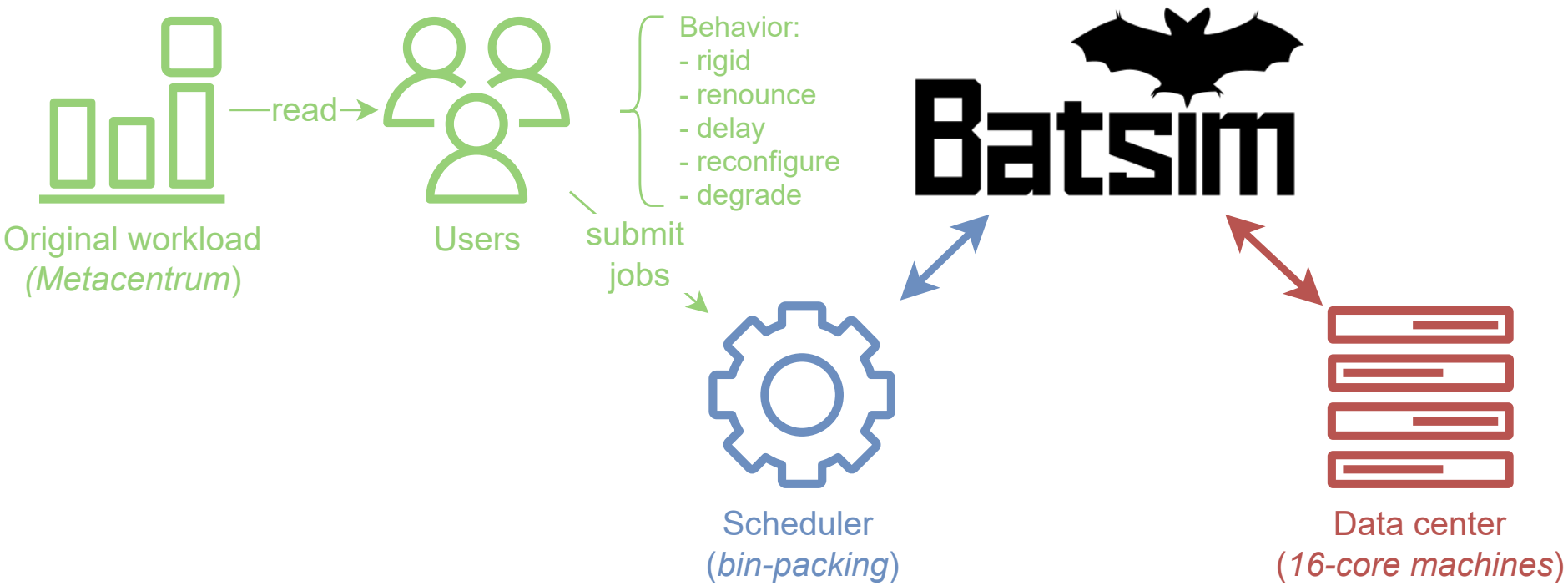








The simulated system



Experimental campaign

Description of the exp. campaign

1. Cleaning the original workload (Metacentrum2, from the Parallel Workloads Archive)
   - keep only requested cores < 16 and execution time < 1 day

Description of the exp. campaign

1. Cleaning the original workload (Metacentrum2, from the Parallel Workloads Archive)
   - keep only requested cores < 16 and execution time < 1 day
2. Sizing the platform to fit the workload

Description of the exp. campaign

1. Cleaning the original workload (Metacentrum2, from the Parallel Workloads Archive)
   - keep only requested cores < 16 and execution time < 1 day
2. Sizing the platform to fit the workload
3. Selecting 105 days to run the experiments
   - all weekdays between Jun 1, 2014 and Oct 23, 2014

Description of the exp. campaign

1. Cleaning the original workload (Metacentrum2, from the Parallel Workloads Archive)
   - keep only requested cores < 16 and execution time < 1 day
2. Sizing the platform to fit the workload
3. Selecting 105 days to run the experiments
   - all weekdays between Jun 1, 2014 and Oct 23, 2014
4. Simulating the five behaviors for each day
   - starting the simulation one day before and stopping one day after to avoid side effects
   - demand response event arises at 16:00 on day 2 and lasts 1 (resp. 4) hour

Description of the exp. campaign

1. Cleaning the original workload (Metacentrum2, from the Parallel Workloads Archive)
   - keep only requested cores < 16 and execution time < 1 day
2. Sizing the platform to fit the workload
3. Selecting 105 days to run the experiments
   - all weekdays between Jun 1, 2014 and Oct 23, 2014
4. Simulating the five behaviors for each day
   - starting the simulation one day before and stopping one day after to avoid side effects
   - demand response event arises at 16:00 on day 2 and lasts 1 (resp. 4) hour

(FYI the exp. campaign ran in <2h on a general purpose machine.)

A characterization of each behavior

Experimental results

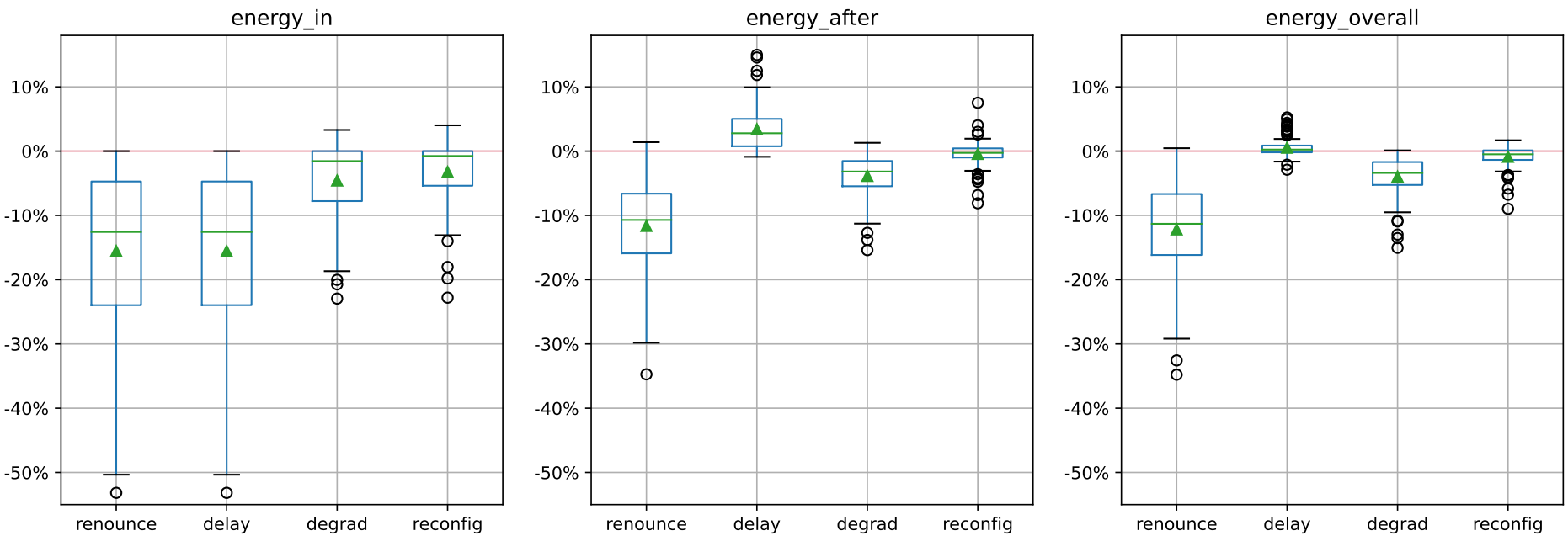
Metrics observed

What can we expect from each behavior in terms of

- energy saved
  - in the window
  - after the window
  - overall
- impact on the scheduling (= QoS)
  - $waiting_{time} = starting_{time} - submission_{time}$
  - $slowdown = (finish_{time} - submission_{time})/execution_{time}$

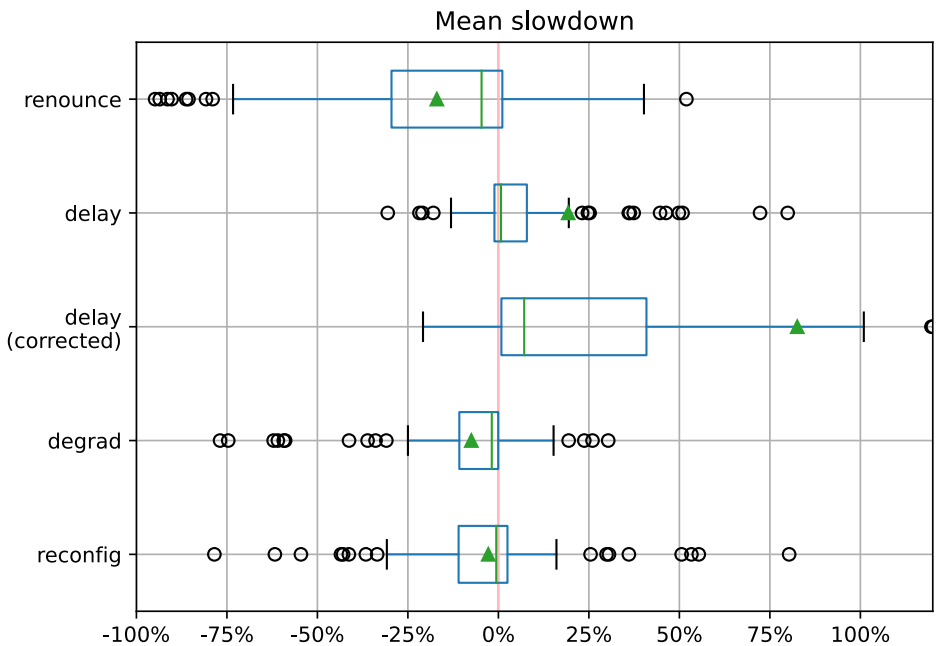
Energy metrics

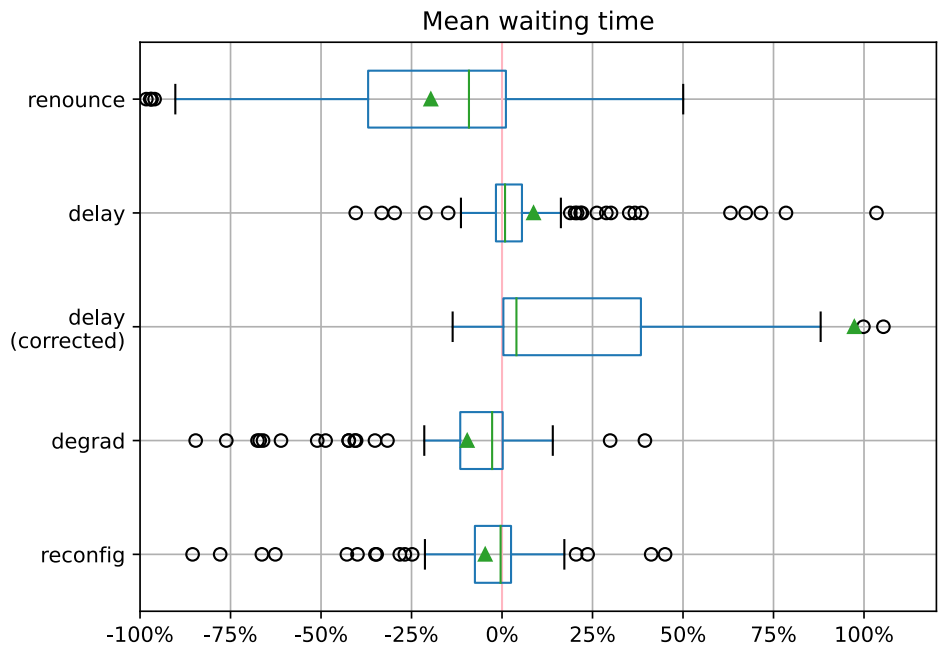
- In % difference from the baseline (“rigid” behavior)
- Window 4 hours:



Scheduling metrics

- In % difference from the baseline (“rigid” behavior)
- Window 1 hour:

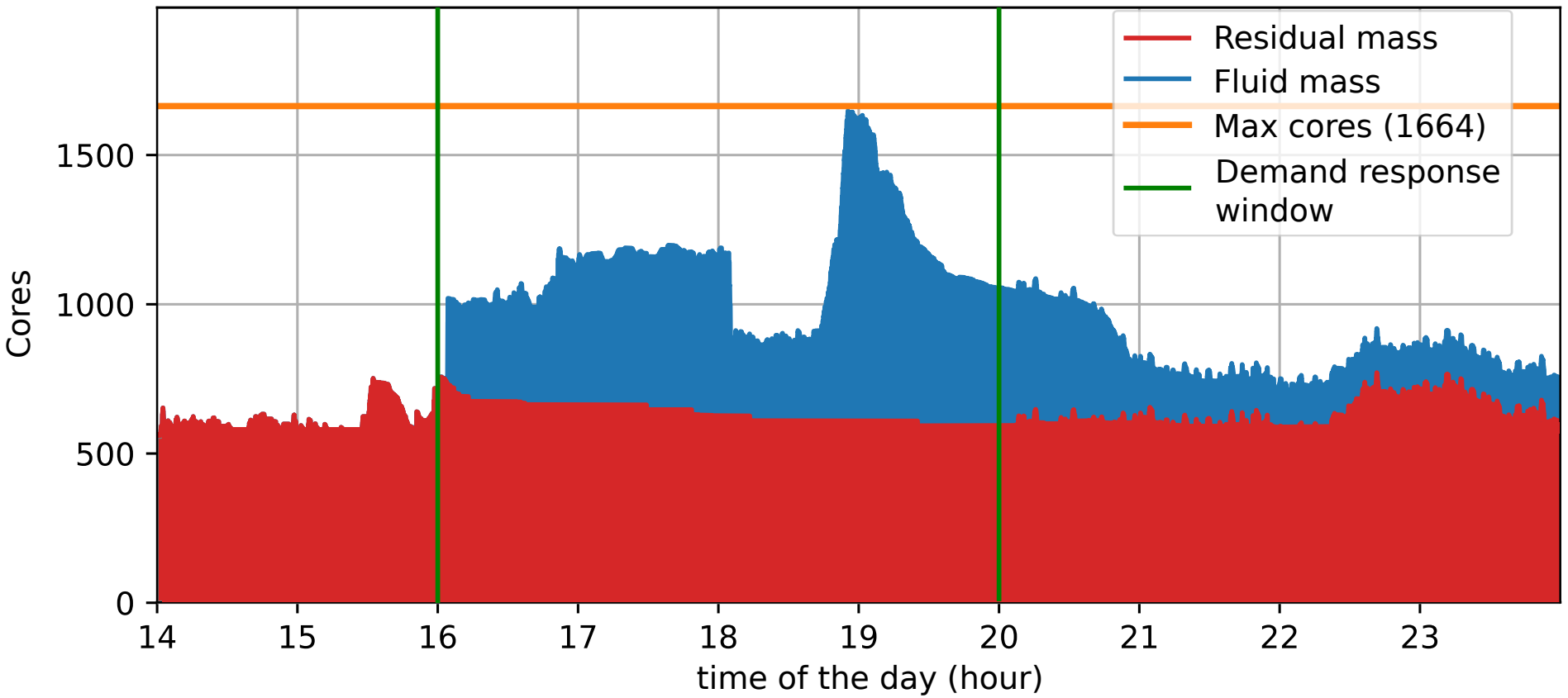




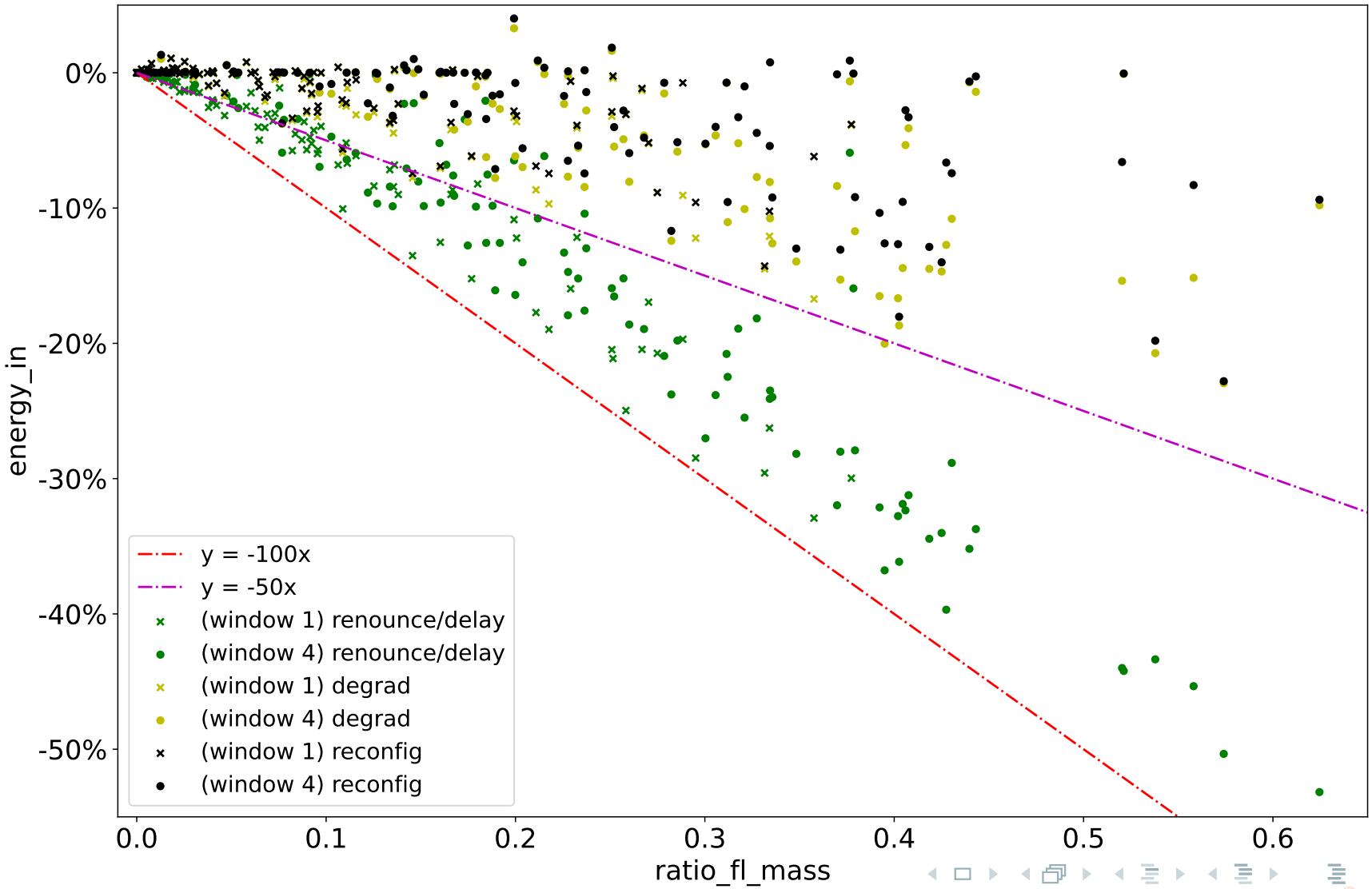
A characterization of each behavior

Explanation of energy results

Fluid and residual mass



Gains wrt fluid-residual ratio



Conclusion and research directions

Conclusion

- Pros and cons of each behavior:

behavior	energy in	energy overall	sched. metrics	"acceptability"
renounce	1st	1st	1st*	4th
delay	1st	4th	4th	2nd
degrad	3rd	2nd	2nd	3rd
reconfig	4th	3rd	3rd	1st

Conclusion

- Pros and cons of each behavior:

behavior	energy in	energy overall	sched. metrics	"acceptability"
renounce	1st	1st	1st*	4th
delay	1st	4th	4th	2nd
degrad	3rd	2nd	2nd	3rd
reconfig	4th	3rd	3rd	1st

- user submission patterns = a lever among others
  - has some inertia
  - leveraging the effort inside the scheduler seems essential
- quantify "acceptability", work on fairness, social sciences?

What's next?

1. Experimental part
   - Improve model (speedup, ...)
   - Try other schedulers (FCFS, easy backfilling, ...)
   - Try with anticipation
   - Home-made scheduler leveraging the user effort

What's next?

1. Experimental part
   - Improve model (speedup, ...)
   - Try other schedulers (FCFS, easy backfilling, ...)
   - Try with anticipation
   - Home-made scheduler leveraging the user effort
2. User model
   - Analyze actual behaviors
   - More realistic user submission patterns (Feitelson et al. Resampling with Feedback)

What's next?

1. Experimental part
   - Improve model (speedup, ...)
   - Try other schedulers (FCFS, easy backfilling, ...)
   - Try with anticipation
   - Home-made scheduler leveraging the user effort
2. User model
   - Analyze actual behaviors
   - More realistic user submission patterns (Feitelson et al. Resampling with Feedback)
3. Going further: what is a "sufficient data center"?
   - reflection to have on our basic digital needs
   - tools and relevant feedback to empower the users

Thank you for your attention!

- All the details in the article
- Material to reproduce the experiments:
  - experiment repository: gitlab.irit.fr/sepia-pub/open-science/demand-response-user
  - plugin "Batmen" for Batsim: gitlab.irit.fr/sepia-pub/mael/batmen
- Contact me:
  - www.irit.fr/~Mael.Madon
  - mael.madon@irit.fr