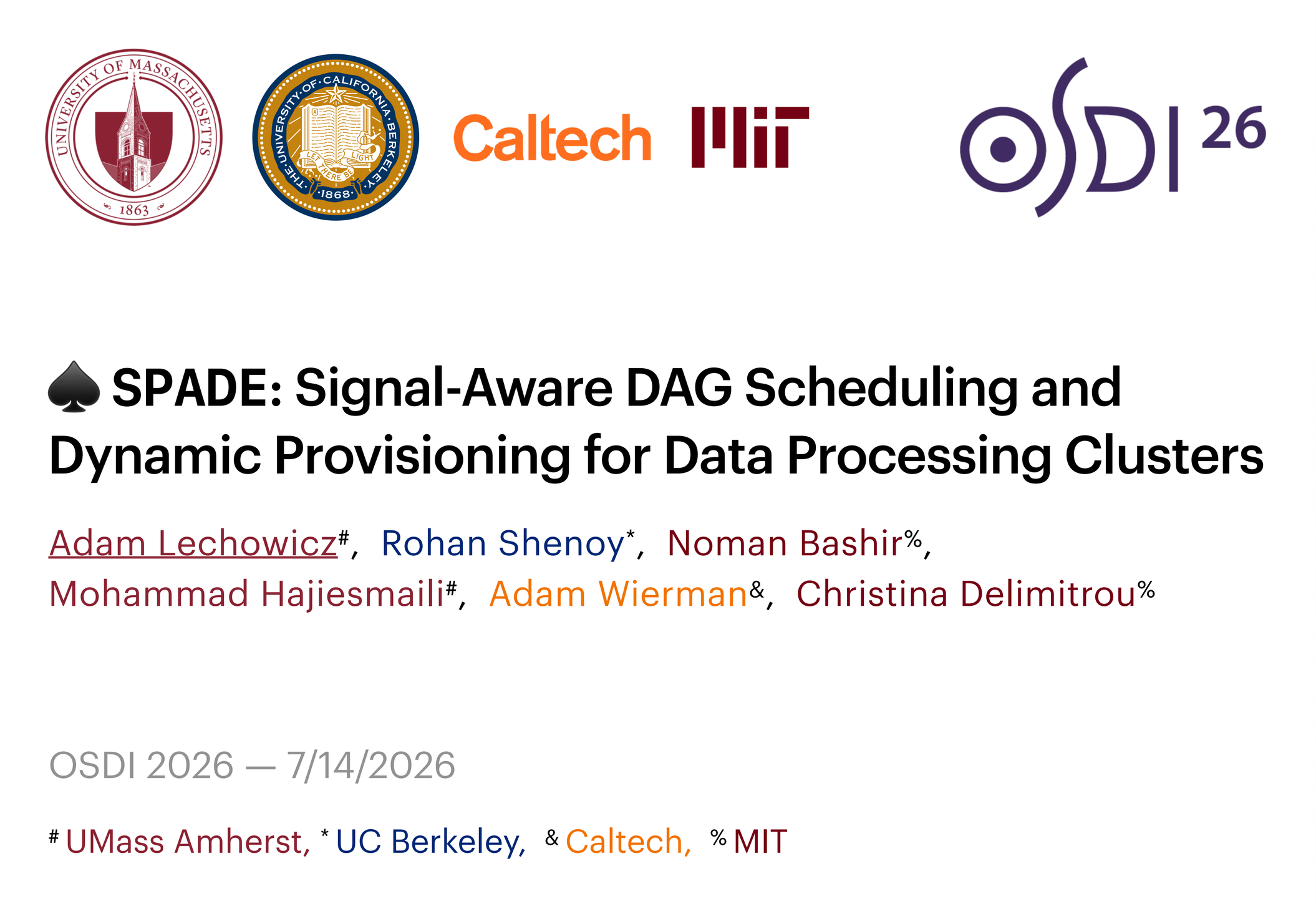


♠ SPADE: Signal-Aware DAG Scheduling and Dynamic Provisioning for Data Processing Clusters

Adam Lechowicz[#], Rohan Shenoy[*], Noman Bashir[%],
Mohammad Hajiesmaili[#], Adam Wierman[&], Christina Delimitrou[%]

OSDI 2026 — 7/14/2026

[#] UMass Amherst, [*] UC Berkeley, [&] Caltech, [%] MIT

Motivation

- Data center growth colliding with shifting expectations & constraints…

Motivation

- Data center growth colliding with shifting expectations & constraints...



Motivation

- **Data center growth colliding with shifting expectations & constraints…**

WIRED

MOLLY TAFT SCIENCE APR 22, 2026 6:30 AM

New Gas-Powered Data Centers Could Emit More Greenhouse Gases Than Entire Nations

A WIRED review of permits for data center projects using natural gas and linked to OpenAI, Meta, Microsoft, and xAI shows they could emit more than 129 million tons of greenhouse gases per year.

Reuters

Stressed US grid forcing data centers to get more flexible

By **Laila Kearney**

March 26, 2026 3:40 PM EDT · Updated March 26, 2026

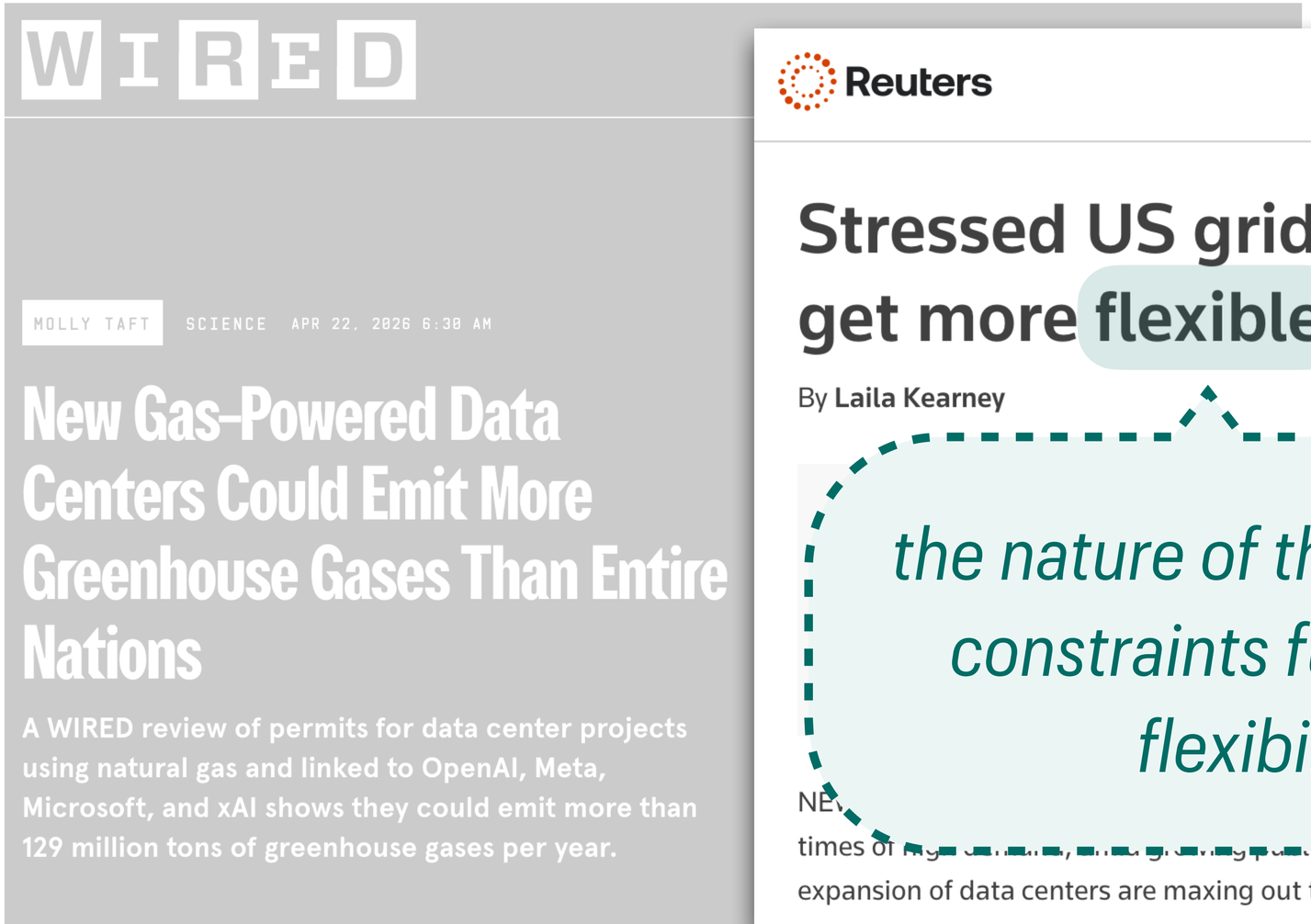
Summary Companies

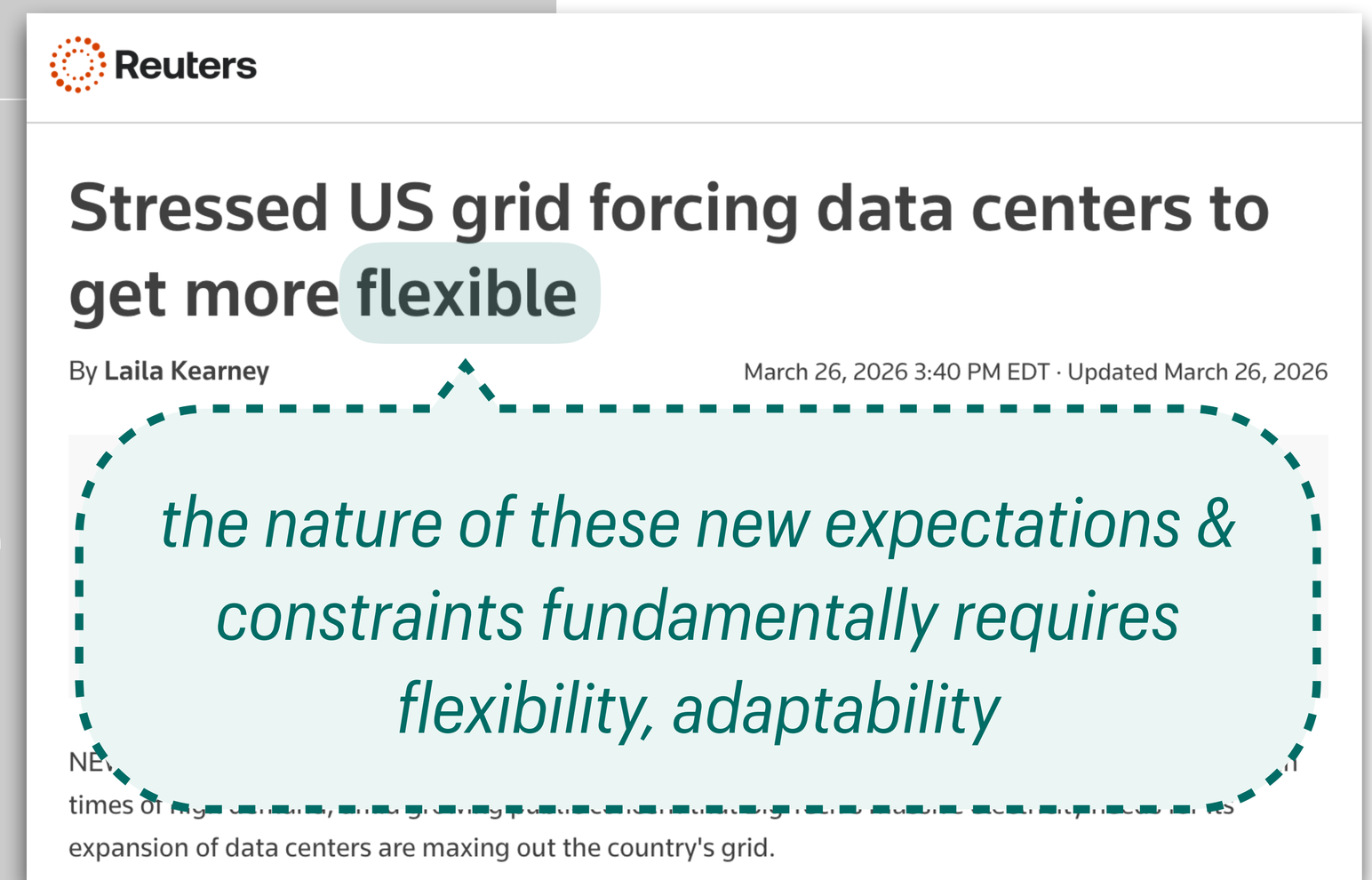
- Data centers urged to reduce power use during peak demand
- US electricity use from data centers could quadruple by decade's end, study says
- Big Tech is beginning to test how it can shrink data center power demand when grid demand peaks

NEW YORK, March 26 (Reuters) - The U.S. technology industry is being pushed to shrink its power use in times of high demand, amid growing public concern that Big Tech's massive electricity needs for its expansion of data centers are maxing out the country's grid.

Motivation

- **Data center growth colliding with shifting expectations & constraints…**





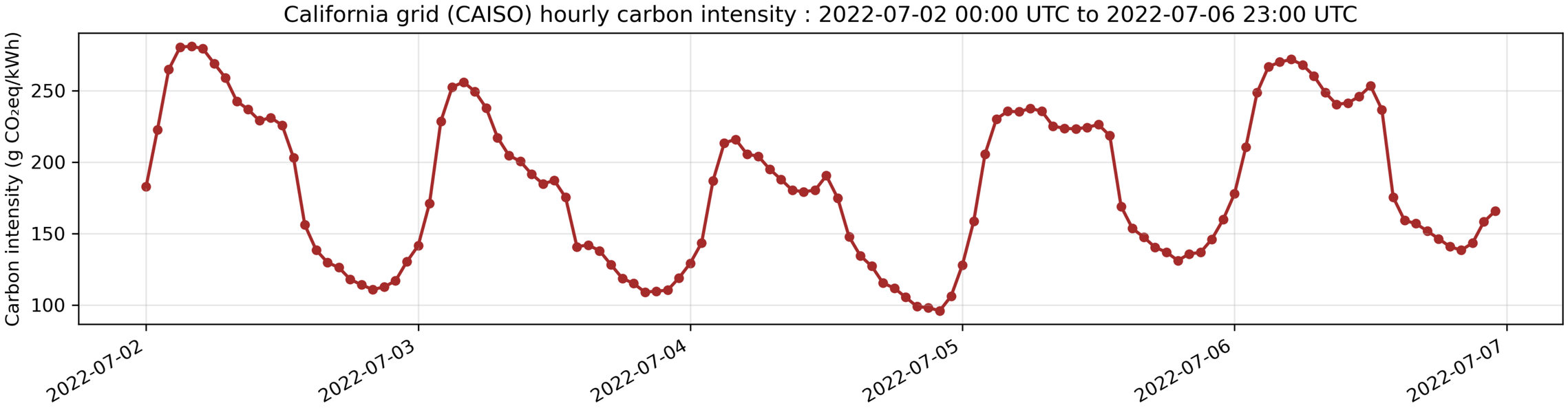
the nature of these new expectations & constraints fundamentally requires flexibility, adaptability

Motivation

- **Data center growth colliding with shifting expectations & constraints…**
  - need flexibility in response to some *time-varying signal*

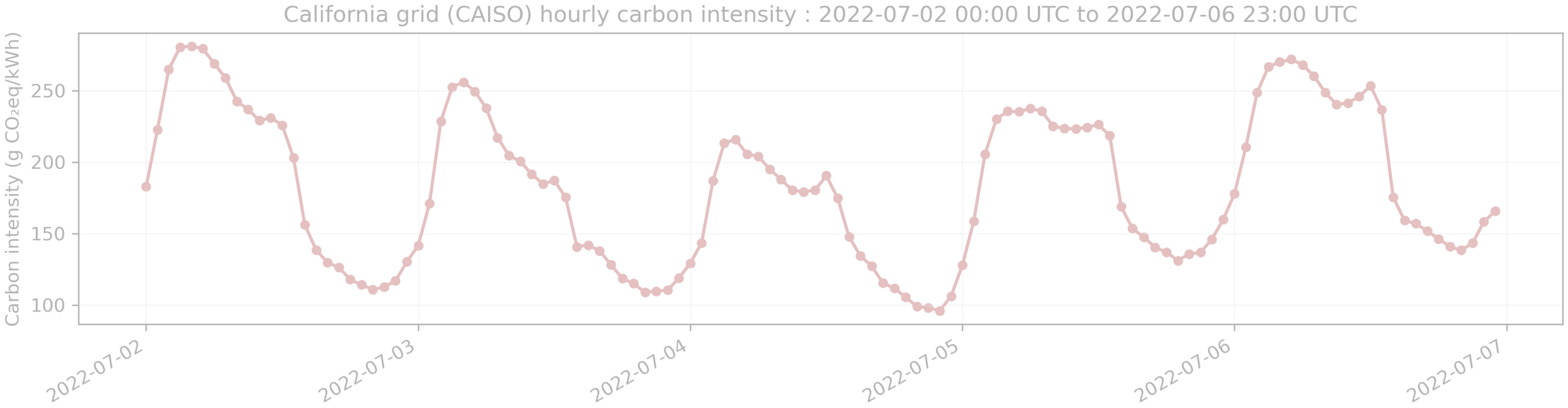
Motivation

- **Data center growth colliding with shifting expectations & constraints…**
  - need flexibility in response to some *time-varying signal*



Motivation

- **Data center growth colliding with shifting expectations & constraints…**
  - need flexibility in response to some *time-varying signal*



  - creates new systems problems in classic topics: resource allocation, scheduling, etc.

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)

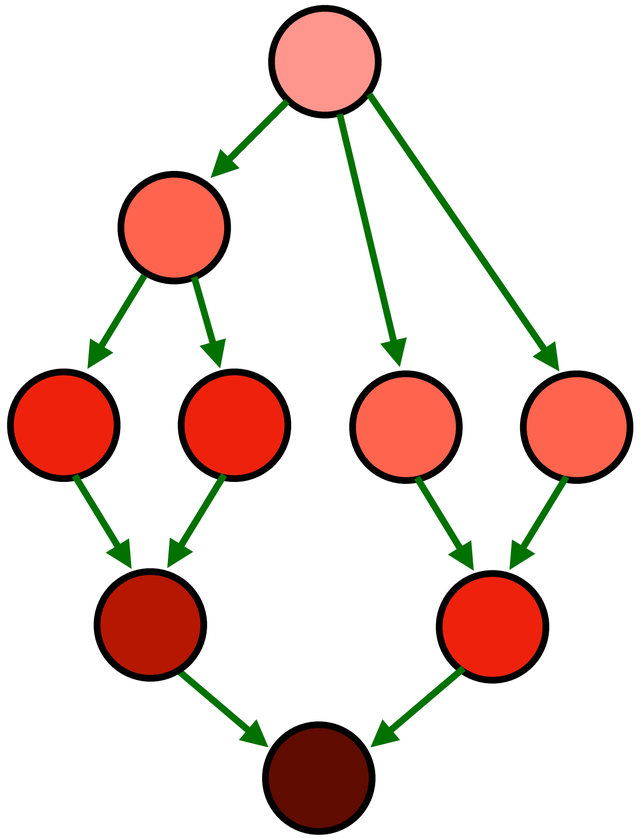
Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)
    - particularly common in big data processing (e.g., 

)

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)

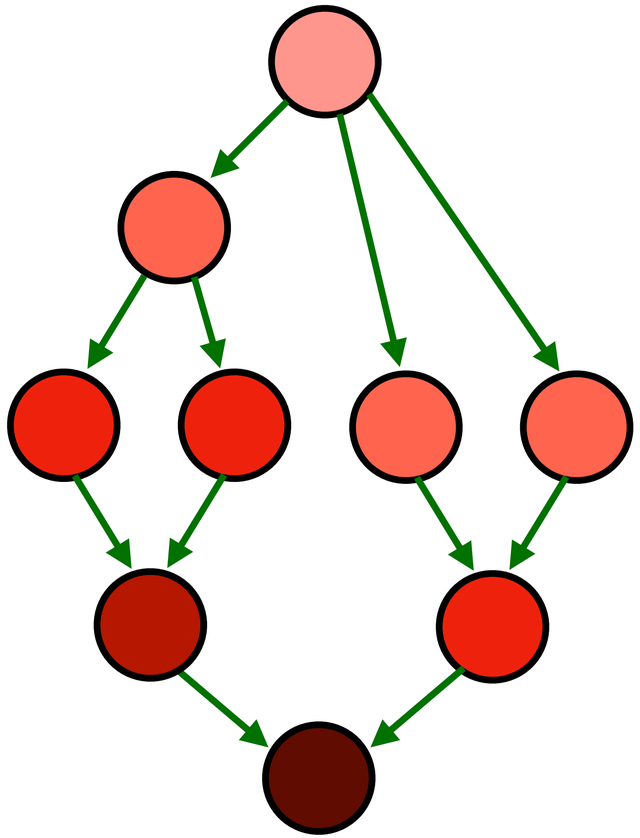


A DAG job

- **nodes** ≡ individual tasks
  (e.g., on data shards)
- **edges** ≡ precedence constraints
  (task must finish before another starts)

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)

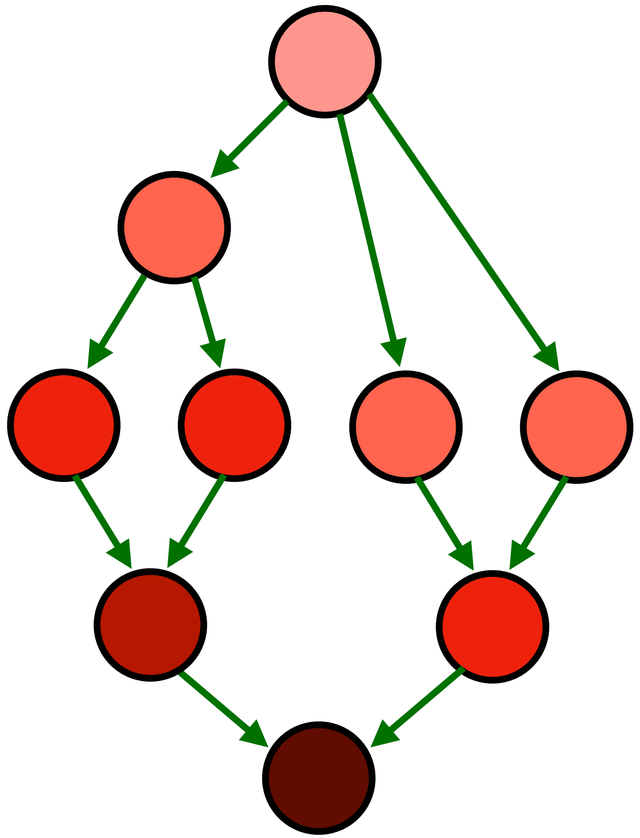


A DAG job

- **Signal-aware DAG execution problem**
  - Given a queue of DAG jobs and a time-varying signal revealed over time, decide:

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)

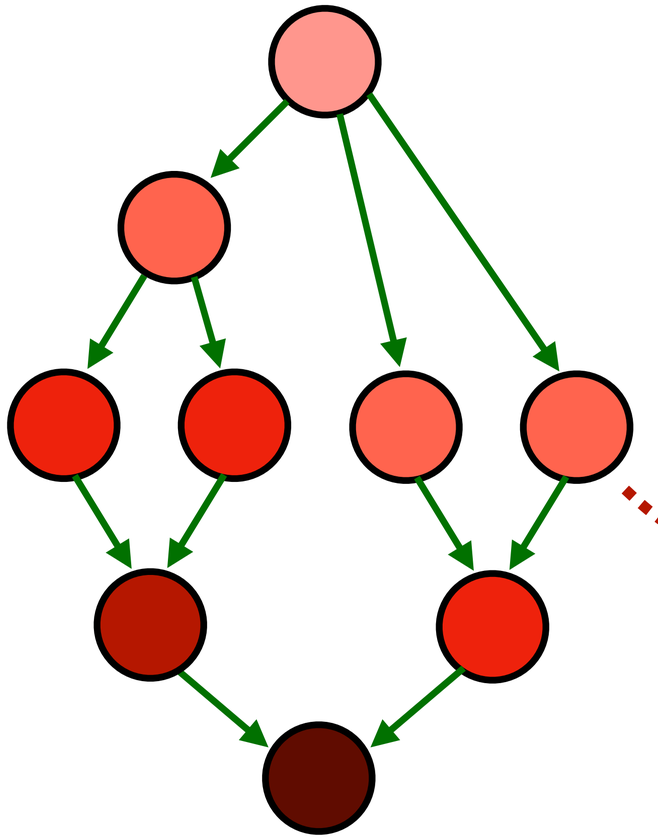


A DAG job

- **Signal-aware DAG execution problem**
  - Given a queue of DAG jobs and a time-varying signal revealed over time, decide:
    1. How many resources to use
       (e.g., no. of busy executors)

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)

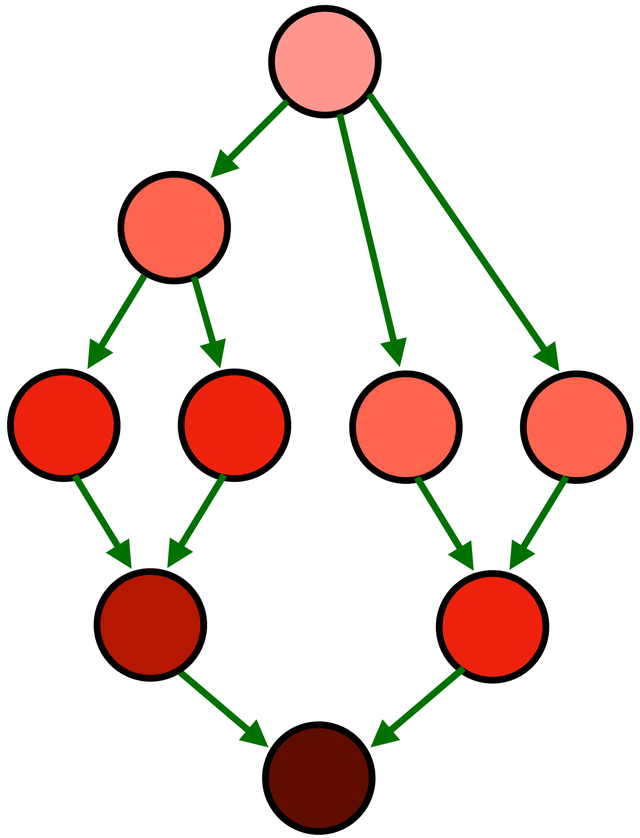


A DAG job

- **Signal-aware DAG execution problem**
  - Given a queue of DAG jobs and a time-varying signal revealed over time, decide:
    1. How many resources to use
       (e.g., no. of busy executors)
    2. Which tasks to execute across all jobs
       (i.e., scheduling the order of execution)

Problem

- Jobs with precedence constrained tasks (as directed-acyclic graphs)



A DAG job

- **Signal-aware DAG execution problem**
  - Given a queue of DAG jobs and a time-varying signal revealed over time, decide:
    1. How many resources to use
    2. Which tasks to execute across all jobs
  - **Objective:** achieve a desired trade-off between performance (e.g., min. makespan) and *secondary objective* (defined by signal)

Context

- Scheduling DAG jobs (under classic systems metrics) is well-studied.[1,2]

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters*.
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters*.

Context

- Scheduling DAG jobs (under classic systems metrics) is well-studied.[1,2]
- Provisioning in response to a signal (e.g., carbon-aware systems) has also been well-studied in recent years.[3,4]

3. Wiesner et al. (Middleware '21) *Let's Wait Awhile: How Temporal Workload Shifting Can Reduce Carbon Emissions in the Cloud.*

4. Colangelo et al. (Nature Energy '25) *AI data centres as grid-interactive assets.*

Context

- Scheduling DAG jobs (under classic systems metrics) is well-studied.[1,2]
- Provisioning in response to a signal (e.g., carbon-aware systems) has also been well-studied in recent years.[3,4]
- Relatively fewer studies have considered scheduling DAGs with a second, **exogenous** objective (e.g., energy cost, variable-price compute).[5,6]

5. Su et al. (ACM TOMPECS '21) *Learning-Augmented Energy-Aware List Scheduling for Precedence-Constrained Tasks.*

6. Park and Ham. (Int. Journal of Production Economics '22) *Energy-aware flexible job shop scheduling under time-of-use pricing.*

Context

- Scheduling DAG jobs (under classic systems metrics) is well-studied.[1,2]
- Provisioning in response to a signal (e.g., carbon-aware systems) has also been well-studied in recent years.[3,4]
- Relatively fewer studies have considered scheduling DAGs with a second, exogenous objective (e.g., energy cost, variable-price compute).[5,6]
- **Key insight of our work:**
  - For signal-aware DAG execution, provisioning and scheduling are tightly coupled: *need an integrated approach to solve the problem*
  - make decisions with knowledge of both DAG structure & current signal

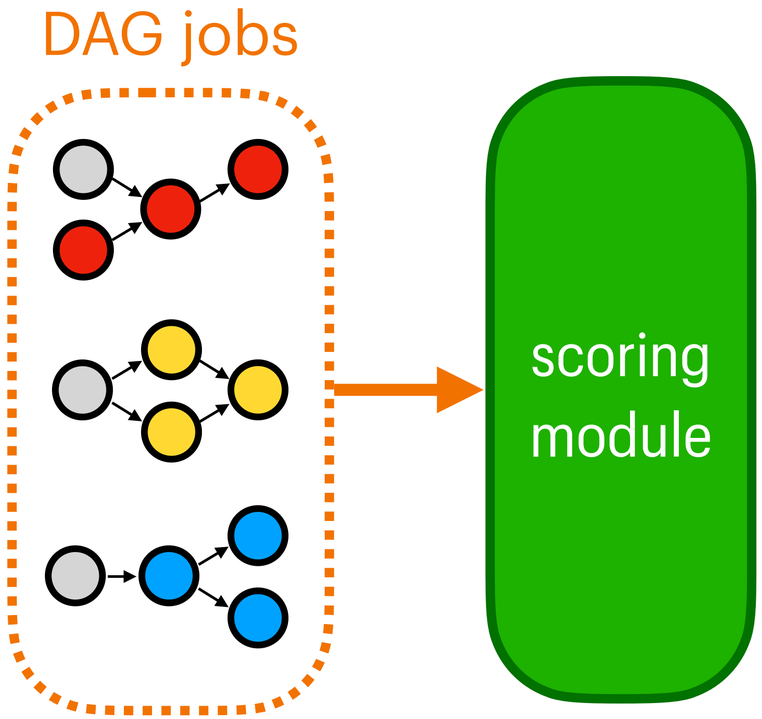
Identifying *bottleneck* tasks

- State-of-the-art techniques for DAG scheduling[1,2] reason over a *score or probability distribution* over tasks at each scheduling event.

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters.*
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters.*

Identifying *bottleneck* tasks

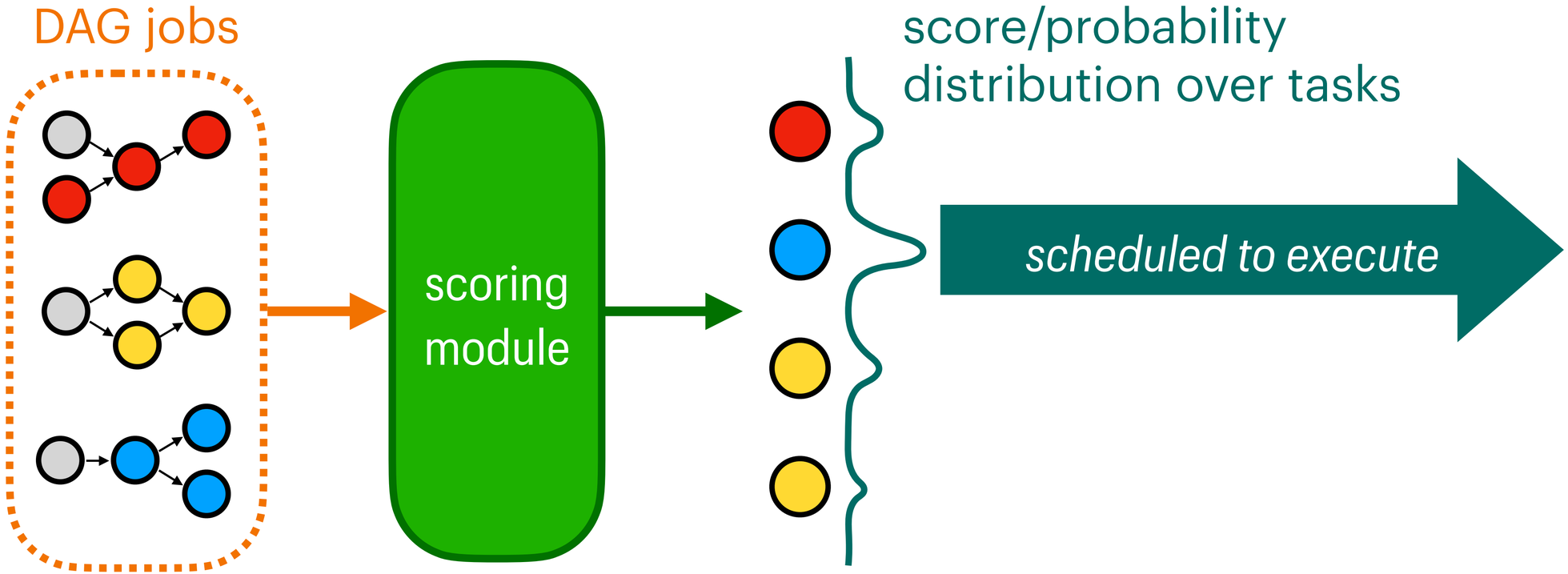
- State-of-the-art techniques for DAG scheduling[1,2] reason over a *score or probability distribution* over tasks at each scheduling event.



1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters.*
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters.*

Identifying *bottleneck* tasks

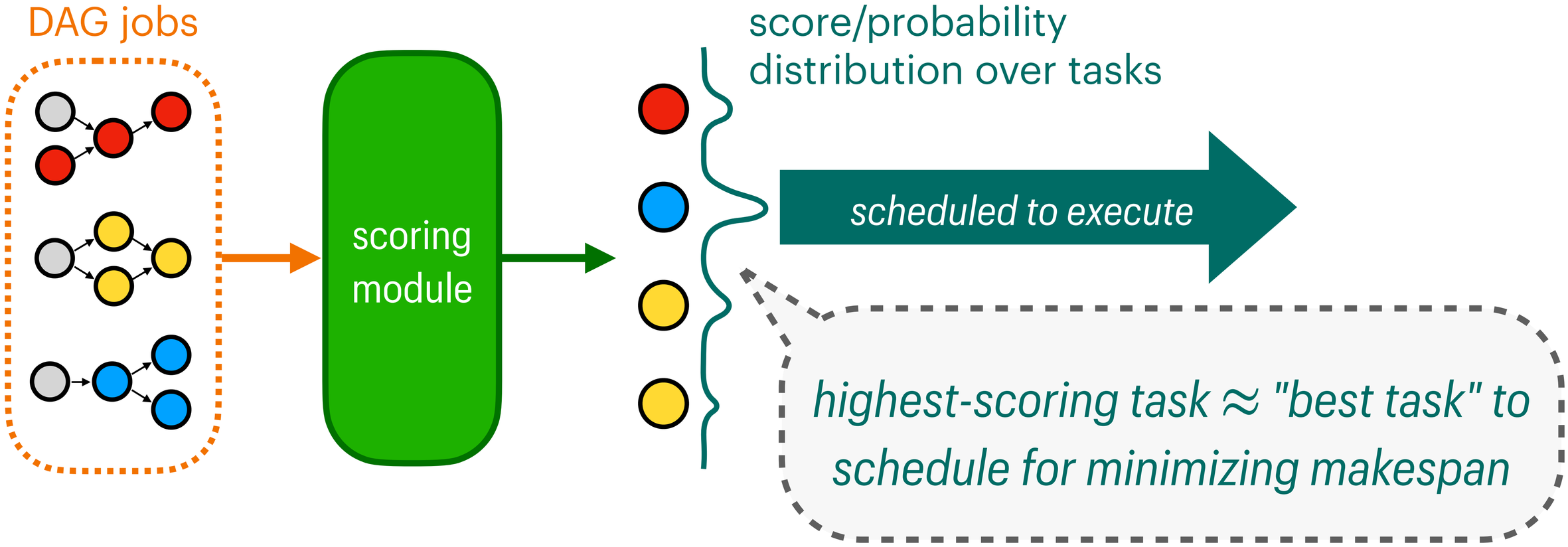
- State-of-the-art techniques for DAG scheduling[1,2] reason over a *score or probability distribution* over tasks at each scheduling event.



1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters.*
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters.*

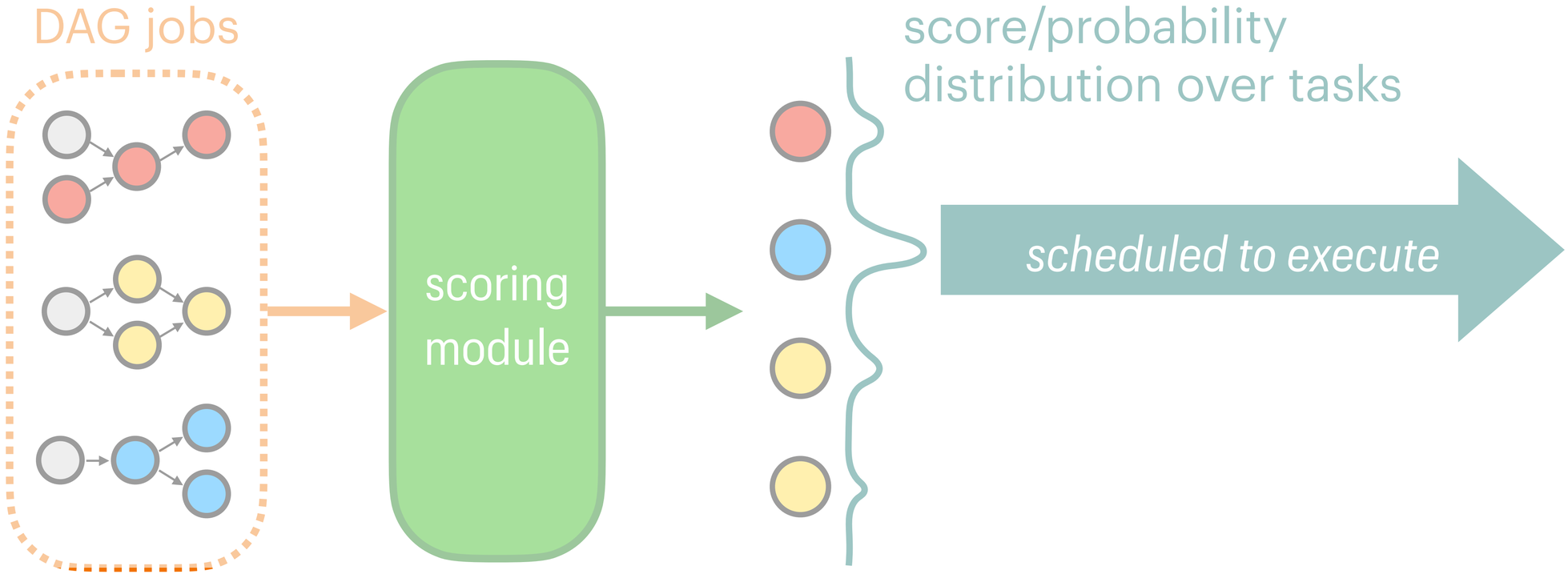
Identifying *bottleneck* tasks

- State-of-the-art techniques for DAG scheduling[1,2] reason over a *score or probability distribution* over tasks at each scheduling event.



Identifying *bottleneck* tasks

- State-of-the-art techniques for DAG scheduling[1,2] reason over a *score or probability distribution* over tasks at each scheduling event.



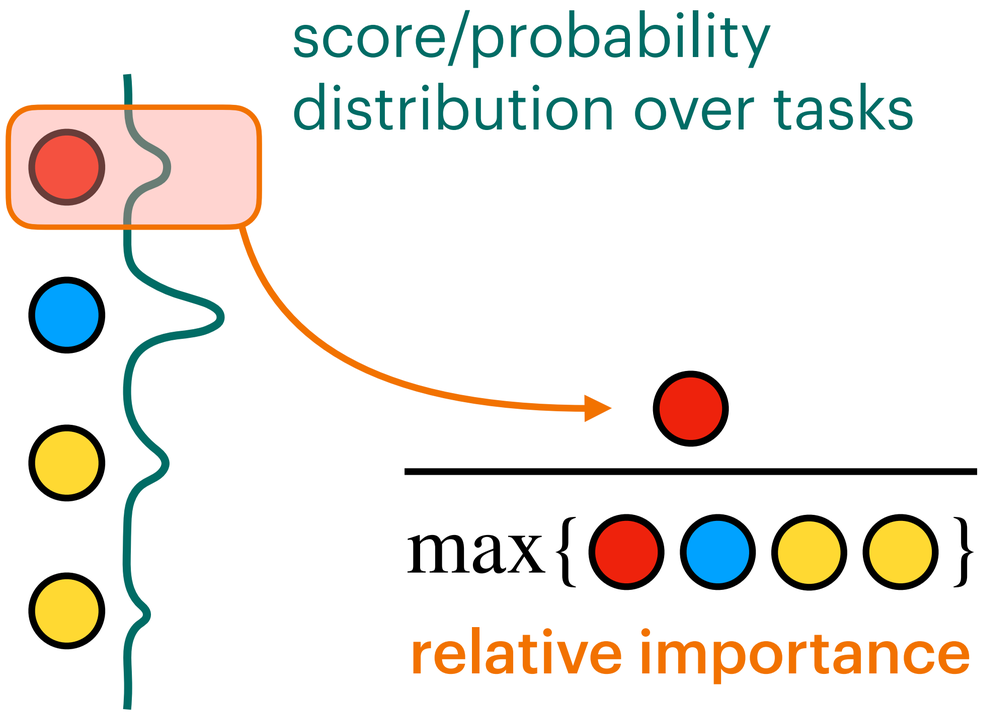
- Our system (SPADE) leverages insights provided by a score/probability scheme to reason about trade-offs (e.g., signal cost vs. makespan)

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.

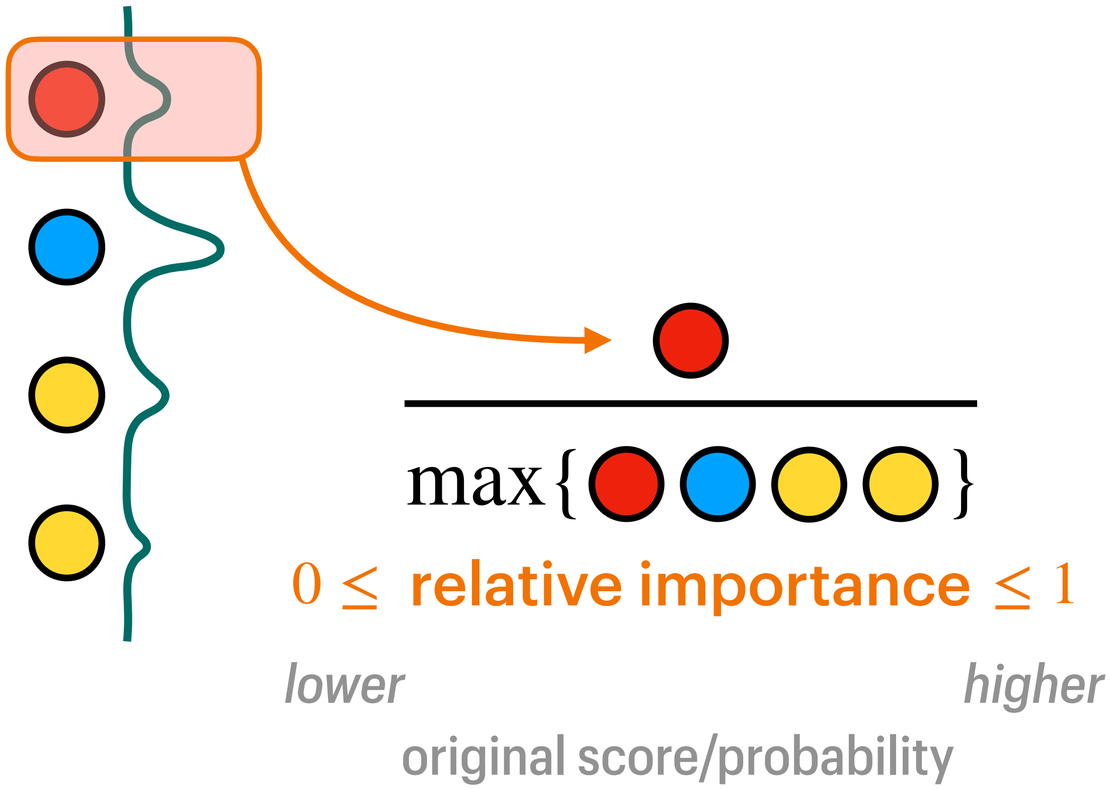
SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.



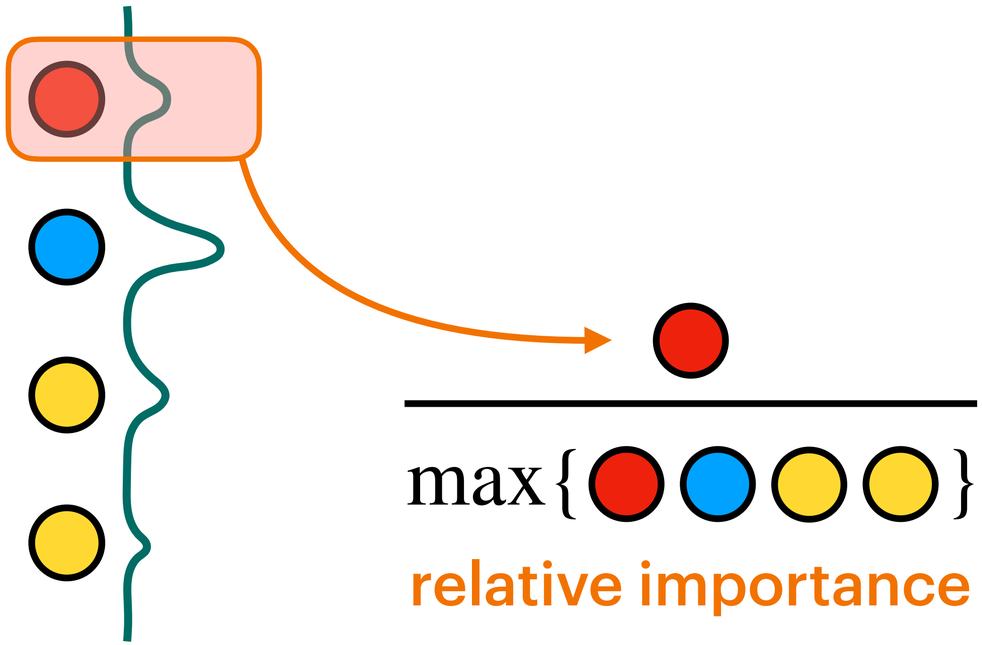
SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.



SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

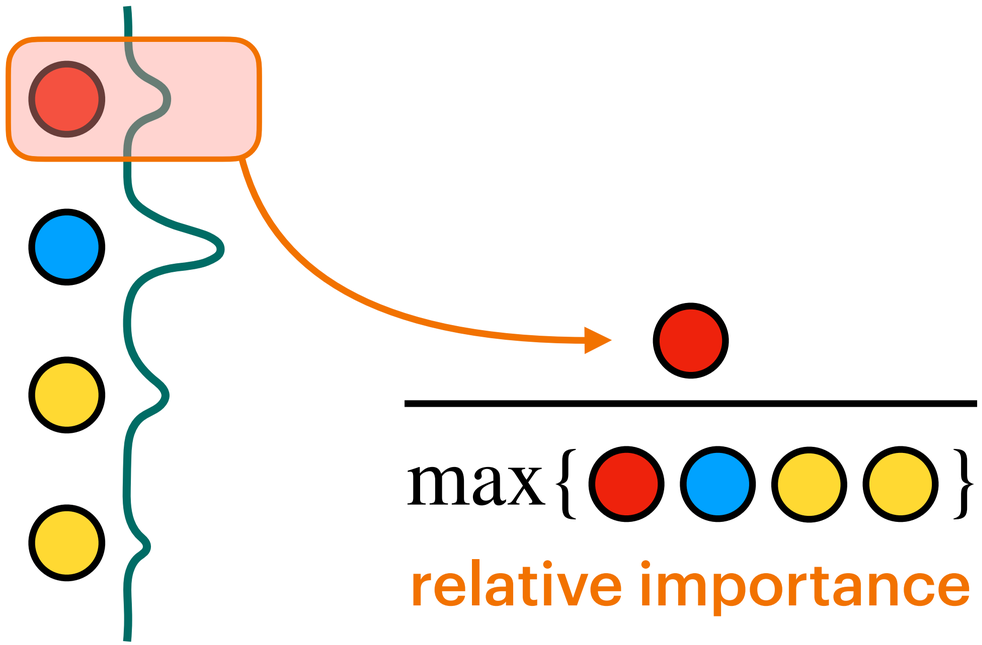
- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.



For a single task: trade-off between running it now (incurring current signal cost) or deferring it (incl. downstream tasks) to avoid using resources.

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.

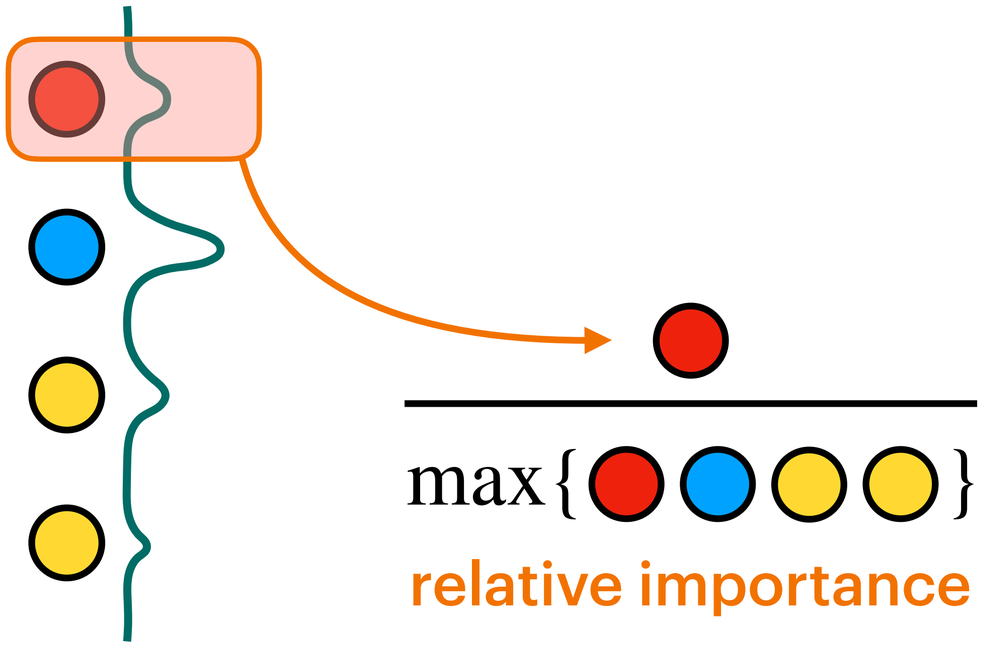


- SPADE quantifies this trade-off using a *signal-aware task filter* informed by theory.[7]

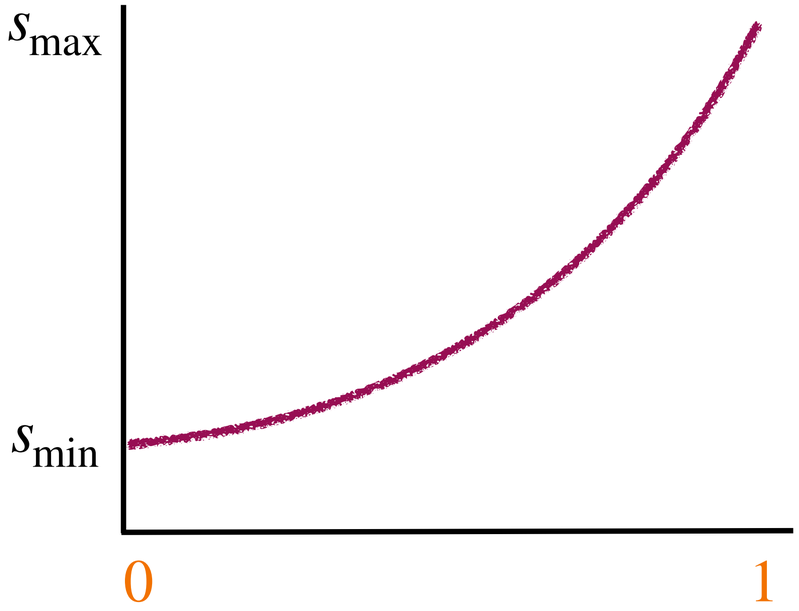
7. **L** et al. (SIGMETRICS/Performance '24) *Online Conversion with Switching Costs: Robust and Learning-Augmented Algorithms.*

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.

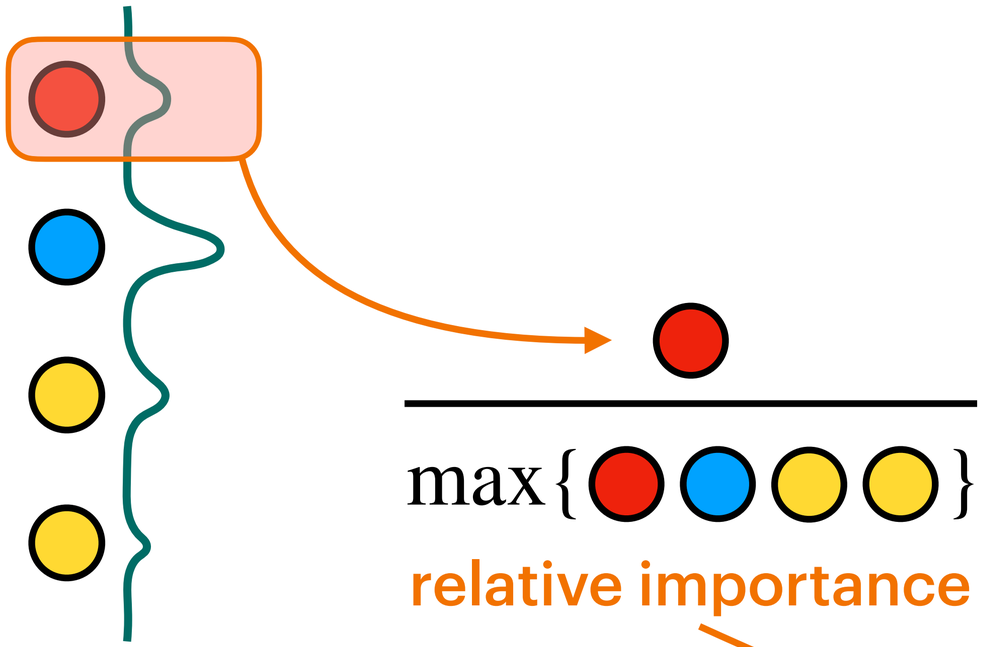


- SPADE quantifies this trade-off using a *signal-aware task filter* informed by theory.[7]



SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

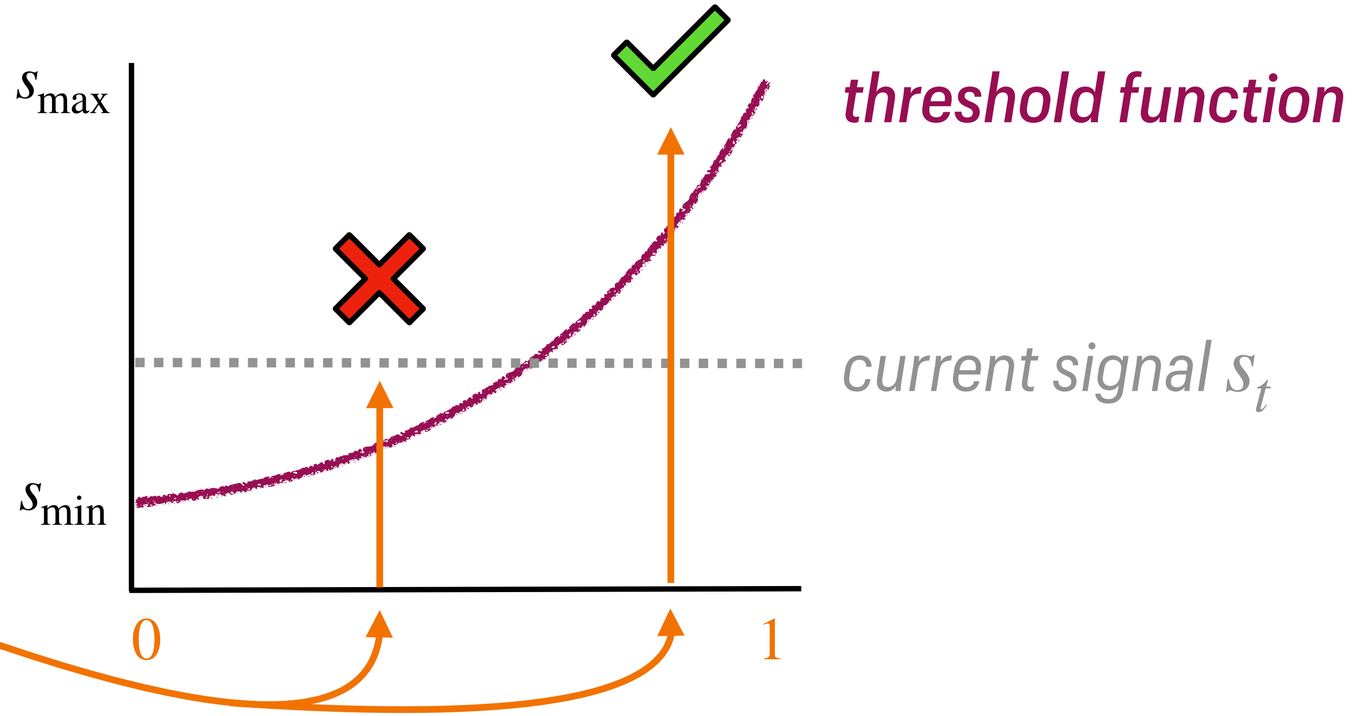
- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.



$$\frac{\bullet}{\max\{\bullet\bullet\bullet\}}$$

relative importance

- SPADE quantifies this trade-off using a *signal-aware task filter* informed by theory.[7]

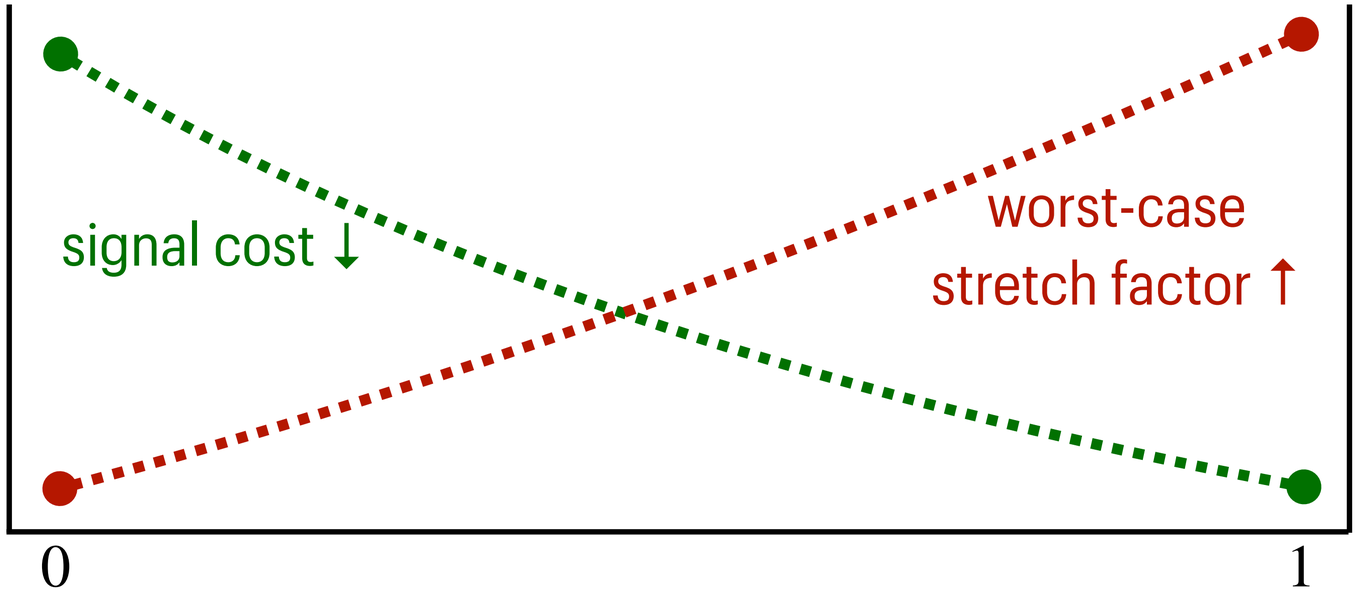


SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.
- We give (worst-case) theoretical guarantees on SPADE's *stretch factor*: how much worse makespan is relative to a signal-agnostic baseline.

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.
- We give (worst-case) theoretical guarantees on SPADE's *stretch factor*: how much worse makespan is relative to a signal-agnostic baseline.

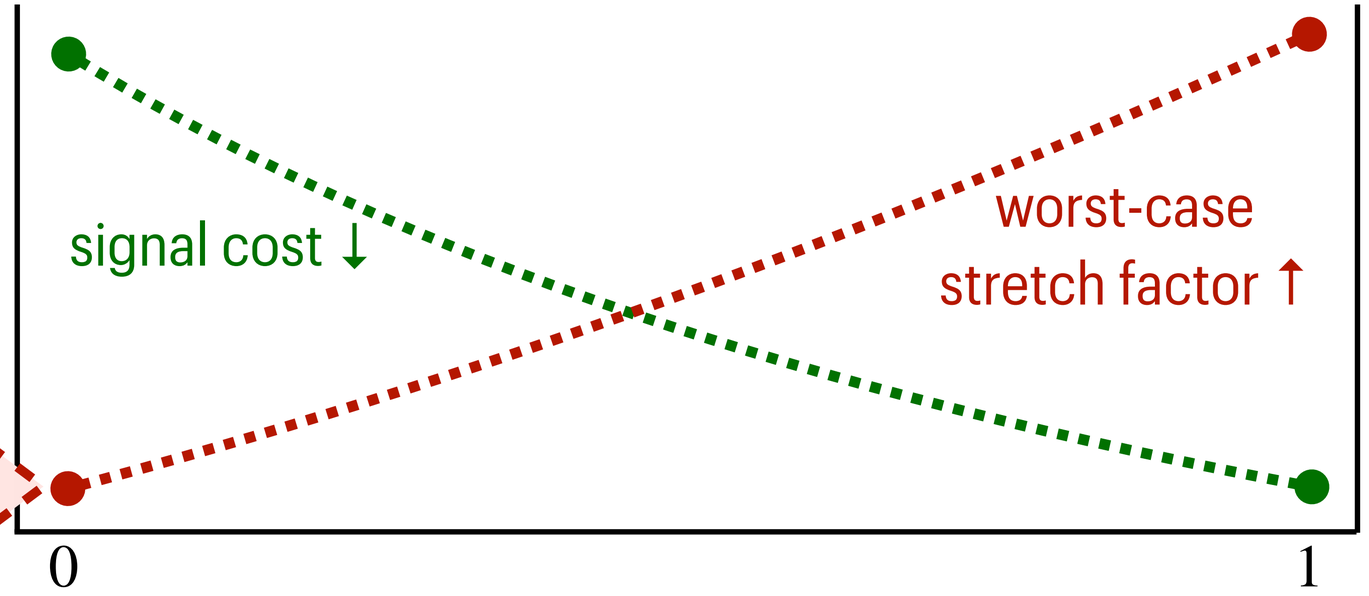


signal-awareness parameter $\gamma \in [0,1]$ (higher = more signal-aware)

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.
- We give (worst-case) theoretical guarantees on SPADE's *stretch factor*: how much worse makespan is relative to a signal-agnostic baseline.

matches makespan of signal-agnostic scheduler



signal-awareness parameter $\gamma \in [0,1]$ (higher = more signal-aware)

SPADE *(Scheduling & Provisioning for Adaptive DAG Execution)*

- Introduce a notion of *relative importance* to relate DAG structure against time-varying signal.
- We give (worst-case) theoretical guarantees on SPADE's *stretch factor*: how much worse makespan is relative to a signal-agnostic baseline.



Why not schedule and provision separately?

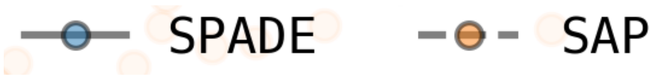
- DAG scheduling & signal-aware provisioning are well-studied in isolation…

Why not schedule and provision separately?

- DAG scheduling & signal-aware provisioning are well-studied in isolation…
- Visibility of *both* signal and DAGs ➜ superior trade-off between performance (e.g., JCT) and secondary objective.

Why not schedule and provision separately?

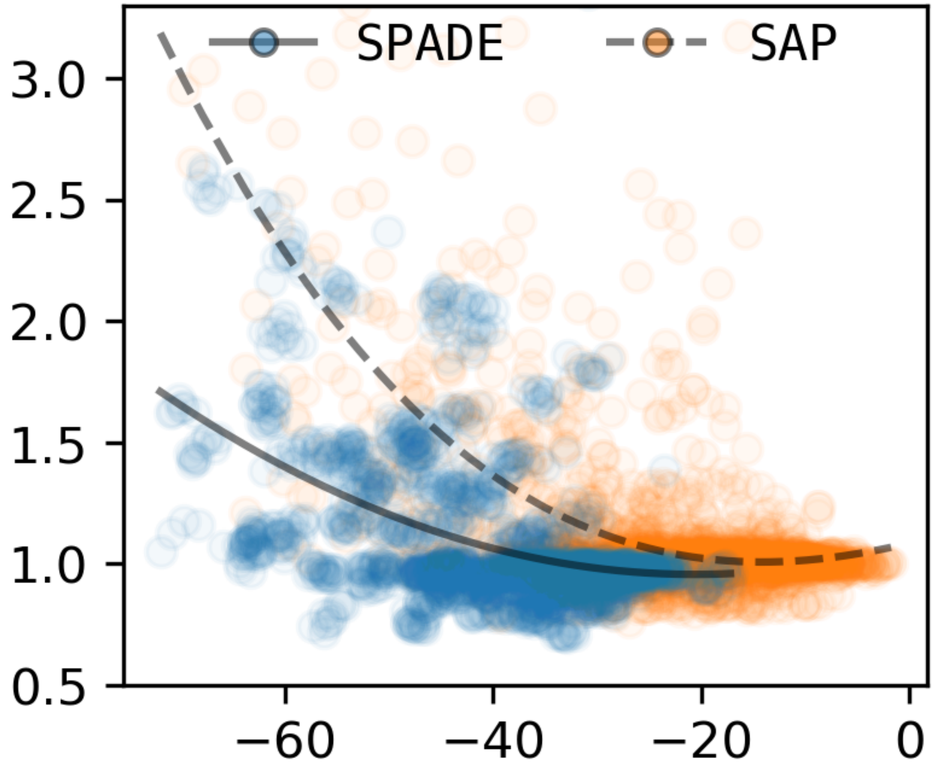
- Visibility of *both* signal and DAGs ➜ superior trade-off between performance (e.g., JCT) and secondary objective.



SAP ablation study:
signal-**a**ware **p**rovisioning layered on top of signal-unaware DAG scheduler

Why not schedule and provision separately?

- Visibility of *both* signal and DAGs ➜ superior trade-off between performance (e.g., JCT) and secondary objective.

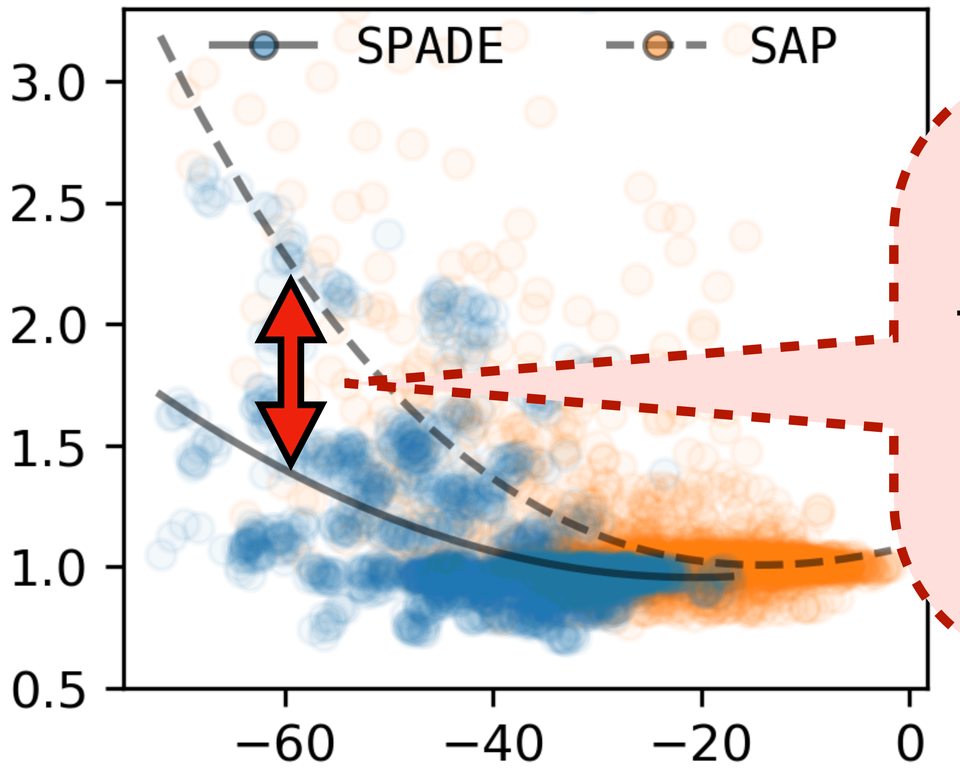


makespan increase (↓better)
(1.0 = default FIFO scheduler)

secondary objective reduction (↓better)
(% with respect to default FIFO scheduler)

Why not schedule and provision separately?

- Visibility of *both* signal and DAGs ➜ superior trade-off between performance (e.g., JCT) and secondary objective.



For the same savings in the secondary objective, SPADE sacrifices *much less* in performance.

Evaluation

- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters*.

Evaluation

- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])
- Signals (& objectives): carbon intensity[8] (carbon footprint reduction),
  power utilization[9] (power overload reduction)

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters.*

8. Electricity Maps. *https://www.electricitymaps.com/*

9. John Wilkes. (Apr. 2020) *Google cluster-usage traces v3. https://github.com/google/cluster-data/blob/master/ClusterData2019.md*

Evaluation

- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])
- Signals (& objectives): carbon intensity[8] (carbon footprint reduction), power utilization[9] (power overload reduction)
- Workloads: TPC-H Spark benchmarks[10], Alibaba trace DAG jobs[11]

10. TPC-H. (2018) *The TPC-H Benchmarks,*

11. Alibaba. (2018) *Cluster data collected from production clusters in Alibaba for cluster management research.*

Evaluation

- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])
- Signals (& objectives): carbon intensity[8] (carbon footprint reduction), power utilization[9] (power overload reduction)
- Workloads: TPC-H Spark benchmarks[10], Alibaba trace DAG jobs[11]
- Baselines: default (FIFO) scheduler, Decima[1], Graphene[2], *more in full paper*

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters*.
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters*.

Evaluation

- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])
- Signals (& objectives): carbon intensity[8] (carbon footprint reduction), power utilization[9] (power overload reduction)
- Workloads: TPC-H Spark benchmarks[10], Alibaba trace DAG jobs[11]
- Baselines: default (FIFO) scheduler, Decima[1], Graphene[2], *more in full paper*
    - SAP adaptations of each existing signal-unaware scheduler

1. Mao et al. (SIGCOMM '19) *Learning Scheduling Algorithms for Data Processing Clusters*.
2. Grandl et al. (OSDI '16) *GRAPHENE: Packing and Dependency-Aware Scheduling for Data-Parallel Clusters*.

Evaluation

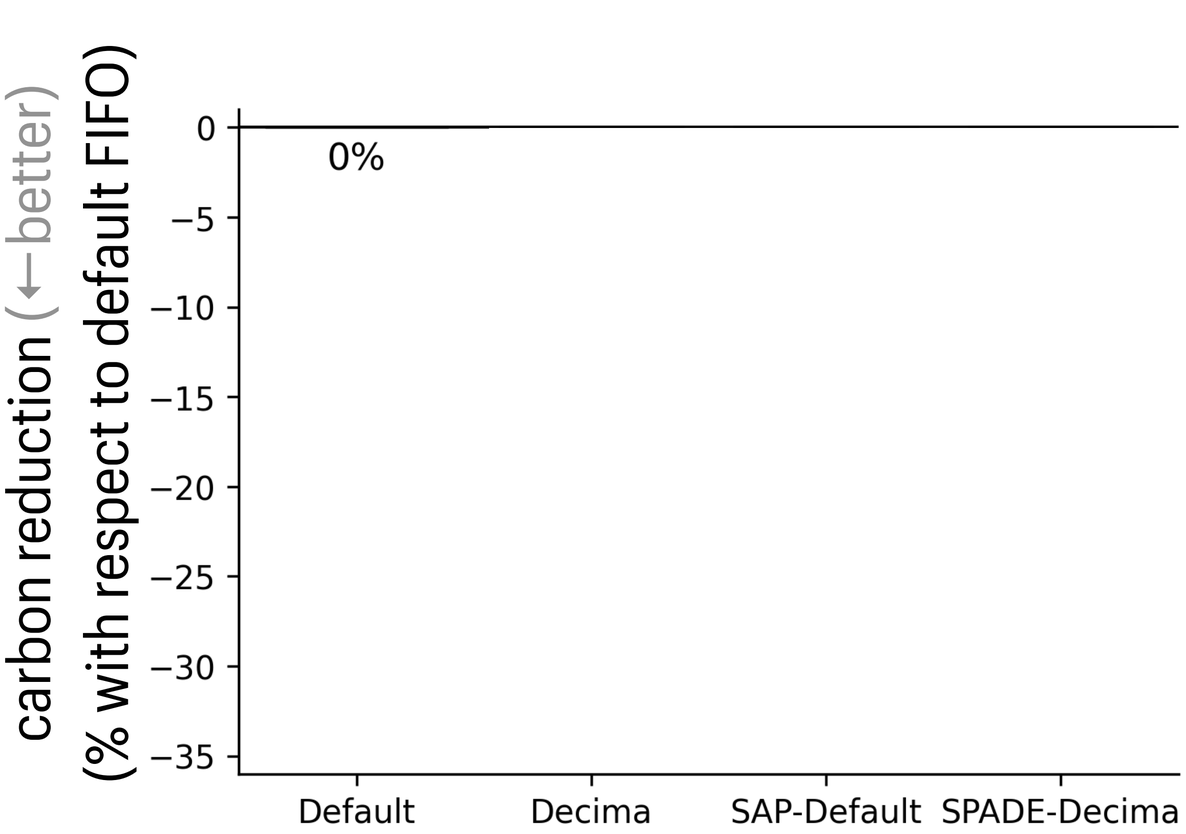
- Application: Apache Spark (100-node Kubernetes testbed + simulation env.[1])
- Signals (& objectives): carbon intensity[8] (carbon footprint reduction), power utilization[9] (power overload reduction)
- Workloads: TPC-H Spark benchmarks[10], Alibaba trace DAG jobs[11]
- Baselines: default (FIFO) scheduler, Decima[1], Graphene[2], *more in full paper*
  - SAP adaptations of each existing signal-unaware scheduler

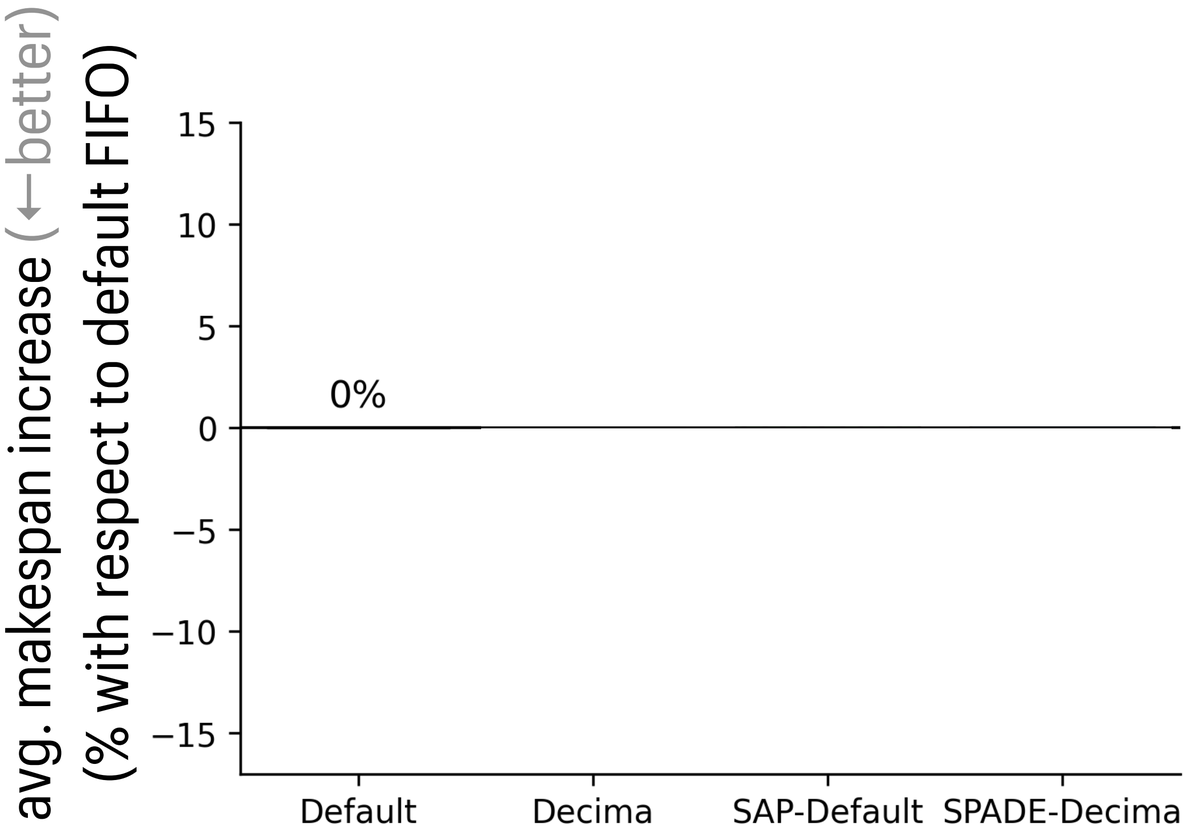


score/probability based: we extract their scoring modules for SPADE's relative importance metric

Evaluation — top-line results

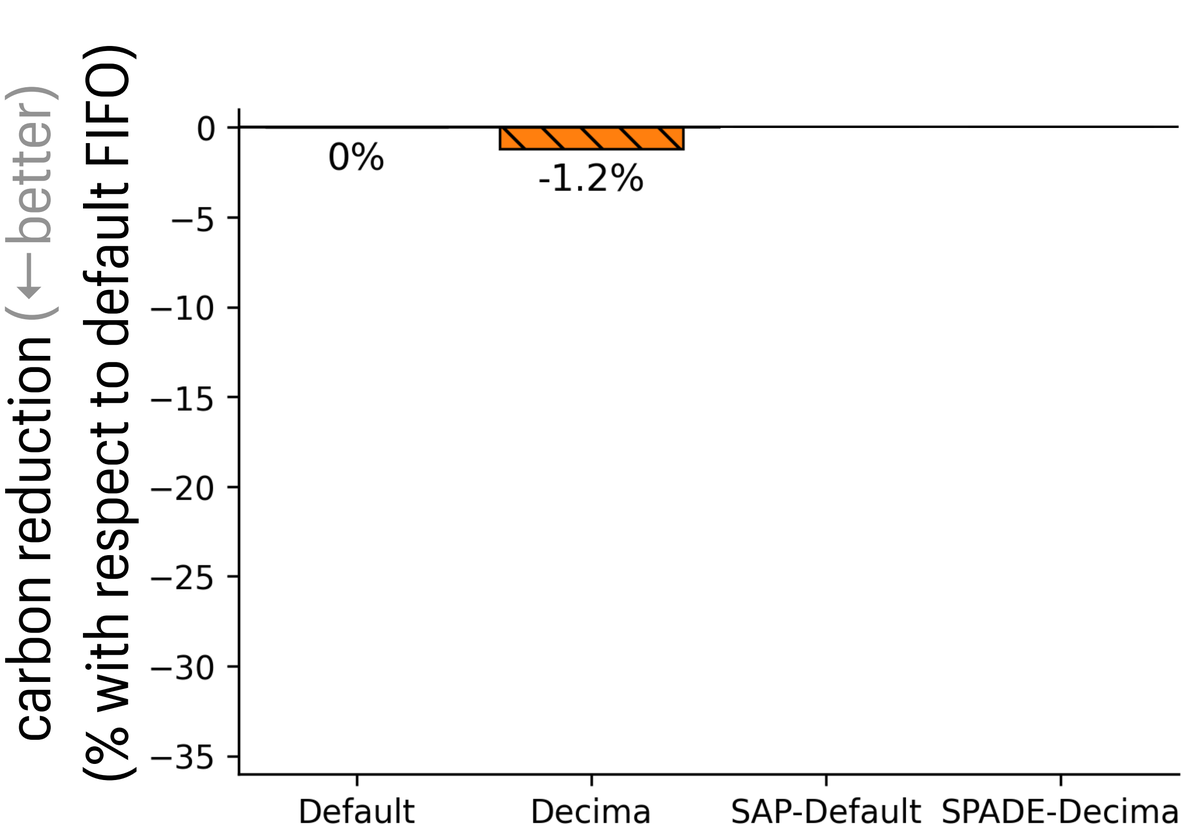
- Results averaged across all prototype experiments
  (TPC-H and Alibaba jobs, signal traces from six different electric grids)

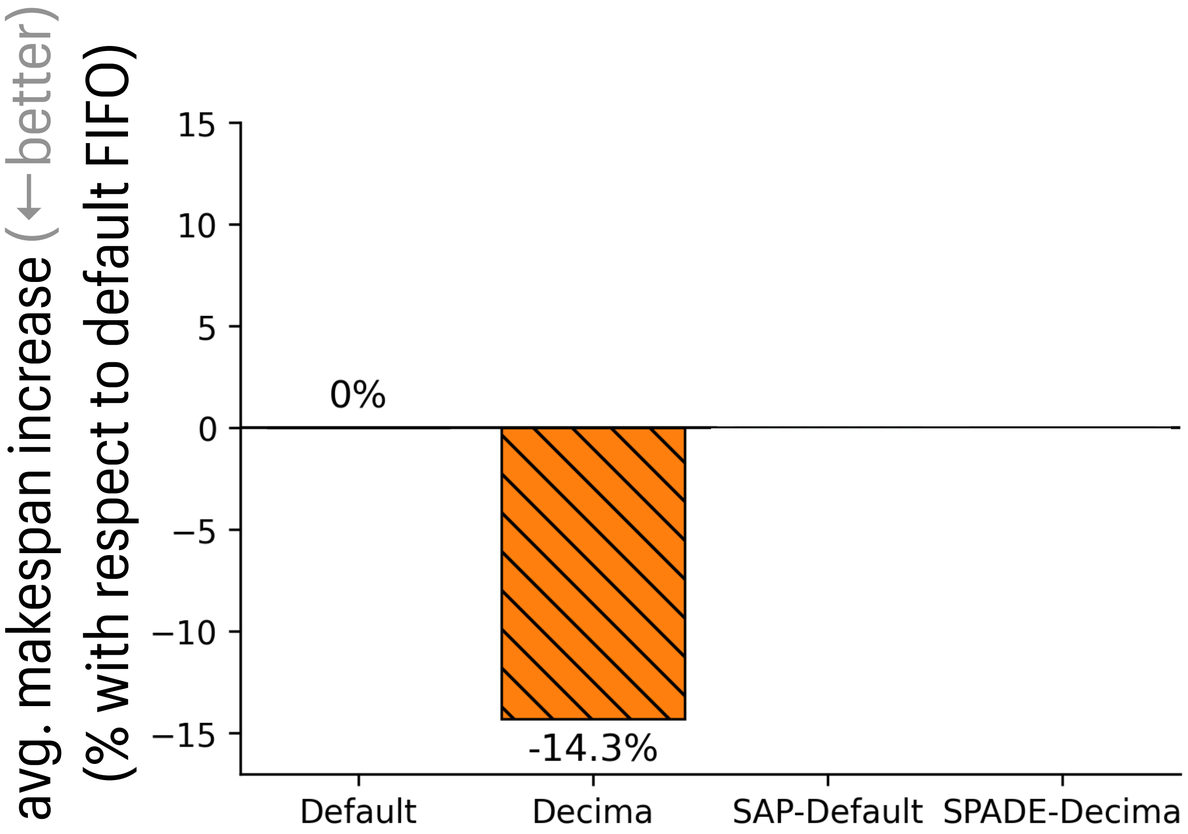




Evaluation — top-line results

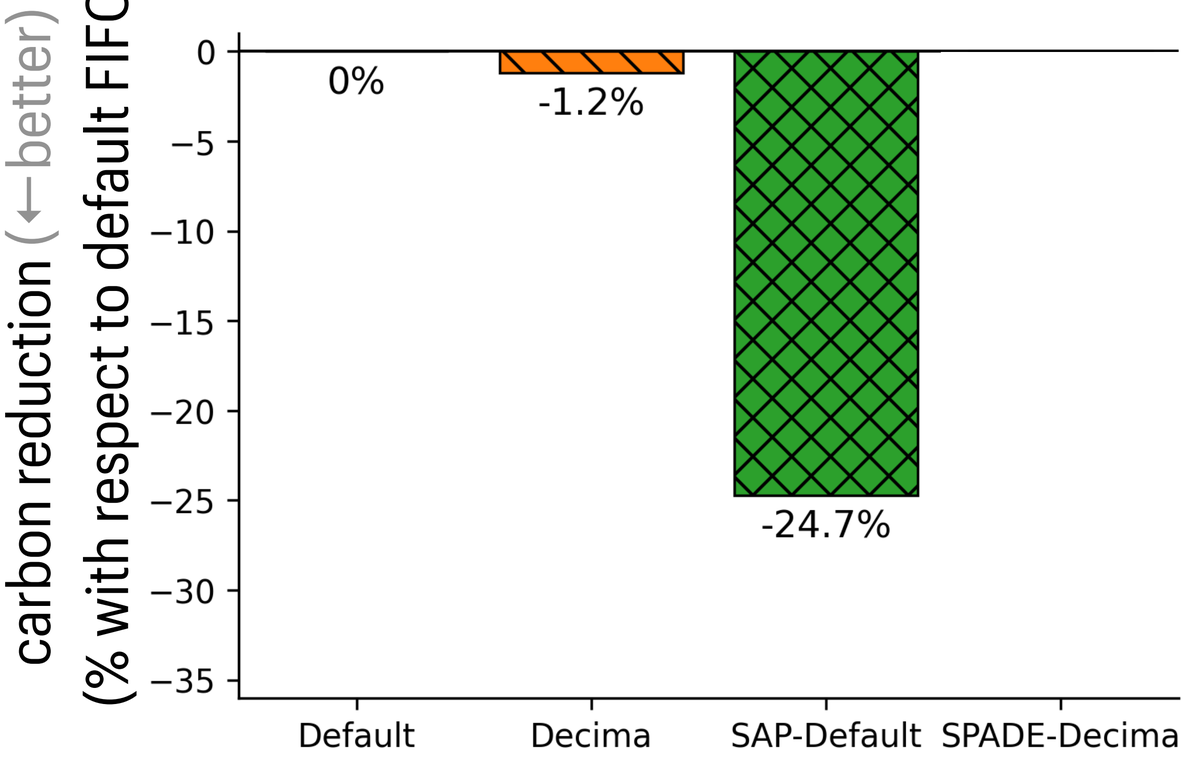
- Results averaged across all prototype experiments
  (TPC-H and Alibaba jobs, signal traces from six different electric grids)

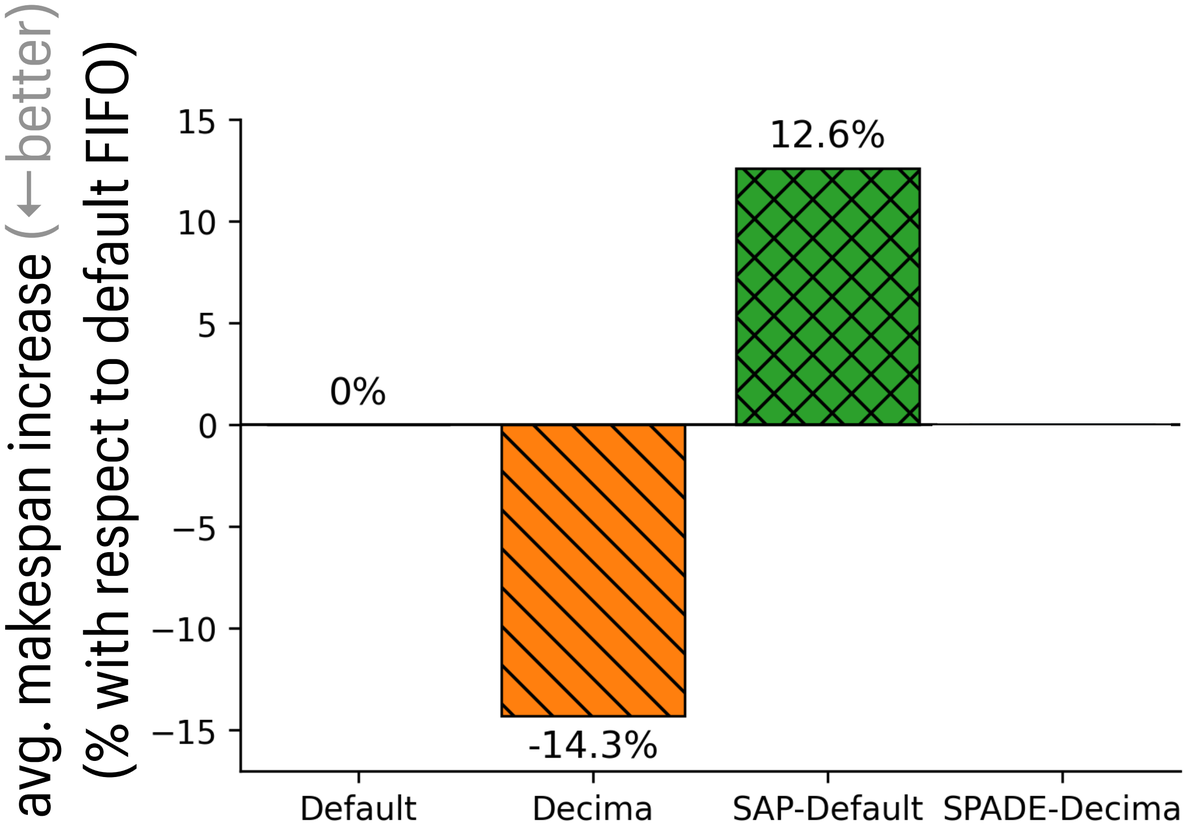




Evaluation — top-line results

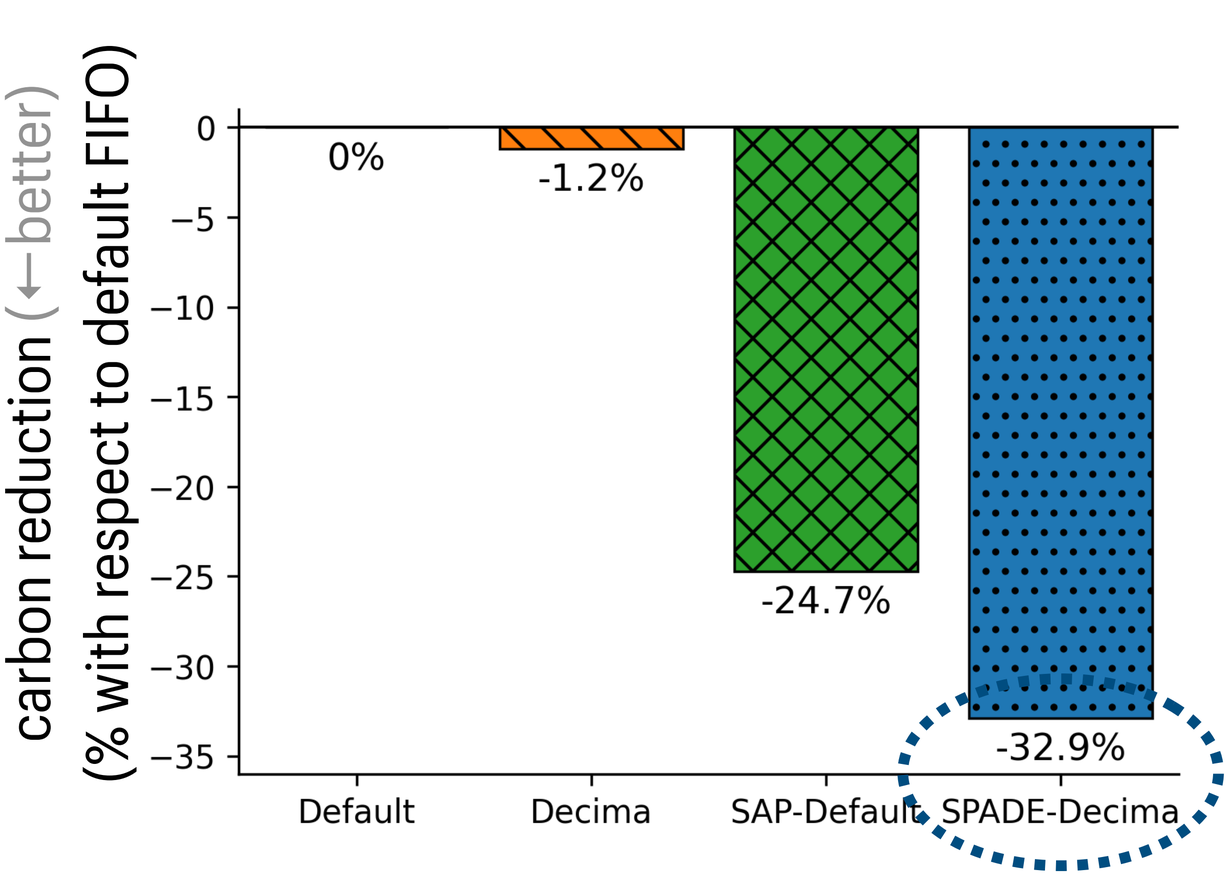
- Results averaged across all prototype experiments
  (TPC-H and Alibaba jobs, signal traces from six different electric grids)

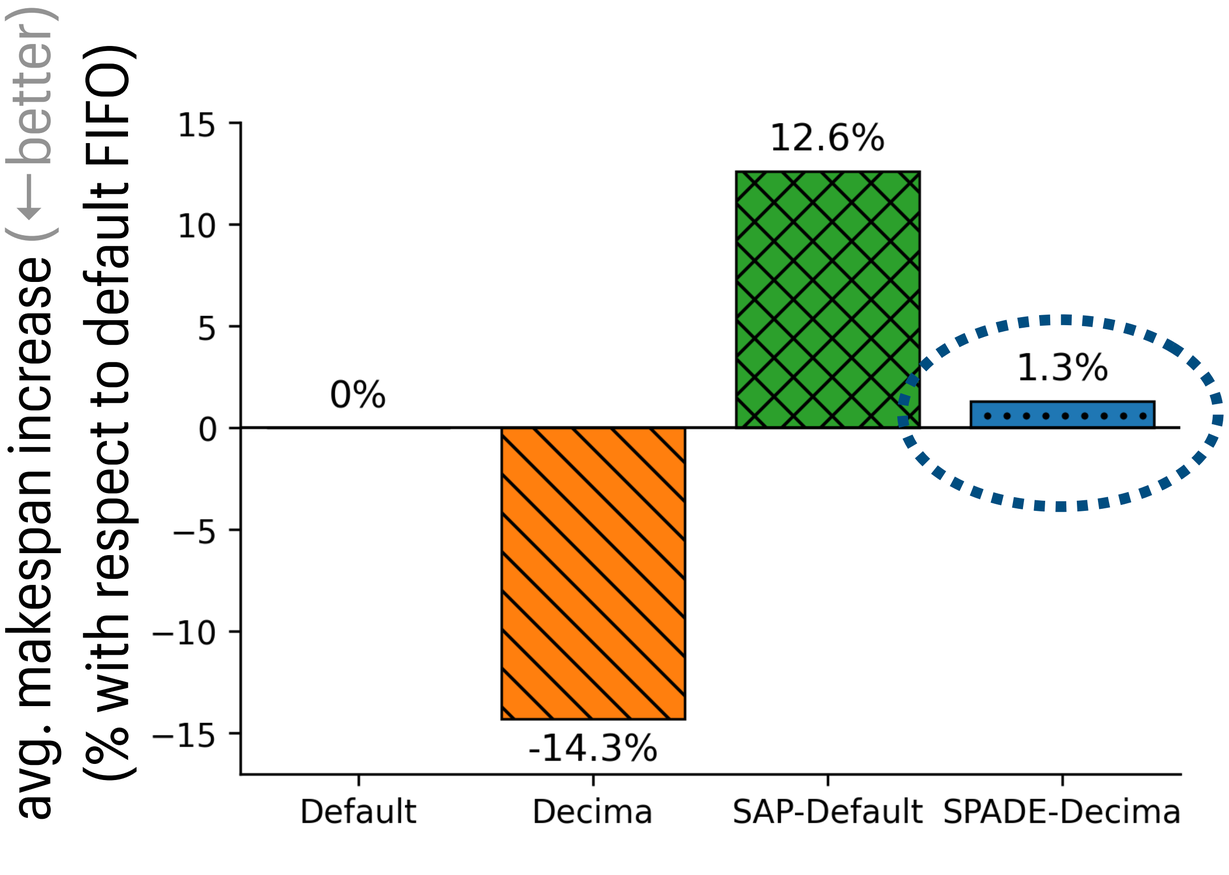




Evaluation — top-line results

- Results averaged across all prototype experiments
  (TPC-H and Alibaba jobs, signal traces from six different electric grids)



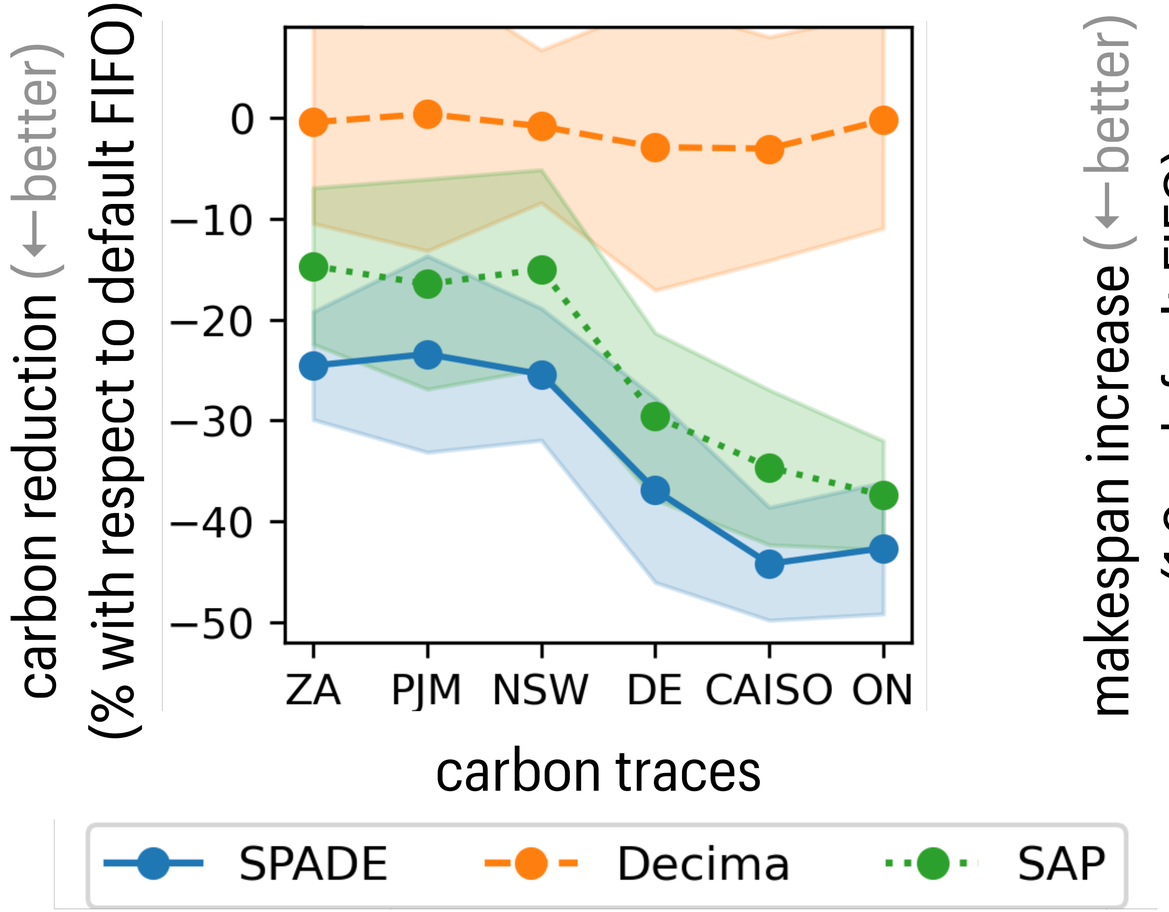


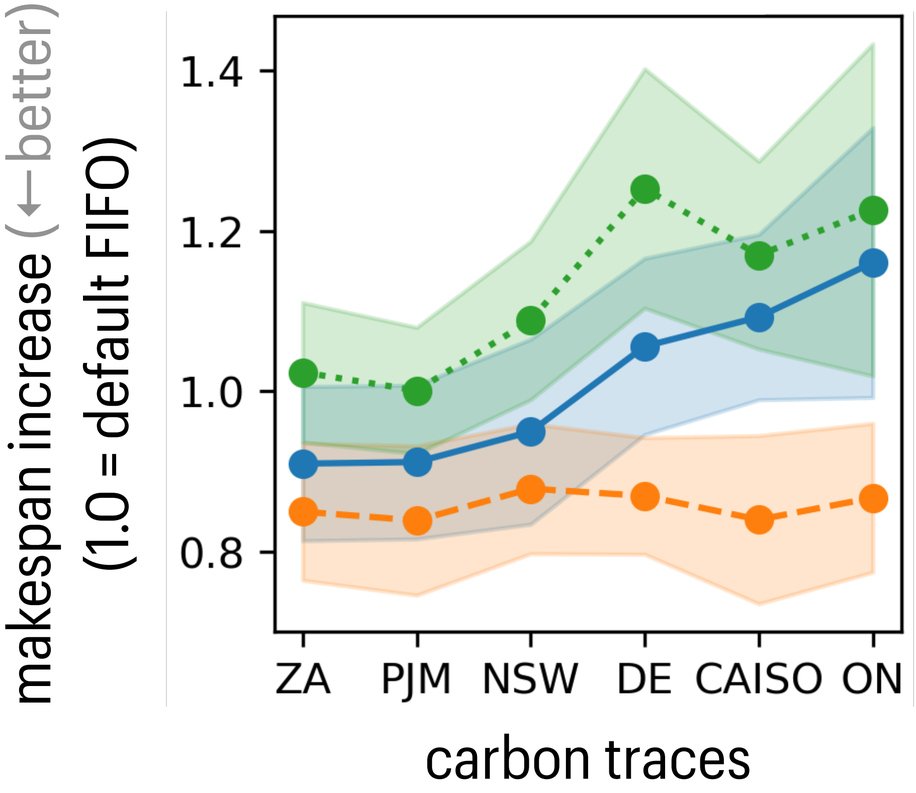
Evaluation — signal characteristics are key

- Signal behavior corresponds to system behavior: a consistent signal implies consistent performance, and vice versa for a highly variable signal.

Evaluation — signal characteristics are key

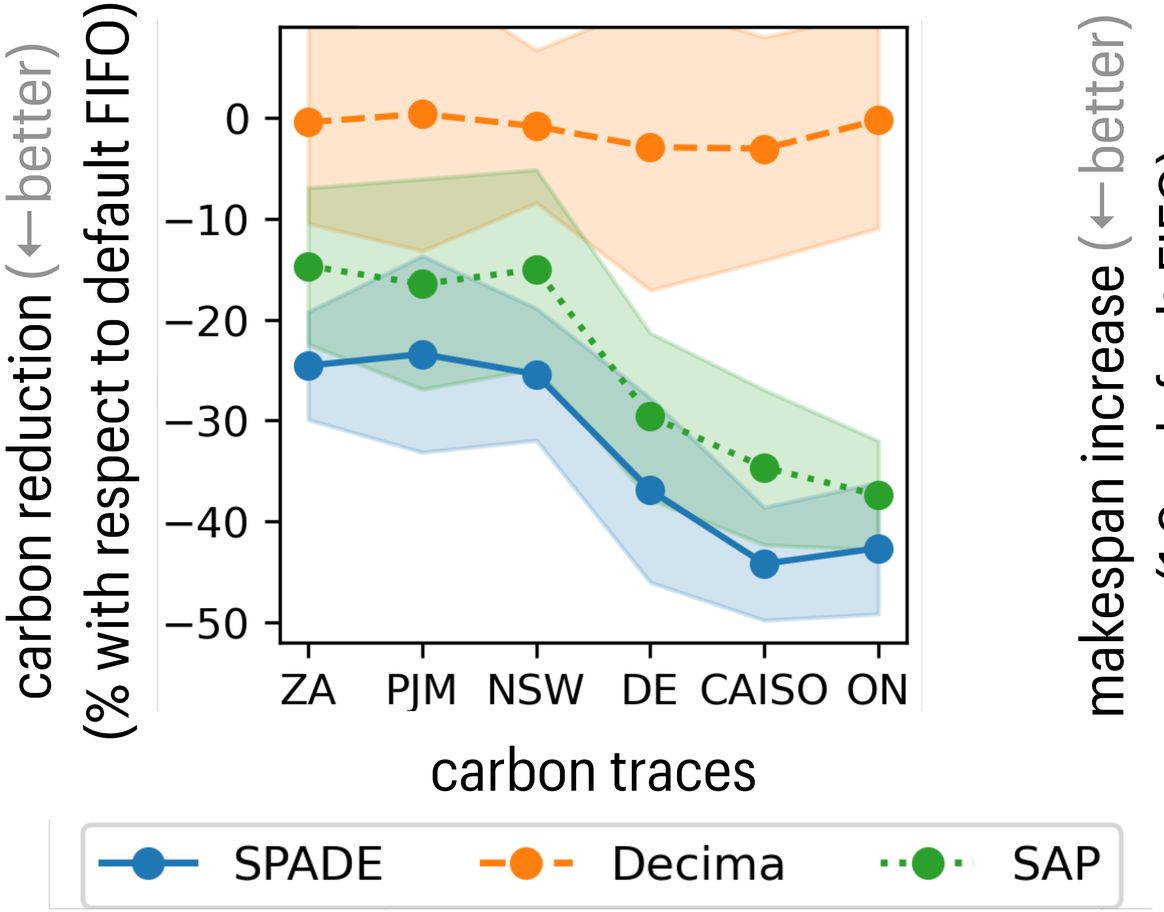
- Signal behavior corresponds to system behavior: a consistent signal implies consistent performance, and vice versa for a highly variable signal.

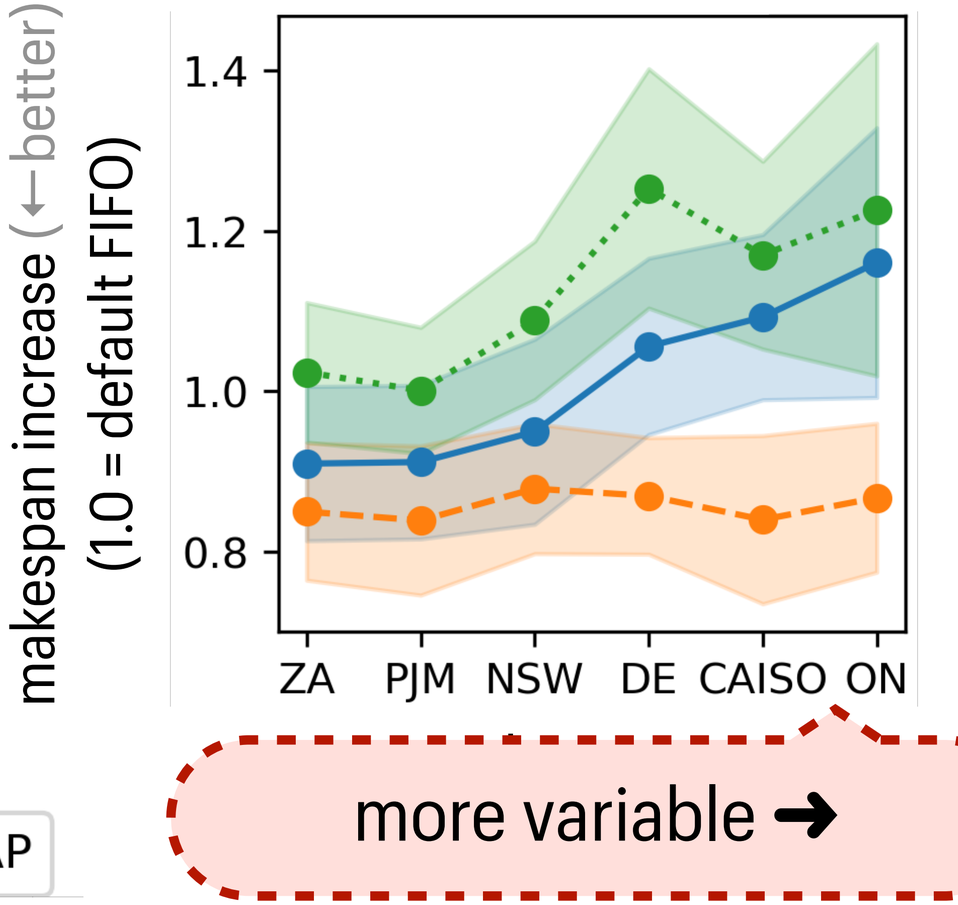




Evaluation — signal characteristics are key

- Signal behavior corresponds to system behavior: a consistent signal implies consistent performance, and vice versa for a highly variable signal.





Summary

- **Signal-awareness and flexibility** gaining importance: data centers are constrained by externalities, and renewables are winning on cost.

Summary

- Signal-awareness and flexibility gaining importance: data centers are constrained by externalities, and renewables are winning on cost.

- We propose **SPADE**, a system for **signal-aware execution of workloads with precedence constraints** (e.g., DAG-structured jobs), which pose a unique challenge / opportunity for compute flexibility.

Summary

- Signal-awareness and flexibility gaining importance: data centers are constrained by externalities, and renewables are winning on cost.
- We propose SPADE, a system for signal-aware execution of workloads with precedence constraints (e.g., DAG-structured jobs), which pose a unique challenge / opportunity for compute flexibility.
- Through our evaluation, we show joint scheduling and provisioning achieves a **superior trade-off between perf. and secondary objectives**.

Summary

- Signal-awareness and flexibility gaining importance: data centers are constrained by externalities, and renewables are winning on cost.
- We propose SPADE, a system for signal-aware execution of workloads with precedence constraints (e.g., DAG-structured jobs), which pose a unique challenge / opportunity for compute flexibility.
- Through our evaluation, we show joint scheduling and provisioning achieves a superior trade-off between perf. and secondary objectives.
- **Future:** other *soft & hard* signals, differently structured workloads; *how to best control?*

Thank you!

SPADE



I will be on the academic job market for postdoc and tenure-track faculty positions this fall.

DOE CSGF

info— alechowicz@umass.edu, adamlechowicz.github.io

code— GitHub: umass-solar/adaptive-dag

NSF