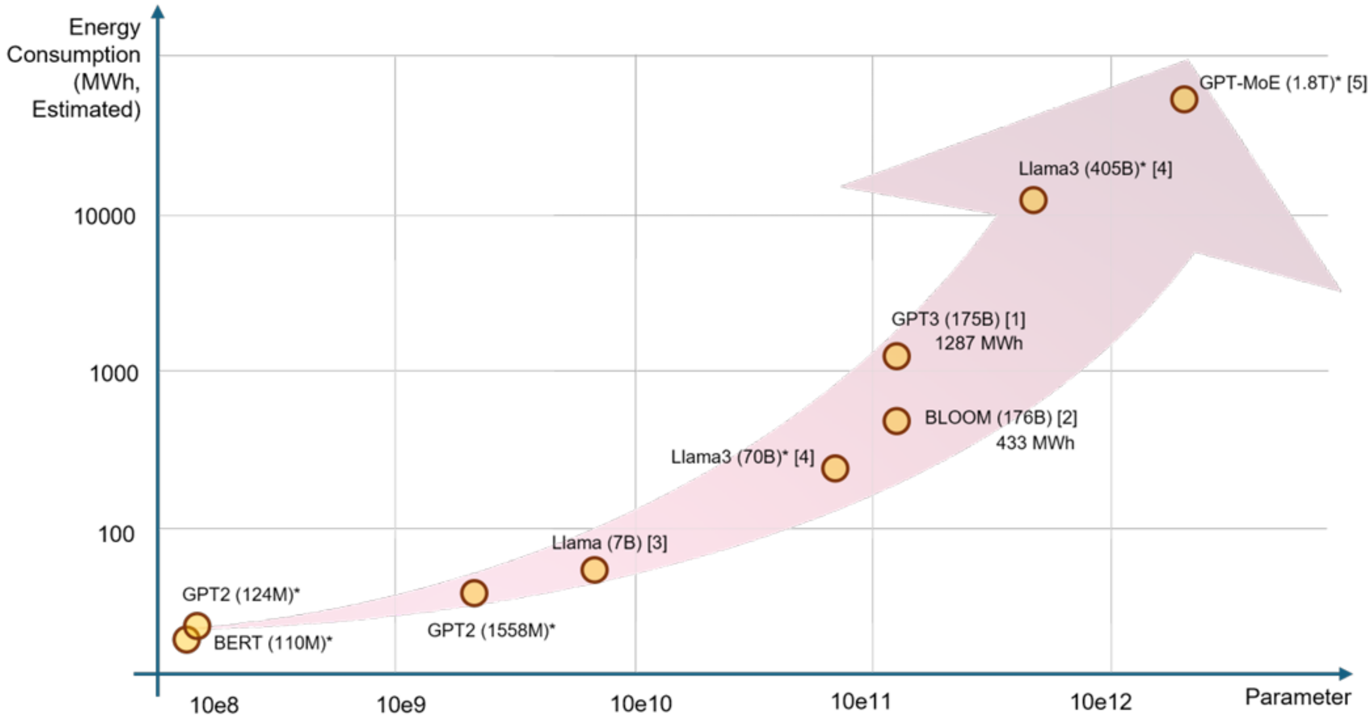
Uncertainty-aware Day-ahead Datacenter Workload Planning with Load-following Small Modular Reactors

Yijie Yang[1], Dan Wang[1], Jian Shi[2], Chenye Wu[3], Zhu Han[2]

[1] The Hong Kong Polytechnic University, [2] University of Houston,
[3] The Chinese University of Hong Kong-Shenzhen

AI Growth Drives Datacenter Energy Demand

- The increase in AI applications leads to increasing energy demands in AI datacenters.
- Training a GPT-3 model needs 1287 MWh.



(Source: [2409.11416] The Unseen AI Disruptions for Power Grids: LLM-Induced Transients)

24/7 Carbon-Free Energy

- 24/7 Carbon-Free Energy:
  - Goal: Matching the consumption with carbon-free energy *every hour*, every day
  - Measurement: Requires **real-time (hourly)** matching

- Net-Zero:
  - Goal: Balance all greenhouse gas emissions (Scope 1, 2, & 3)
  - Measurement: Focuses on **annual** accounting

Nuclear Power as a Solution

- Nuclear energy is carbon-free, and safe
  - Death per Terawatt-hour, Nuclear 0.07 vs. Lignite 32.72 (accidents and air pollution)

- Big tech companies are pursuing nuclear power solutions.
  - Google: Partnered with Kairos Power to launch by 2030.
  - Microsoft: Restart the Three Mile Island energy plant.
  - Amazon: Purchased a nuclear-powered datacenter facility in Pennsylvania.
  - Meta: Seeking proposals for 4GW of new nuclear capacity, early 2030s.

[1]. Why nuclear power is safer than ever. https://www.gisreportsonline.com/r/nuclear-energy-safe/
[2]. Data center owners turn to nuclear as potential electricity source https://www.eia.gov/todayinenergy/detail.php?id=63304#

Load-following Small Modular Reactors (LF-SMRs)

- Load-following Small Modular Reactors (LF-SMRs):
  - 20-300 MW
  - Capable to adjust its power output dynamically in response to real-time demands

- Standard Nuclear Power Plants:
  - > 1000 MW
  - 1+ square mile and massive infrastructure (e.g., cooling towers)
  - Greater human intervention

LF-SMR is very suitable for co-location with datacenters

[1]. Going Nuclear: A Guide to SMRs and Nuclear-Powered Data Centers: https://www.datacenterknowledge.com/energy-power-supply/going-nuclear-a-guide-to-smrs-and-nuclear-powered-data-centers
[2]. Do SMRs and Microreactors embody a nuclear renaissance? https://www.aquaswitch.co.uk/blog/smrs-and-microreactors/

Related work

- SMRs as energy sources
  - SMR for electricity market [1]
  - SMR for electricity-hydrogen integrated systems [2]
  - SMR for electricity-heating multi-energy systems [3]
  - SMR for datacenters [4]

SMRs are more cost-effective than stardard NPPs for DCs in minimizing capital and operations costs

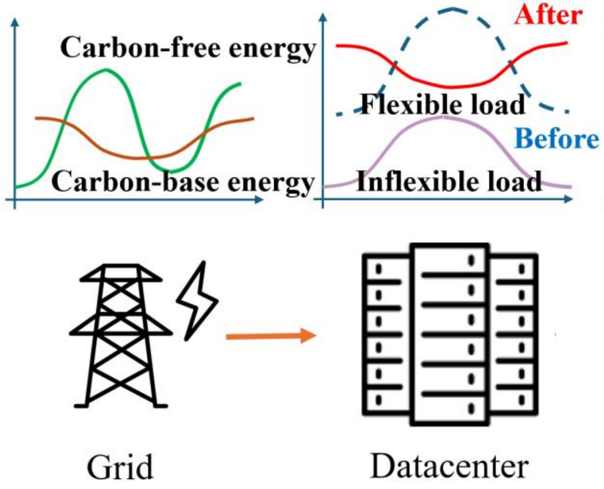
- Emphasis on investment level, i.e., SMRs can increase the revenue of energy systems and minimize energy generation costs

There is a lack of study on operation level for collocated datacenter & LF-SMR

[1]. Jubeyer Rahma, *et al.*, "Optimization of nuclear-renewable hybrid energy system operation in forward electricity market," IEEE GreenTech 2021
[2]. Jubeyer Rahman, *et al.*, "Multi-timescale power system operations for electrolytic hydrogen generation in integrated nuclear-renewable energy systems," Applied Energy 2025
[3]. Jubeyer Rahman *et al.*, "Multi-timescale operations of nuclear-renewable hybrid energy systems for reserve and thermal product provision," Journal of Renewable and Sustainable Energy 2023
[4]. Gabriel Jose Soto Gonzalez, *et al.*, "Powering Data Centers with Clean Energy: A Techno-Economic Case Study of Nuclear and Renewable Energy Dependability," Technical Report. Idaho National 2023

Carbon-aware datacenter operations

- Day-ahead datacenter operations
  - Step 1: Day-ahead renewable energy forecasting (through PPAs)
  - Step 2: Day-ahead workload planning: shifting the flexible workloads (e.g., AI training)

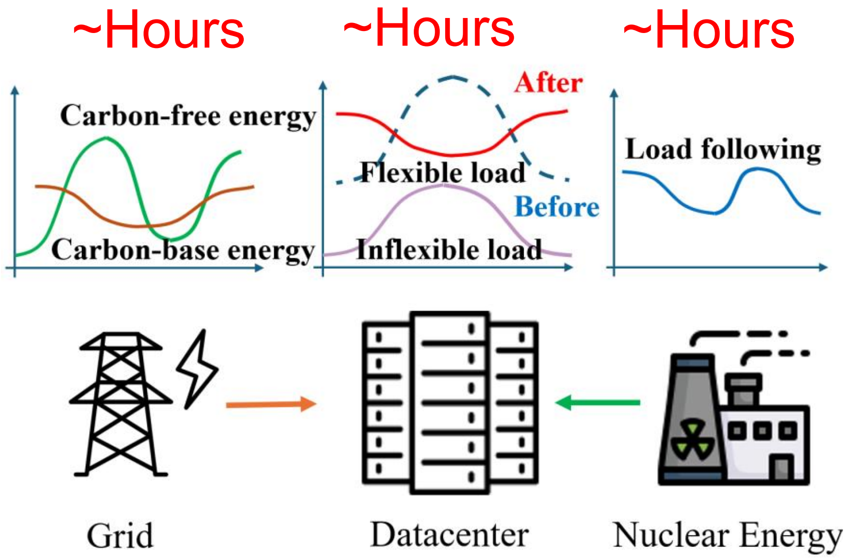


[1]. Bostandoost, Roozbeh, *et al.*, "Data-driven Algorithm Selection for Carbon-Aware Scheduling," ACM SIGENERGY Energy Informatics Review 2024
[2]. Diandian Gu, *et al.*, "GreenFlow:A Carbon-Efficient Scheduler for Deep Learning Workloads," IEEE Transactionson Parallel and Distributed Systems 2024
[3]. Yejia Liu, *et al.*, "Geographical Server Relocation: Opportunities and Challenges," ACM SIGENERGY Energy Informatics Review 2024
[4]. Ana Radovanović, *et al.*, "Carbon-aware computing for datacenters," IEEE Transactions on Power Systems 2022

Carbon-aware datacenter operations

- Day-ahead datacenter operations
  - Step 1: Day-ahead renewable energy forecasting (through PPAs)
  - Step 2: Day-ahead workload planning: shifting the flexible workloads (e.g., AI training)

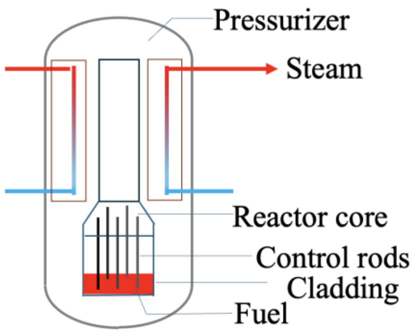
- Day-ahead workload planning with LM-SMRs: to adapt to the renewable energy dynamics, and co-planning the datacenter workload planning and SMR energy generation



[1]. Bostandoost, Roozbeh, *et al.*, "Data-driven Algorithm Selection for Carbon-Aware Scheduling," ACM SIGENERGY Energy Informatics Review 2024
[2]. Diandian Gu, *et al.*, "GreenFlow:A Carbon-Efficient Scheduler for Deep Learning Workloads," IEEE Transactionson Parallel and Distributed Systems 2024
[3]. Yejia Liu, *et al.*, "Geographical Server Relocation: Opportunities and Challenges," ACM SIGENERGY Energy Informatics Review 2024
[4]. Ana Radovanović, *et al.*, "Carbon-aware computing for datacenters," IEEE Transactions on Power Systems 2022

Challenges

- LF-SMR is not an extra energy source immediately dispatchable
  - SMR in a nut shell



- Nuclear fission generates heat, producing steam to drive a turbine to generate electricity.
- Control rods regulate the fission rate and power output.

  - Ramping limit: Rapid changes cause mechanical stress on fuel cladding, leading to damage.
  - Stable power period: After a reactor reduces its power generation, the xenon buildup rate still exceeds decay, causing the Xe-135 to increase for hours before decline to a new equilibrium
  - Excess energy cost: if the generation is too much, unused energy affects grid stability and introduces economic penalties

System Models and Problem Formulation

- The LF-SMR Model

Ramp-down limit and ramp-up limit

$$p_{t-1} - p_t \leqslant \overline{P}_{RD} \times Rd_t - \delta \times Up_t, \quad \forall t,$$

$$p_t - p_{t-1} \leqslant \overline{P}_{RU} \times Up_t - \delta \times Rd_t, \quad \forall t,$$

$$Rd_t + Up_t + St_t = 1, \quad \forall t,$$

$$Rd_t, Up_t, St_t \in \{0, 1\}, \quad \forall t,$$

Minimum stable period hour

$$(Up_t - Up_{t-1}) \times T_h \leqslant \sum\nolimits_{tt=t-T_h}^{t-1} (St_t + Up_t), \quad \forall t,$$

- The Datacenter Power Supply Model

$$w_t = (1 + \epsilon_t) W_t, \quad g_{grid,t} = q_t + w_t.$$

- The Datacenter Power Demand Model

$$z_t = \sum\nolimits_{c \in C} \sum\nolimits_{k \in \mathcal{H}} Y_{k,c,t} \cdot s_{k,c}, \qquad v_t = e(z_t),$$

$$\sum\nolimits_{t \in \mathcal{T}} Y_{k,c,t} \geq 1, \quad \forall k \in \mathcal{H}, c \in C,$$

$$Y_{k,c,t} = 0, \quad \forall k \in \mathcal{H}, c \in C, t \notin \mathbb{Z}_{[k:k+h_c]}.$$

- The Datacenter Power Cost Model

$$EPC = \sum\nolimits_{t=1}^{T} \alpha q_t + p_{co_2} q_t I_t + \kappa w_t,$$ → Energy Purchased Cost

$$NGC = \sum\nolimits_{t=1}^{T} \gamma p_t,$$ → Nuclear Generation Cost

$$EEC = \sum\nolimits_{t=1}^{T} \beta (p_t - d_t),$$ → Excess Energy Cost

Problem Formulation

- Given Datacenter workload profile $s_{c,k}$, SMR operation parameters T_h, $\bar{P}_{RU}$, $\bar{P}_{RD}$, Renewable energy of the grid
- Minimize Total Cost (Energy Purchase Cost + Nuclear Generation Cost + Energy Excess Cost)

$$\min EPC + NGC + EEC$$

$$\text{s.t.} \quad (1) - (11),$$

$$p_t \geqslant d_t, \quad \forall t,$$

$$g_{grid,t} + d_t = v_t, \quad \forall t,$$

$$w_t + d_t \geqslant \eta v_t, \quad \forall t,$$

Methodology

- Uncertainty-aware Day-ahead Datacenter Workload Planning with LF-SMR
  - A robust chance-constraint optimization

- A Two-Stage Approach:
  - Stage 1 (Renewable Energy Prediction): Forecast the availability of the renewable energy and quantify the uncertainty

  - Stage 2 (Co-optimization): Optimize the SMR energy generation under the physics-constraints and datacenter workload scheduling based on the forecast with uncertainty

Stage 1: Uncertainty-aware Renewable Forecast

- Method: Conformal Prediction (CP)
  - What CP Does: Uses historical data to construct prediction intervals [L, U] for future renewable energy availability.

- Given a statistical confidence level ε, find the prediction intervals in history with confidence ε.

$$P(Actual\ w_t \in Predicted\ Interval) \geq \varepsilon$$

- For example, ε = 90%, we carry out renewable energy prediction in history and find the interval that can cover 90% of the predictions, which can be ±25.
- We use this uncertainty set for stage 2 robust optimization problem.

Stage 2: Co-optimization

- Reformulation: Robust Optimization (RO) Problem
  - Use the prediction intervals from Stage 1 to reformulate the problem as a *robust optimization (RO) problem* with a traditional box uncertainty set

- Method: Mixed Integer Linear Programming (MILP)
  - The RO problem can be transformed into a MILP problem, solved by commercial MILP solvers such as CPLEX and Gurobi.

Algorithm Analysis

*Corollary 4.1 (**Green Energy Coverage Guarantee**). If w_t is obtained using conformal prediction to determine the interval $Y_{w_t} \in \hat{C}\left(X_{w_t}\right)$, the green energy coverage constraint Eq. (20) can be guaranteed with the probability of ϵ, namely,*

$$\mathbb{P}\left(Y_{w_t} + d_t \geqslant \eta v_t\right) \geq \epsilon$$

- Explanation: If the datacenter requires a green energy coverage of 80% (η=0.8), and employing a 95% confidence level (ϵ=0.95), our operational strategy (workload and SMR) guarantees: the 80% coverage target is achieved with the probability of 95%.

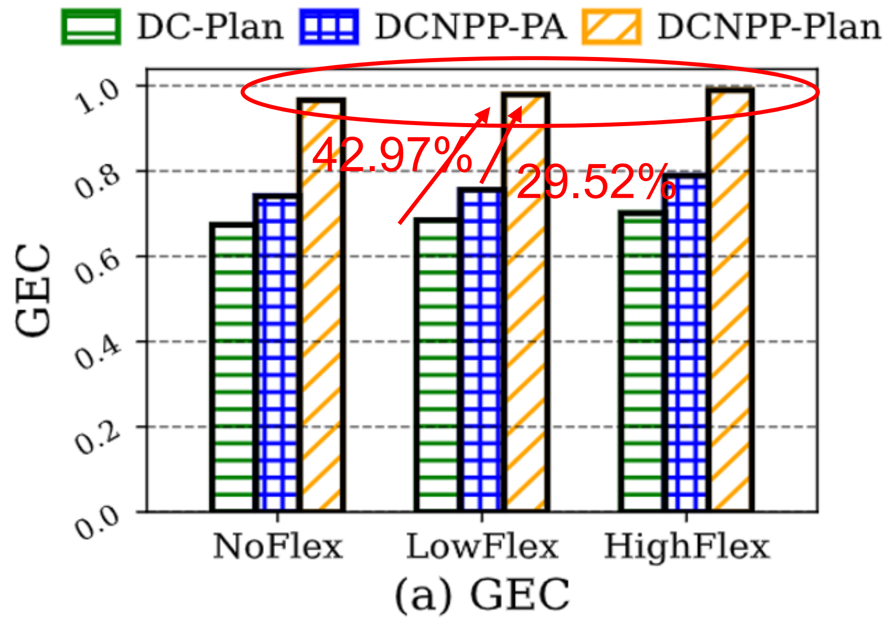
Evaluation Setup

- Datacenter Workloads: Google datacenter trace (May 2019)
  - Peak capacity: 20MW
  - Job types: Inflexible (No delay), Flexible (5h maximum delay)
- SMR:
  - Capacity: 20MW
  - Ramp-rate: 10%, stable power time: 3 hours
- Power Grid:
  - Renewable energy: Wind/solar based on historical weather data
  - Brown energy: All other energy (demand – renewable – nuclear)

Baselines and Metrics:

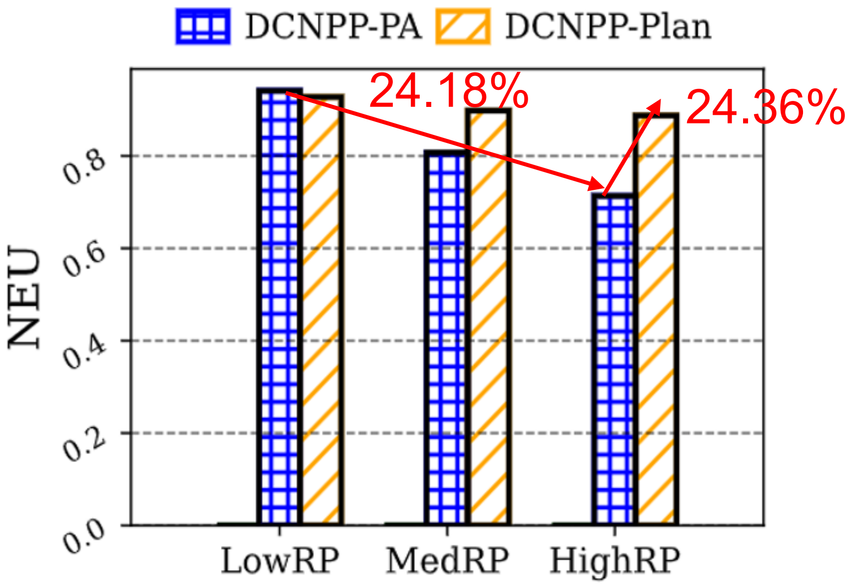
- Baselines:
  - DC-Plan: Grid only (No nuclear energy)
  - DCNPP-PA: SMR fixed output (Physics-Agnostic, no load-following)
  - DCNPP-Plan: DC-SMR with deterministic renewable forecast (No uncertainty)
- Metrics:
  - Green Energy Coverage (GEC): proportion of carbon-free energy (renewable + nuclear) in the total datacenter energy consumption
  - Nuclear Energy Utilization (NEU): proportion of nuclear energy consumed to total nuclear energy produced
- Scenarios:
  - Varying Workload Flexibility: (NoFlex 0%, LowFlex 10%, HighFlex 30%)
  - Varying Renewable Penetration levels: (LowRP: 14 MW solar capacity, MedRP: 28 MW solar capacity, and HighRP: 42 MW solar capacity)

Results



- DCSMR-Plan can achieve almost 100% GEC
- DCSMR-Plan outperforms DC-Plan (no nuclear) in GEC 42.97% and DCNPP-PA (fixed SMR) by 29.52%

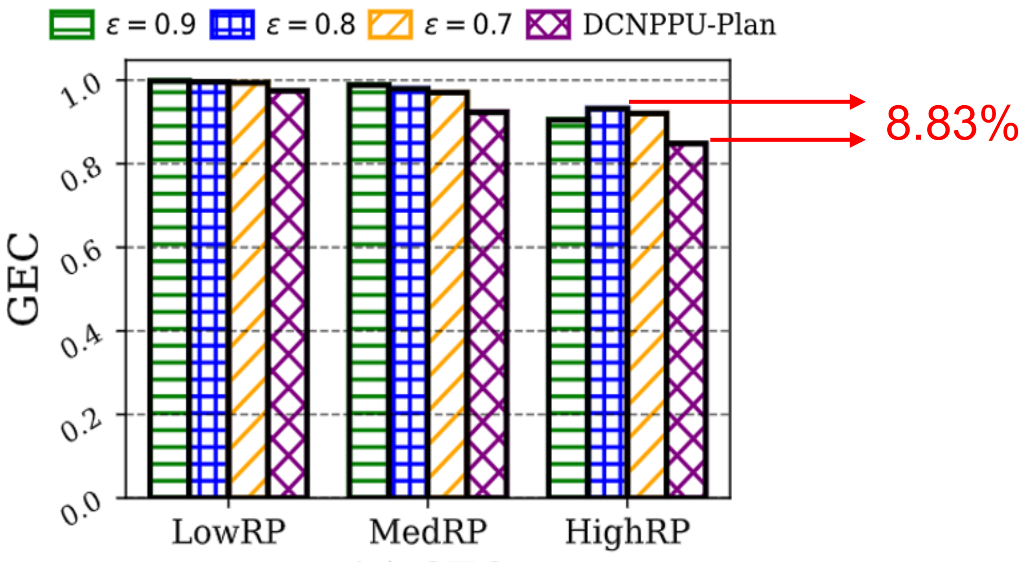
Results: Impact of Renewable Penetration



(b) NEU

- With more renewable (solar, wind), the NEU (nuclear utilization) of DCNPP-PA (fixed SMR) decreases
  - DCNPP-PA has no model of the physical constraints & no load following
- We maintains high NEU in HighRP with 24.36% greater NEU than DCNPP-PA

Results: Impact of Uncertainty Handling



(a) GEC

- Handling uncertainty is important when there is High Renewable Penetration
- We (ε=0.8) outperform DCNPP-Plan (without uncertainty handling) by 8.83% under HighRP

Conclusion

- We studied co-located datacenters and SMR at the operation level.
- We formulated a problem of uncertainty-aware day-ahead datacenter workload planning with LF-SMR; where we explicitly incorporated the physics and operational constraints of LF-SMRs.
- We employed predict and optimize approach where we use conformal prediction to handle grid renewable uncertainty.

Future Work

- One SMR can support multiple datacenters and one datacenter can use multiple SMR supplies
  - The mathematical tools will be different, Game theory.

- Multi-output LF-SMRs (co-generation of thermal energy, hydrogen)
  - Absorption chiller, a type of chiller that uses heat to produce chilled water.

Thank you!

Q&A