

NEW MEXICO

**ECAM**

ENERGY CONSERVATION  
AND MANAGEMENT

# Data Centers

## Large Load Energy Use

WNRC – Roswell

July 13, 2026



# EMNRD's Role in the Data Center Landscape

- EMNRD helps develop energy policy for the state (e.g., CETS, Grid Modernization Plan, Energy Security Plan)
- EMNRD does not permit energy assets for data center generation
- Forestry Division may issue incidental take permits for endangered plant species
- EMNRD may intervene in PRC dockets to support zero-carbon goals, grid planning, and the Governor's agenda broadly

# Large Loads: Cooling without water

- Closed-loop liquid cooling
- Immersion cooling – immersing chips and boards in liquid
- Free cooling -- using ambient outside air during winter months or in cooler climates.
- "Dry" or air cooled, using heat pumps, refrigeration, or chillers)

.. many low- or no-water techniques require more electricity

# What does data center energy use look like?

“Let’s say you’re running a marathon as a charity runner and organizing a fundraiser to support your cause. You ask an AI model 15 questions about the best way to fundraise..”

“Then you make 10 attempts at an image for your flyer before you get one you are happy with, and three attempts at a five-second video to post on Instagram.

“You’d use about 2.9 kilowatt-hours of electricity—enough to ride over 100 miles on an e-bike (or around 10 miles in the average electric vehicle) or run the microwave for over three and a half hours.”

# How much electricity does one **text** query consume?

“The smallest model in our Llama cohort, Llama 3.1 8B, has 8 billion parameters—essentially the adjustable “knobs” in an AI model that allow it to make predictions. When tested on a variety of different text-generating prompts, like making a travel itinerary for Istanbul or explaining quantum computing, the model required about 57 joules per response, or an estimated 114 joules when accounting for cooling, other computations, and other demands. This is tiny—about what it takes to ride six feet on an e-bike, or run a microwave for one-tenth of a second.”

“The largest of our text-generation cohort, Llama 3.1 405B, has 50 times more parameters. More parameters generally means better answers but more energy required for each response. On average, this model needed 3,353 joules, or an estimated 6,706 joules total, for each response. That’s enough to carry a person about 400 feet on an e-bike or run the microwave for eight seconds.”

# How much electricity does one text query consume?

“So model size is a huge predictor of energy demand. One reason is that once a model gets to a certain size, it has to be run on more chips, each of which adds to the energy required. The largest model we tested has 405 billion parameters, but others, such as DeepSeek, have gone much further, with over 600 billion parameters. The parameter counts for closed-source models are not publicly disclosed and can only be estimated. GPT-4 is estimated to have over 1 trillion parameters.

“But in all these cases, the prompt itself was a huge factor too. Simple prompts, like a request to tell a few jokes, frequently used nine times less energy than more complicated prompts to write creative stories or recipe ideas.”

# How much electricity does one **image** consume?

“AI models that generate images and videos work with a different architecture, called diffusion. Rather than predicting and generating words, they learn how to transform an image of noise into, let’s say, a photo of an elephant. They do this by learning the contours and patterns of pictures in their training data and storing this information across millions or billions of parameters. Video-generator models learn how to do this across the dimension of time as well.

“The energy required by a given diffusion model doesn’t depend on your prompt—generating an image of a skier on sand dunes requires the same amount of energy as generating one of an astronaut farming on Mars. The energy requirement instead depends on the size of the model, the image resolution, and the number of “steps” the diffusion process takes (more steps lead to higher quality but need more energy).”

# How much electricity does one **image** consume?

“Generating a standard-quality image (1024 x 1024 pixels) with Stable Diffusion 3 Medium, the leading open-source image generator, with 2 billion parameters, requires about 1,141 joules of GPU energy. With diffusion models, unlike large language models, there are no estimates of how much GPUs are responsible for the total energy required, but experts suggested we stick with the “doubling” approach we’ve used thus far because the differences are likely subtle.

“That means an estimated 2,282 joules total. Improving the image quality by doubling the number diffusion steps to 50 just about doubles the energy required, to about 4,402 joules. That’s equivalent to about 250 feet on an e-bike, or around five and a half seconds running a microwave. That’s still less than the largest text model.”

# How much electricity does one **video** consume?

“An older version of the [Code Carbon], released in August, made videos at just eight frames per second at a grainy resolution—more like a GIF than a video. Each one required about 109,000 joules to produce. But three months later the company launched a larger, higher-quality model that produces five-second videos at 16 frames per second (this frame rate still isn't high definition; it's the one used in Hollywood's silent era until the late 1920s).

“The new model uses more than 30 times more energy on each 5-second video: about 3.4 million joules, more than 700 times the energy required to generate a high-quality image. This is equivalent to riding 38 miles on an e-bike, or running a microwave for over an hour.

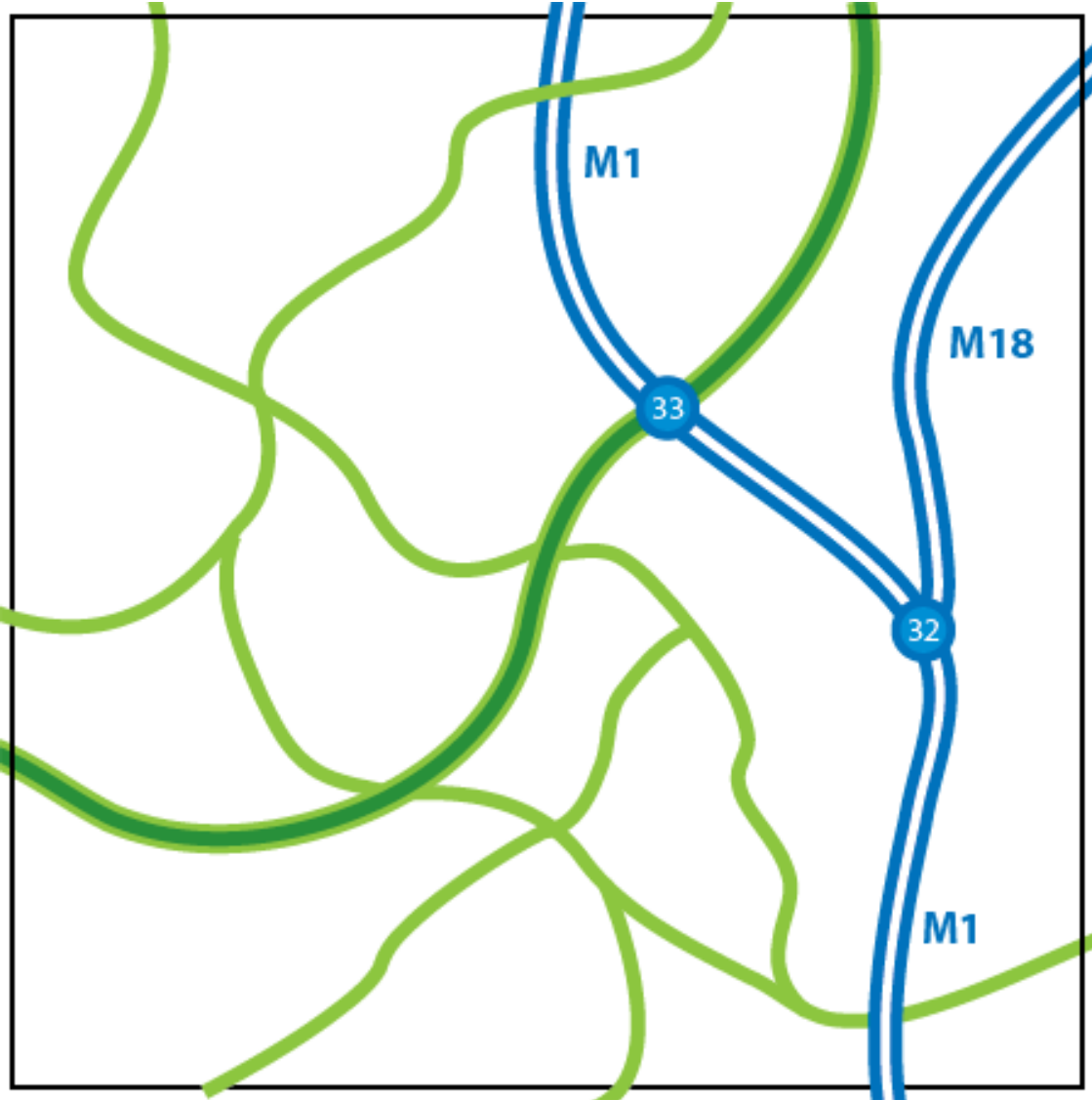
# City-sized electricity usage...



...in ~one~ location on the grid...



...which can be anywhere on the grid, even grid edge...



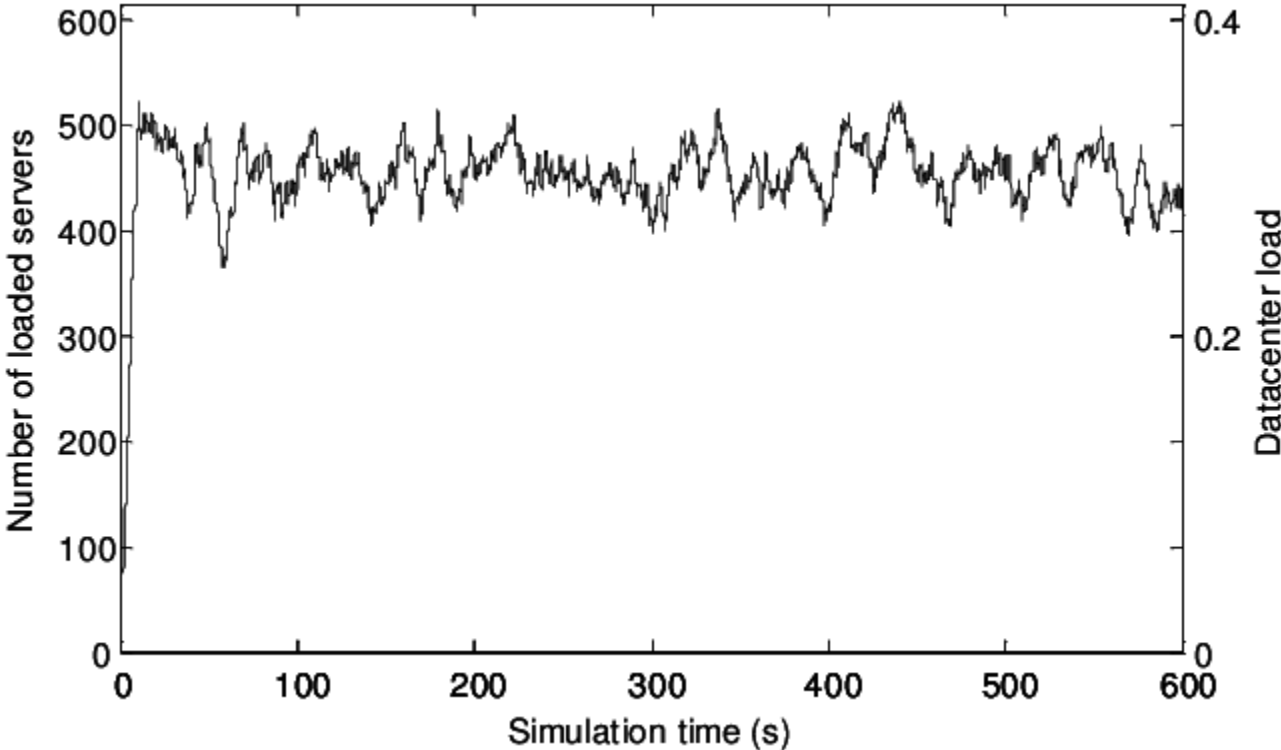
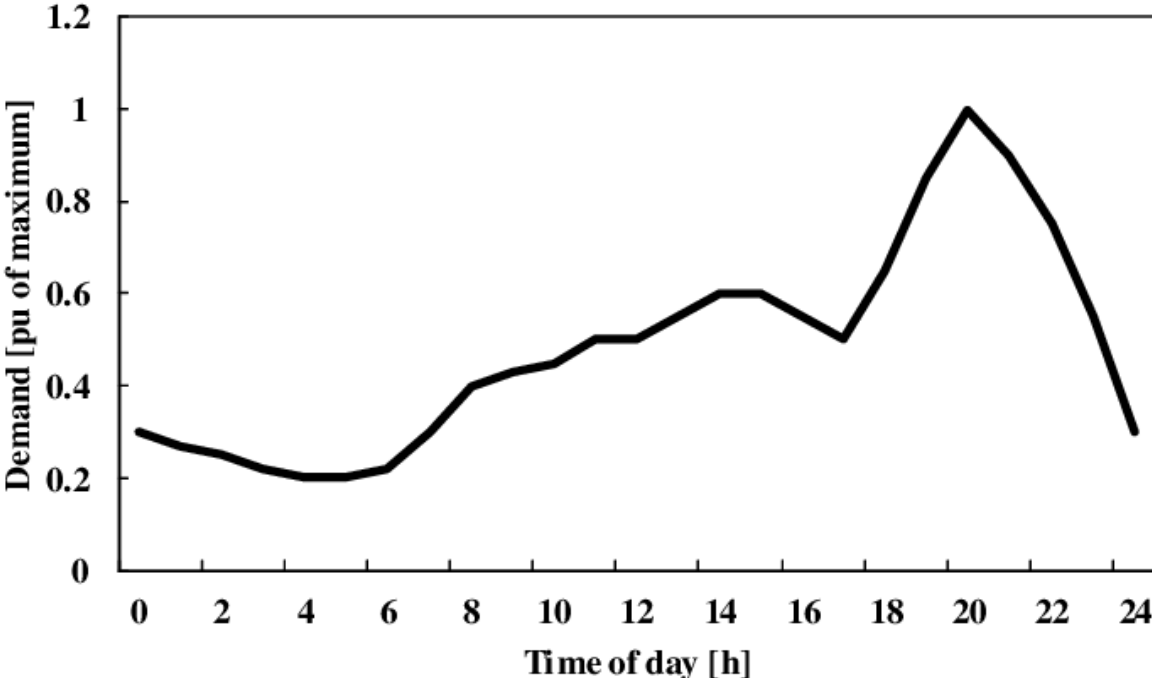
...needed much faster than usual...



VS

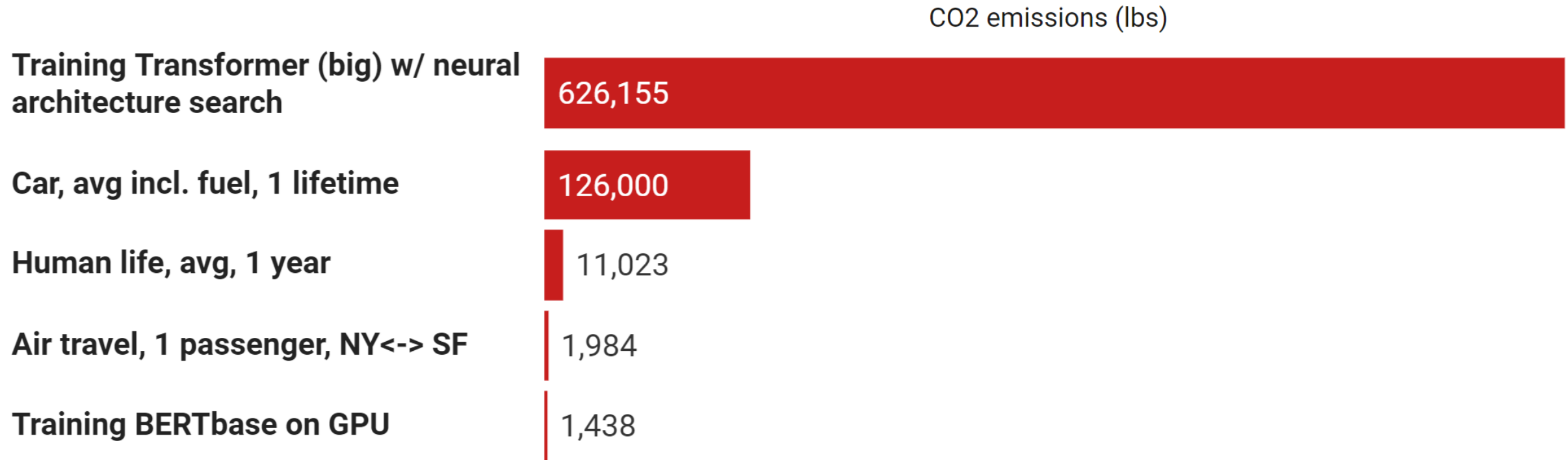


# .... unlike typical load, consumed 24/7



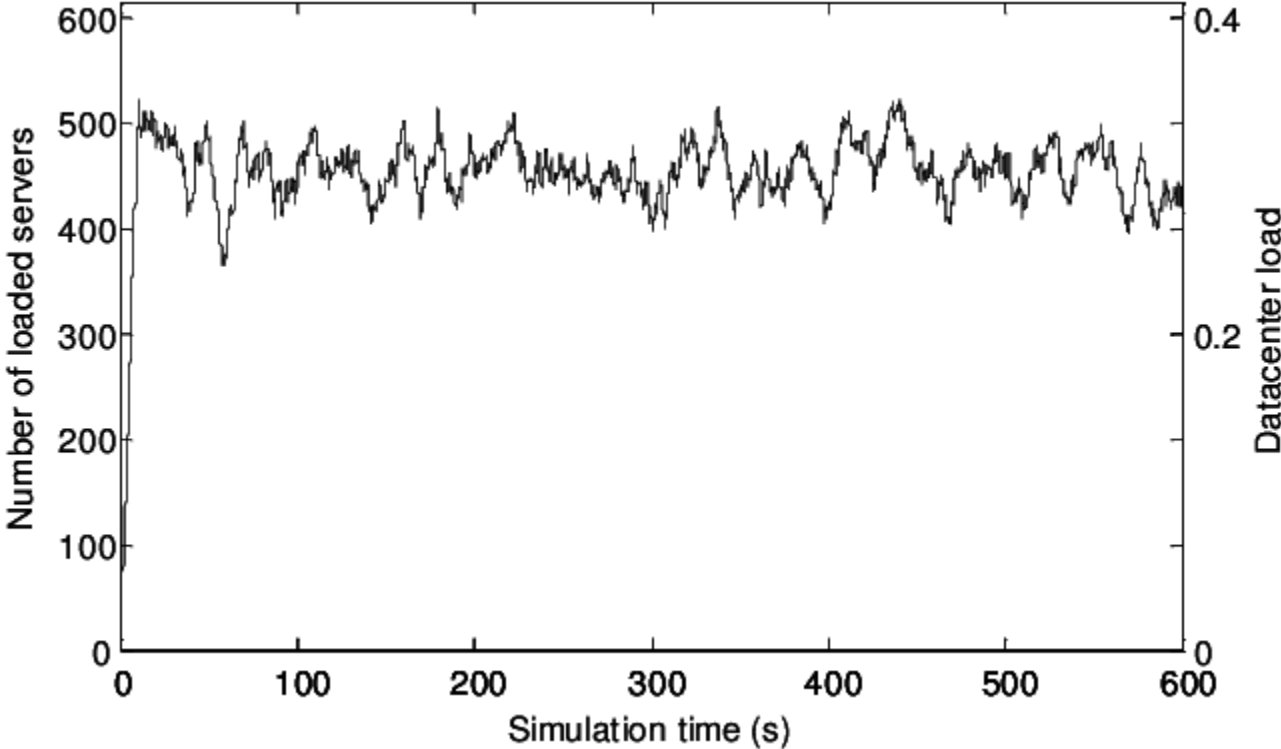
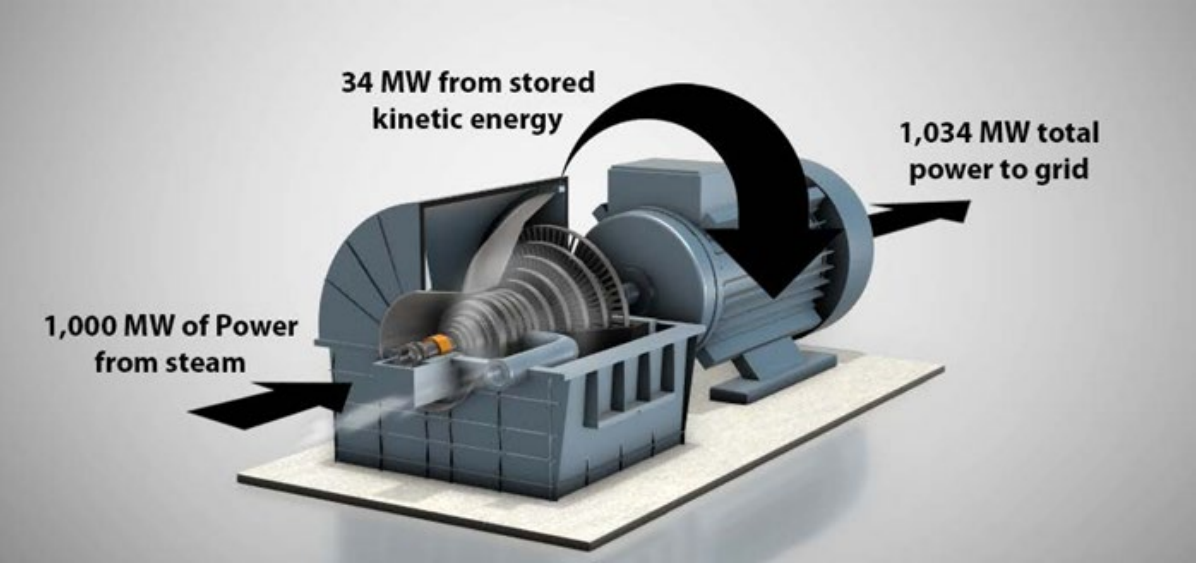
# Carbon footprint comparison

Source: Strubell et al, 2019.

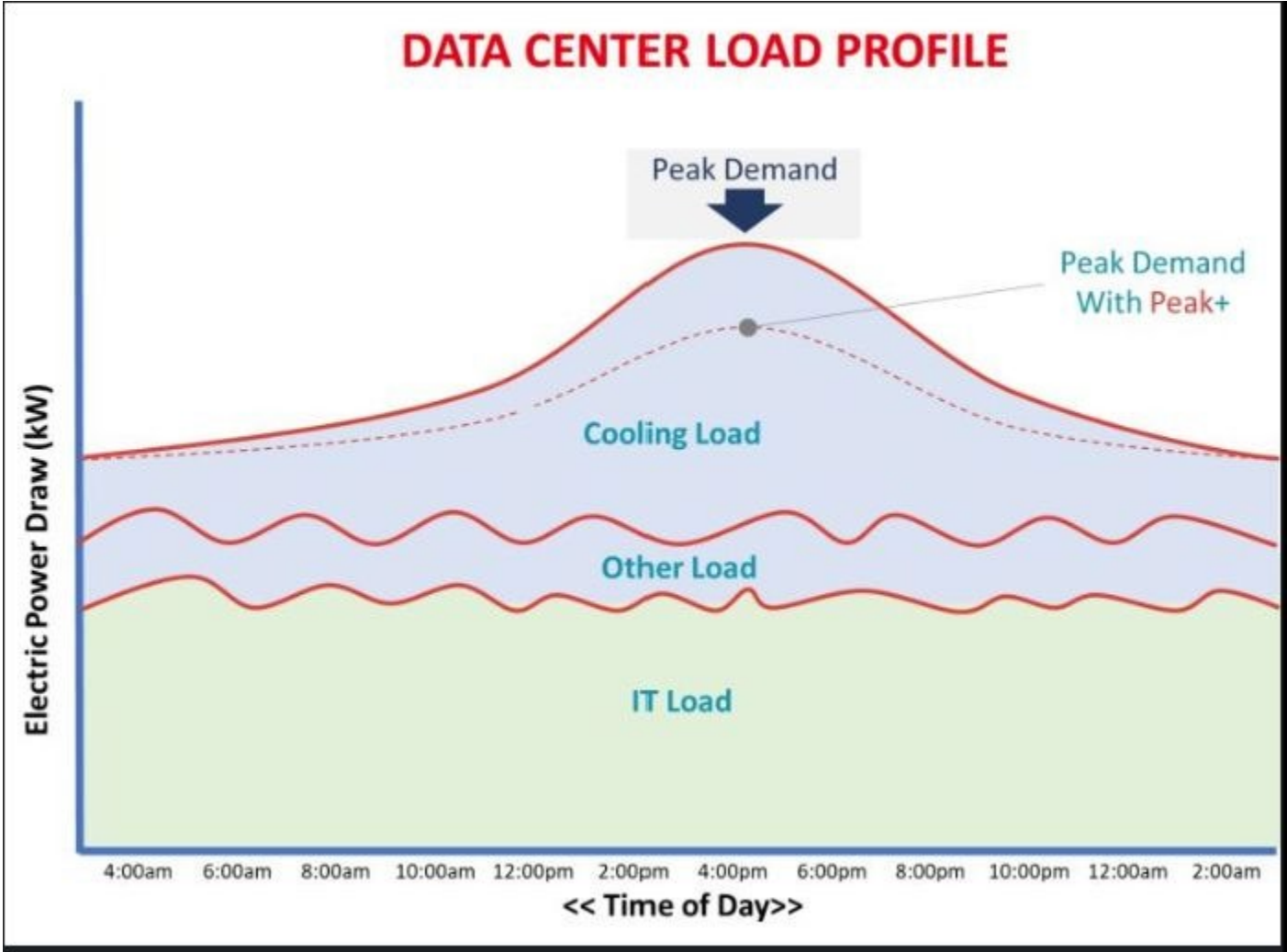


Reconstructed from: <http://arxiv.org/abs/1906.02243>

# ...and “spiky” ...



....plus data centers need electricity for cooling.



# Large Loads: Challenges and Opportunities



# Large Loads: Challenges

- Utility load forecasts have increased dramatically in recent years.
- Traditional utility interconnection and infrastructure-upgrade processes can struggle to deliver power on the accelerated timelines requires by large-load customers.
- Uncertainty around the timing, size and location of large loads can generate significant planning and reliability risks for grid operators, who must then make long-lead infrastructure decisions without predictable load forecasts.
- Significant affordability risks for other ratepayers may arise when utilities must make large capital investments to serve new high-demand loads while still meeting RPS requirements.

# Large Loads: Opportunities

- 1) **When the already-built-and-paid-for grid has excess capacity (which can be often), large-load customers pay for a significant share of that existing capacity, lowering rates for others.**
- 2) **Large-load growth can drive grid modernization when these customers help fund flexible, cost-saving solutions that strengthen the system and lower rates, such as:**
  - 1) Accelerated renewable energy development
  - 2) Accelerated battery or other storage.
  - 3) Utility partnership or self-generation by data centers to reduce grid strain.
  - 4) Contributing to grid modernization tools like digital twins, real-time load balancing, reconductoring, etc.

## **Potential guardrails include:**

- Large load specific tariffs (including flexible large load tariffs) to ensure large customer costs are fairly allocated and not shifted onto or subsidized by other rate classes
- Demand-response programs to shift energy use

# Current Proposals Before the PRC

In their latest rate case before the Public Regulation Commission, El Paso Electric has proposed two new large load tariffs for the PRC to review and approve:

- **Rate No. 50 – High Load Factor Service Rate**
  - Max monthly demand of 30MW, and averaging above 25.5MW
  - 20-year minimum term
  - If the customer triggers grid upgrades, the customer pays.
  - Early termination *may* result in exit fees
- **Rate No. 51 - Large Load Power Service Rate**
  - Max monthly demand of 10MW; no "average" requirement.
  - Same minimum term, facility cost charge and potential penalties for early termination as in rate No. 50
- **For customers to receive these tariffs, they must execute an Energy Service Agreement (ESA) with EPE**
  - This special version of an ESA would include specific language on triggered grid upgrades, financial security (if needed), cost allocation parameters to prevent cost shifting, and other technical requirements as applicable.

# Large Load Tariff Best Practices

## Prevent stranded-asset and repayment risk

- Long-term service commitments (e.g., 20-year minimum term)
- Early-exit fees to prevent cost-shifting if a large load customer leaves.
- Up-front demand charges to secure cost recovery *before* construction.
- Collateral requirements to protect utility and ratepayer finances.

## Prevent cost-shifting for any required grid upgrades

- Ensure large loads pay for the upgrades they trigger so households and small businesses aren't subsidizing them.

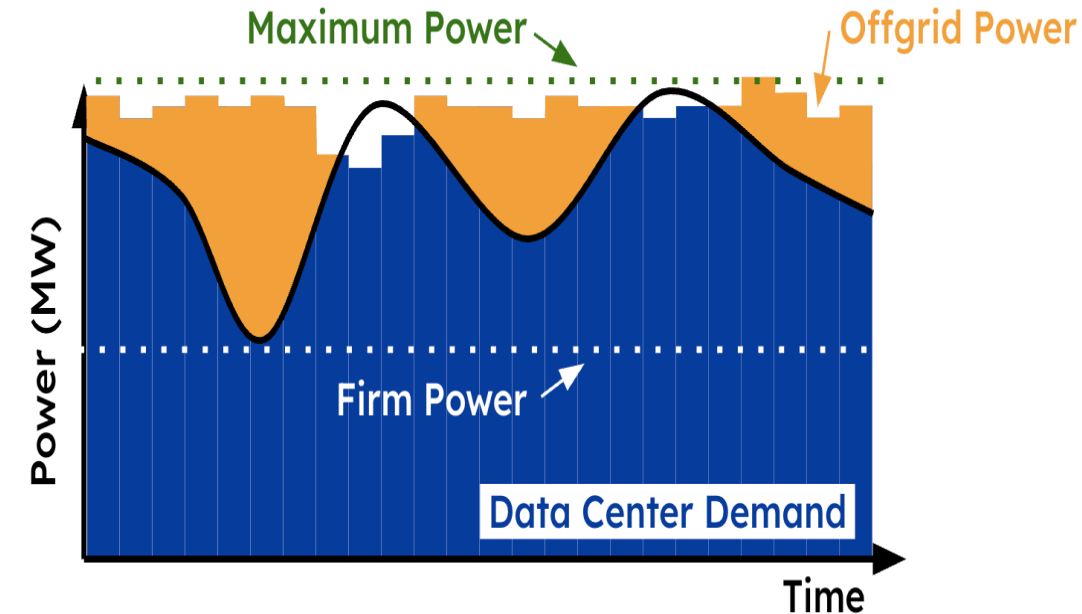
## Ensure large-load customers pay for the costs they trigger, including adjust charges when some customers place greater demands on the grid.

- Establish an *inflexible* large-load rate class and a separate *flexible* large-load rate class, as these two types of customers have vastly different infrastructure upgrade needs.

# Flexibility Solutions

**Some large loads (including data centers) can make use of behind-the-meter resources to avoid triggering grid upgrades**

- Co-locating renewables and storage at data centers can significantly reduce the infrastructure build necessary to interconnect new large customers quickly.
- Duke University found that if all new loads seeking interconnection in the SPP RTO accepted a 0.25% (22 hrs/yr) curtailment rate, 7.7 GWs of additional load could be interconnected without triggering massive upgrades. **Even small amounts of operational flexibility from large customers can dramatically reduce the need for expensive grid upgrades.**





Puck Stair  
Director  
Energy Conservation and Management Division

[Rebecca.Stair@emnrd.nm.gov](mailto:Rebecca.Stair@emnrd.nm.gov)