Decarbonizing Scientific Computing

Dr. Andrew Grimshaw
Lancium Compute
Emeritus Professor of Computer Science, University of Virginia

September 8, 2022

LANCIUM
BALANCING ENERGY

Agenda

- The message
- The renewable transformation of the Texas electricity markets
- The challenges of renewable power
- Addressing the challenges
- Lancium
- OSG/CMS and CVMFS

LANCIUM
BALANCING ENERGY

The Message

- Renewables are the least expensive way to produce power in the world today.
- The pace of renewable deployment is staggering.
  - And gated only by the ability of electrical grids to safely absorb the power
- Controllable loads, particularly loads that can be rapidly reduced, are critical to maintaining grid stability in the presence of non-constant solar and wind power.
- Computation can be used as the controllable load.
- If we, as a community, can adapt a bit, there are gigawatts of very inexpensive, green, electricity available.

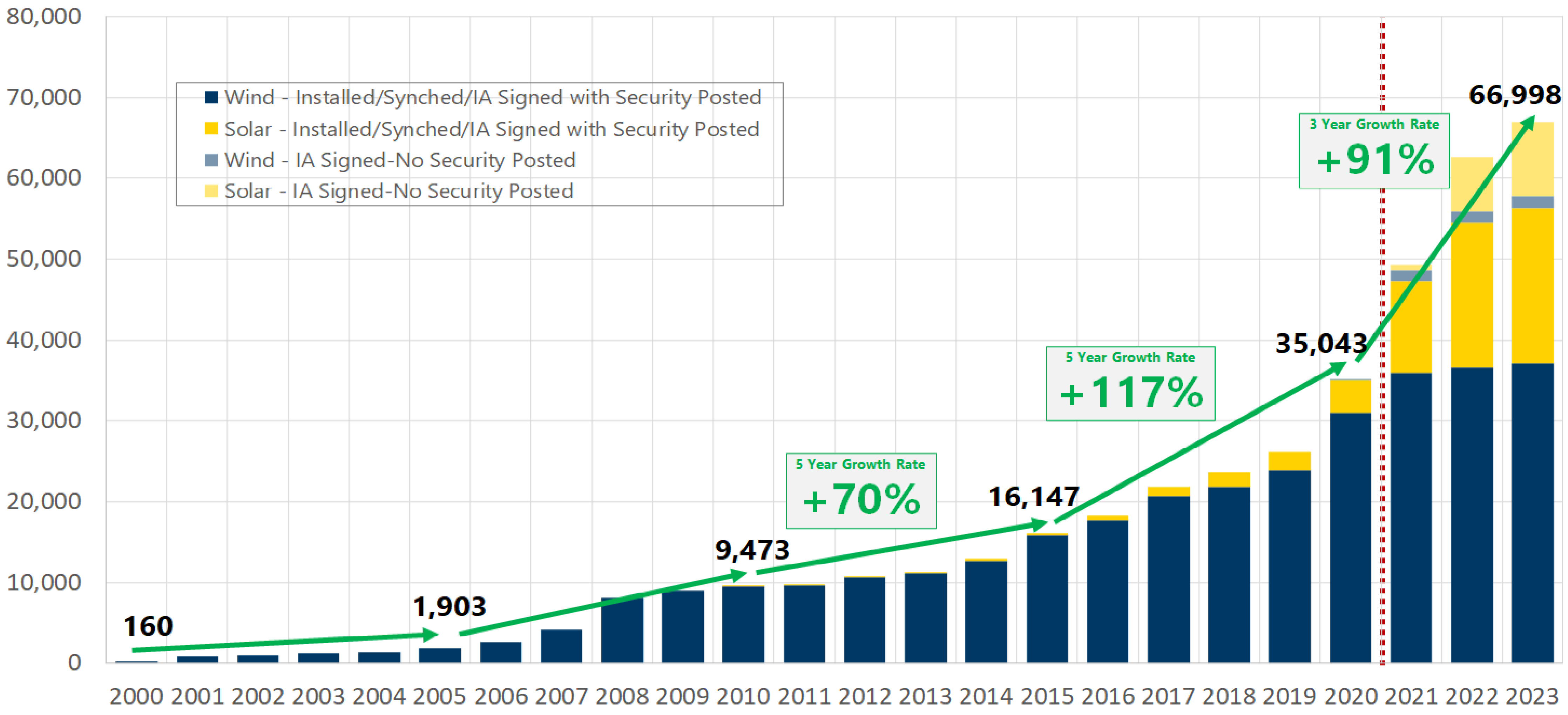
1: “Renewables 2020: Analysis and forecast to 2025”, International Energy Agency
2: Lazard, Levelized Cost of Energy and Levelized Cost of Storage – 2020

LANCIUM
BALANCING ENERGY

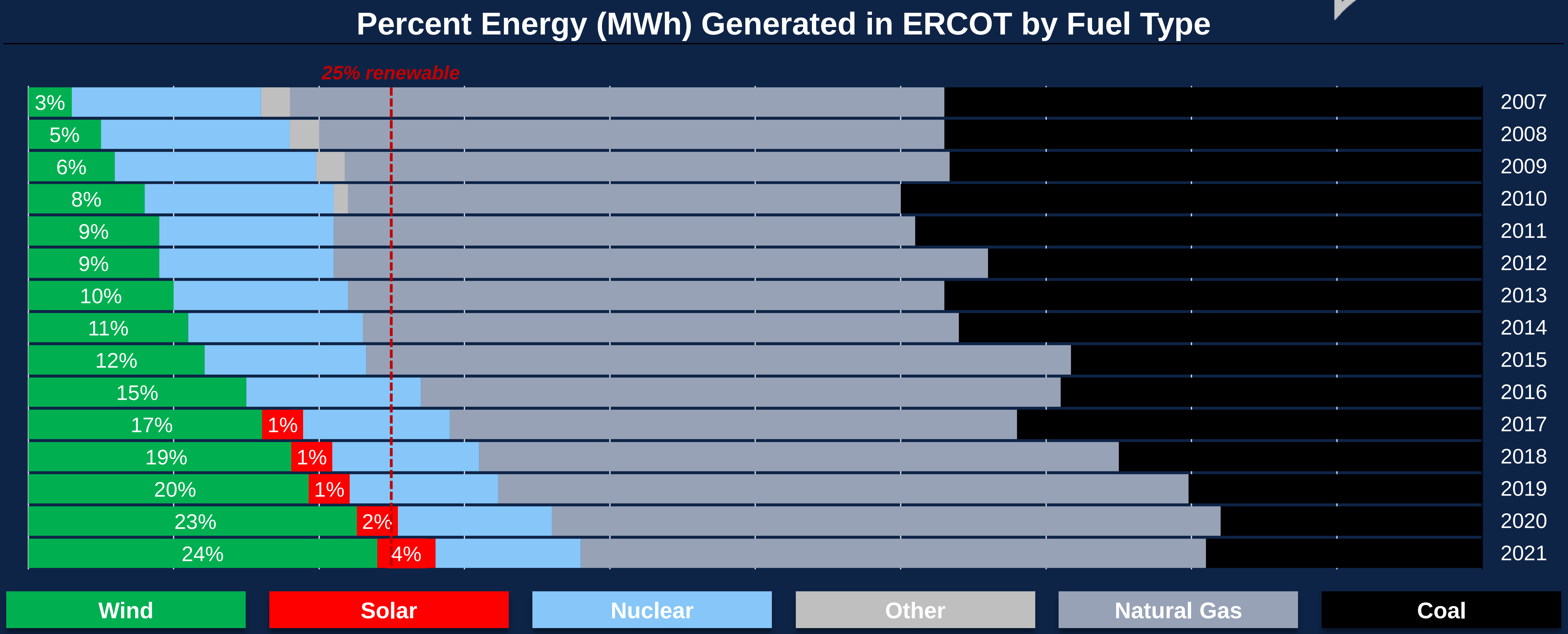
The Renewable Transformation - Texas

LANCIUM
BALANCING ENERGY

ERCOT WIND & SOLAR – ANNUAL ADDITIONS



ERCOT's Rapid Renewable Generation Growth...

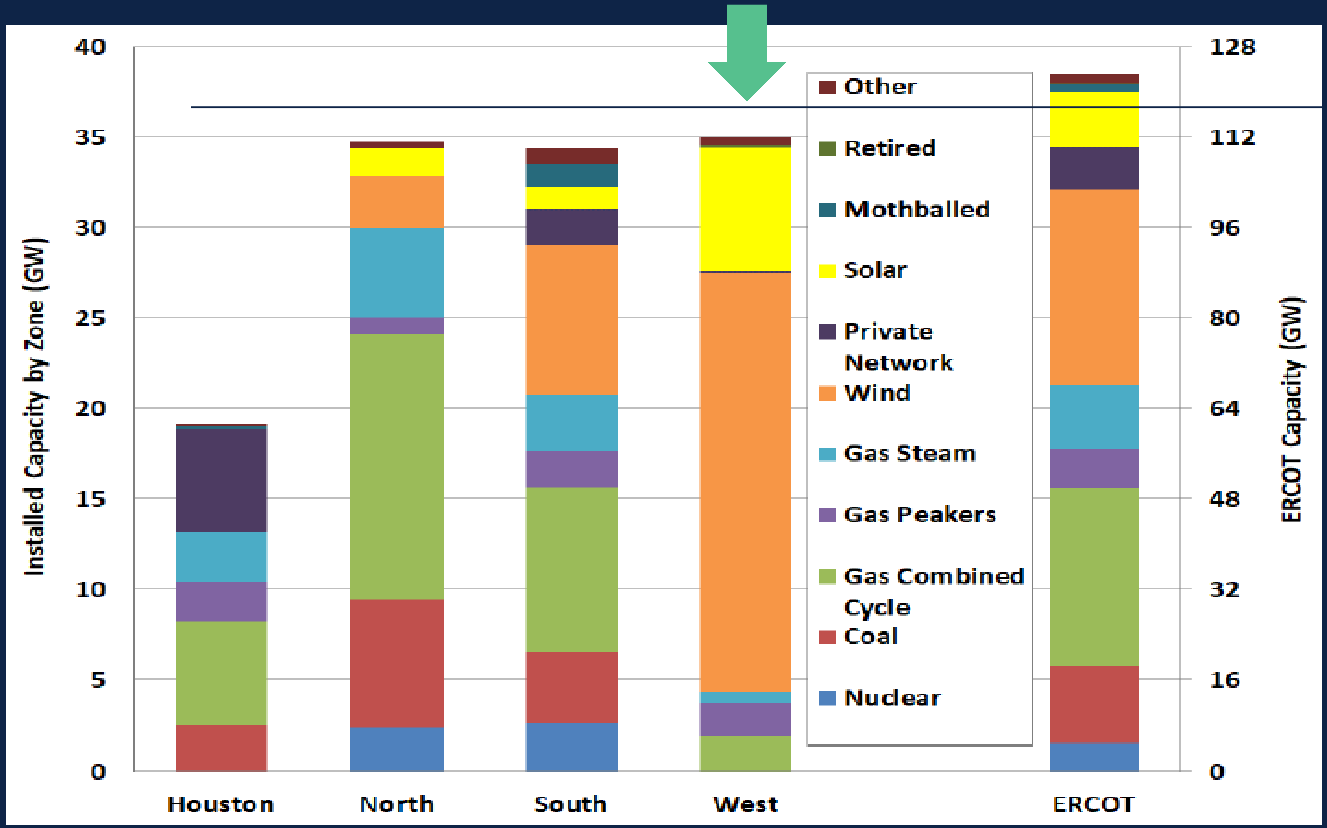


LANCIUM
BALANCING ENERGY

Source: ERCOT - Market Data

West Texas Installed Capacity by Technology

West Texas Supply
Wind: 24GWs
Solar: 7GWs
NatGas: 4GWs



In 2021, West Texas load averaged 6GWs with installed capacity at 35GWs (24GWs Wind, 7GWs Solar, 4GWs NatGas)

LANCIUM
BALANCING ENERGY

Challenges with Wind/Solar

LANCIUM
BALANCING ENERGY

Challenges

- Congestion – the load (people, industry) is far from the generation.
- They are variable and intermittent.
  - Solar is (obviously) diurnal, varies with the seasons, and can change rapidly due to clouds.
  - Wind varies constantly, though has trends:
    - Stronger in winter than summer. Stronger at night.
- As load approaches generation, frequency starts to drift (called an excursion). If load not reduced, or generation increased, generators can be destroyed. The grid operator will not let this happen. They are taken offline to prevent damage. This can lead to blackouts to reduce load.
  - Because of this grid operators (ERCOT) keep a significant amount of reserve capacity that can come on-line very quickly (~15s), quickly (~15 minutes), slowly (~3 hrs).
  - They pay for this .. It is called an “ancillary” service.
  - “Capacity” can be increases of generation OR decrease of load.
- In the absence of massive (and expensive) energy storage systems, or the ability to rapidly shed load, or increase production, the amount of renewable power the grid can absorb is limited without risking instability and blackouts.

Addressing the Challenges

LANCIUM
BALANCING ENERGY

To address congestion and variability we need:

- To move load close to production.
  - There are many industries that consume huge amounts of power: aluminum, steel (arc furnaces), heavy manufacturing, computing (think about training large models).
  - But most require significant infrastructure and logistics to operate. Computation does not.
- A controllable load. We need to be able to ramp the controllable load down when the renewable resources are not producing enough – or are forecast to be unable to produce.

LANCIUM
BALANCING ENERGY

Scientific Computing

- Computing meets the criteria
  - All we need is power and fiber …
  - Most machines can be remotely managed
    - Still need hands
- Scientific computing specifically
  - The bulk of applications are “batch” and have no interactions with humans.
  - Users are already used to queueing delays … suspending a job for a bit is not the end of the world.
  - Need the ability to stop/start, persist/restart, and migrate. More on this tomorrow.

Lancium Compute

Unhook Computation from fossil fuels by dancing with the grid.

LANCIUM
BALANCING ENERGY

Lancium Compute Offerings

HPC PaaS

1. **HPC PaaS Offering**: Pre-configured HPC Clusters to focus your efforts on HTC and loosely-coupled parallel jobs
2. **Just Bring your Workloads and Data:** We make it easy to launch your workloads and retrieve the results for analysis

Custom HPC PaaS

1. **Customize HPC PaaS**: Modify the HW infrastructure of our HPC PaaS to suit your workload needs (and leverage our SW Stack)
2. **Just Bring your Workloads and Data:** Leverage our state-of-the-art HPC Software Stack to obtain the PaaS benefit on custom HW Config

Custom HPC System (BYODC)

1. **Customize HPC:** A data center to suit the specific requirements of your HPC environment and business model (HW, SW, cooling, space design etc..)
2. **Carbon-Free Environment**: Our clean campuses are **powered 100% by renewable energy** to host and maintain your HPC Infrastructure

LANCIUM
BALANCING ENERGY

300 MW at Ft Stockton, TX, January 2023.
200 MW in Abilene, TX in May 2023.
Another 1 GW at Abilene approved.
Another 1 GW in approval process in Childress, TX.

PaaS “run a container” service

- Singularity based.
- Users can upload their own images or use provided images.
  - Users specify the job resource requirements, e.g., 8GB of memory, 8 vCPUs, and the command line to run, files to stage in and out, e.g.,
- CLI, API, browser
- Single node codes and low-degree parallel MPI – currently up to 96 nodes

```
lcli job run  --name "BLAST Tutorial job” --command "blastp –query \
        cow_db_small/cow.1000.protein.faa -db human_db/human.1.protein.faa –evalue 1e-5 \
        -max_target_seqs 1 -num_threads 4" --image lancium/blast --cores 4 --mem 8 \
        --input-file cow_db_small.tar.gz --input-file human_db.tar.gz
```

LANCIUM
BALANCING ENERGY

Lancium Compute Today

- Providing ~ 1M vCPU hrs/week for Open Science Grid .. 42M+ vCPU hours so far.
- Have run CERN/CMS jobs for US CMS.
- Running low-degree parallel MPI CFD jobs for commercial customer– less than 100 nodes.
- Rendering for Grid Markets.
- Financial & business analytics – essentially ML/AI.

LANCIUM
BALANCING ENERGY

OSG/CMS and CVMFS

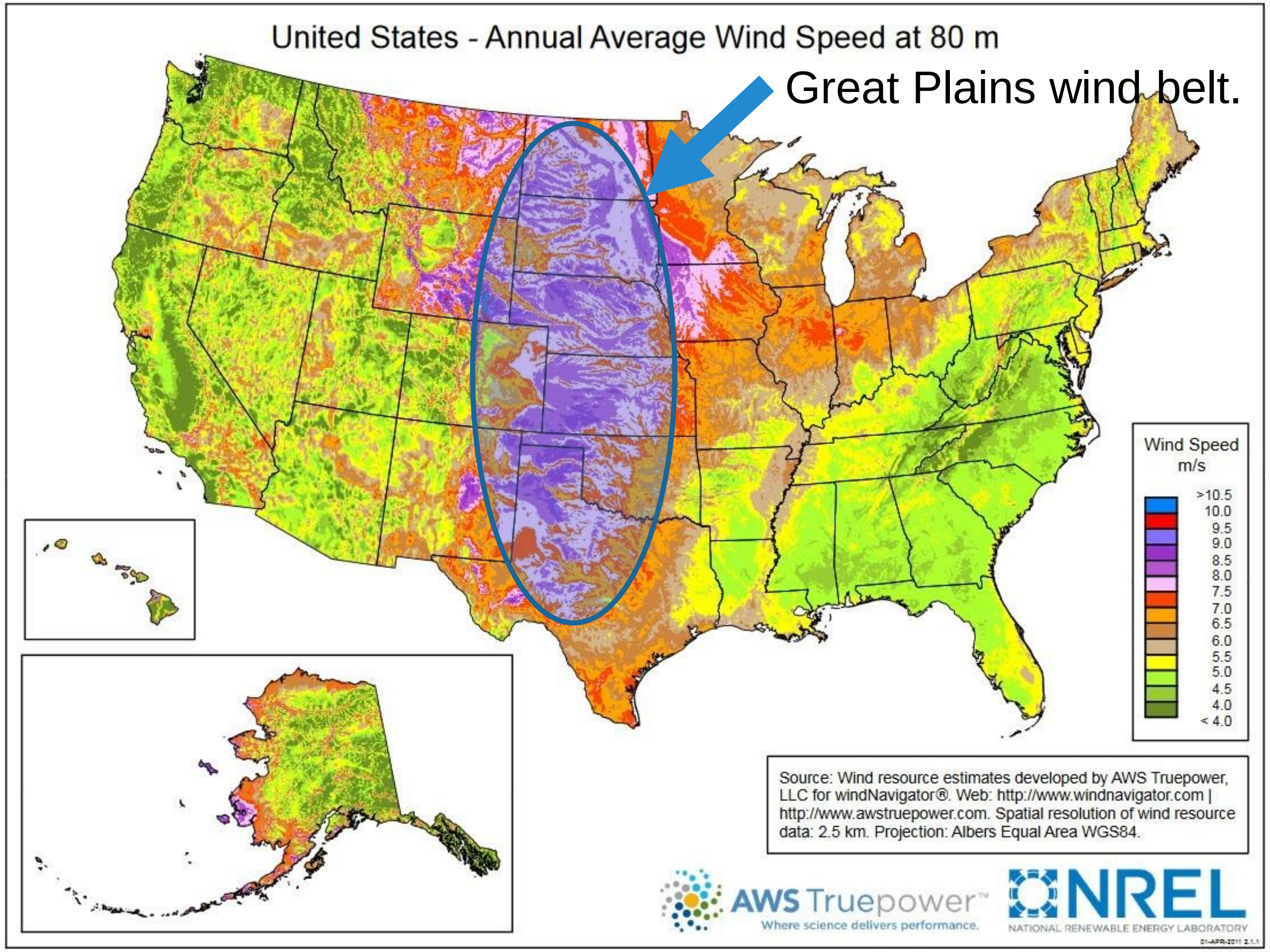
- To port OSG onto we Lancium Compute we installed Squid and mounted CVMFS on all compute nodes, and then bound it into Singularity container instances.
  - Initially only a single centralized Squid instance.
  - Later we went hierarchical with caches on each head node.
  - Getting CVMFS operational was pretty easy.
- To support CVMFS we installed Frontier everywhere.

LANCIUM
BALANCING ENERGY

Conclusion

Our vision is to build geographically dispersed data centers.

- In the “wind belt”.
- When the wind does not blow in Texas, we’ll migrate north to the Dakota’s, or somewhere else where the wind is blowing, the sun is shining, and the electricity is cheap and carbon-neutral.
- Further, with decent forecasting we can begin migrations before the power levels shift, particularly with jobs that are being scheduled for the first time.



LANCIUM
BALANCING ENERGY

I believe we are at the beginning of a new industrial revolution; just as oil, gas and coal replaced water and animal power as the power sources of industry, wind and solar will gradually yet surely replace oil and gas as the primary power source.

And just as in the first industrial revolution, it is the efficiency of the new technology that is the main driver in the transition. While it is true that it is in our collective best interest to turn away from carbon-emitting energies, it is the fact that electricity is simply cheaper when generated from renewables that is securing a carbon-free, or carbon-reduced, future.

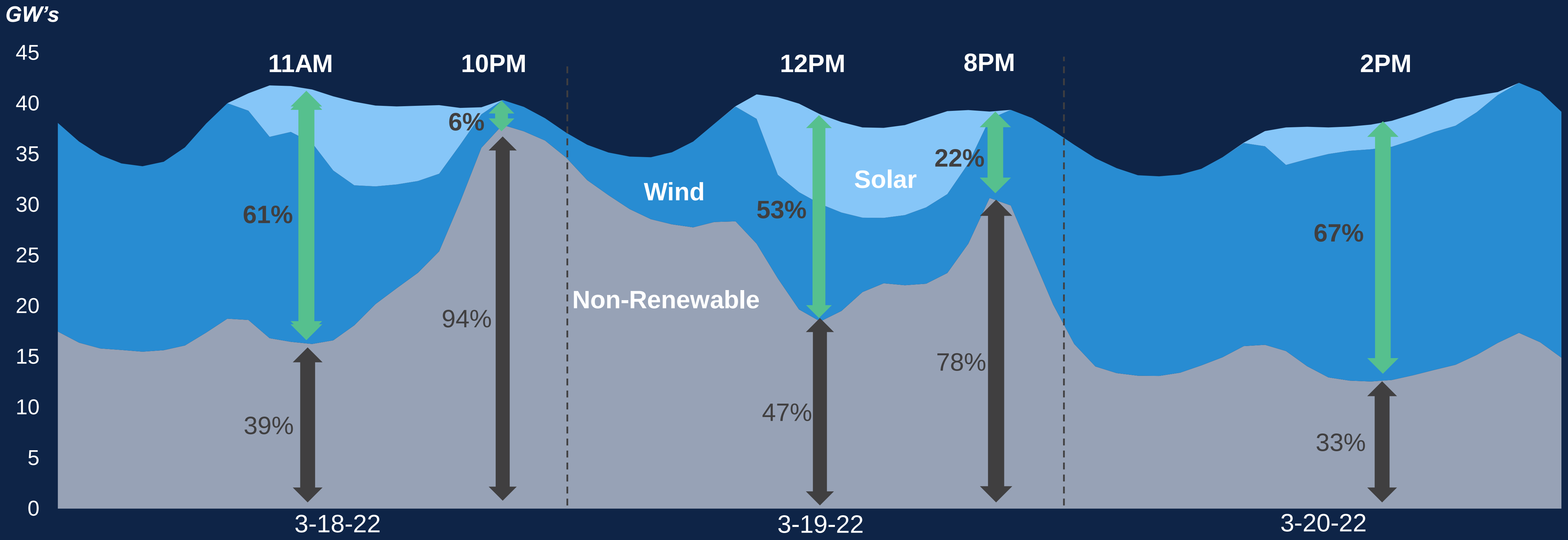
Lancium is at the center of this transformation.

END

Texas West Load Zone

LANCIUM
BALANCING ENERGY

Hourly balancing challenge for ERCOT – March 18, 2022 – March 20, 2022



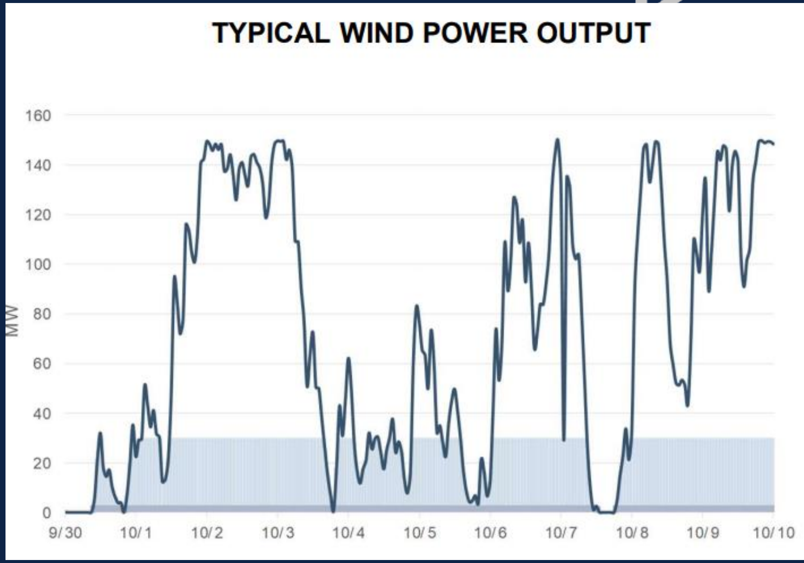
Batteries and load flexibility can provide short-term balancing throughout the day

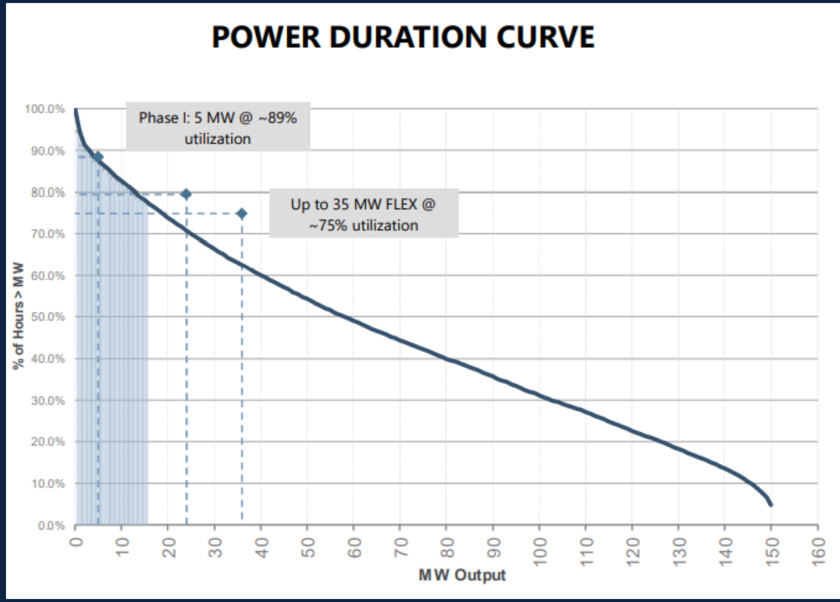
LANCIUM
BALANCING ENERGY

Sources and Notes: ERCOT

Wind power production functions

- On the right is data from a small, 150MW, facility in the panhandle near Lubbock.
- One rarely gets the full capacity out of a wind system.
- Below that is a graph that shows what % of time you get a certain amount of power.
- Wind power can vary very rapidly.





LANCIUM
BALANCING ENERGY

Wind and Solar are complimentary

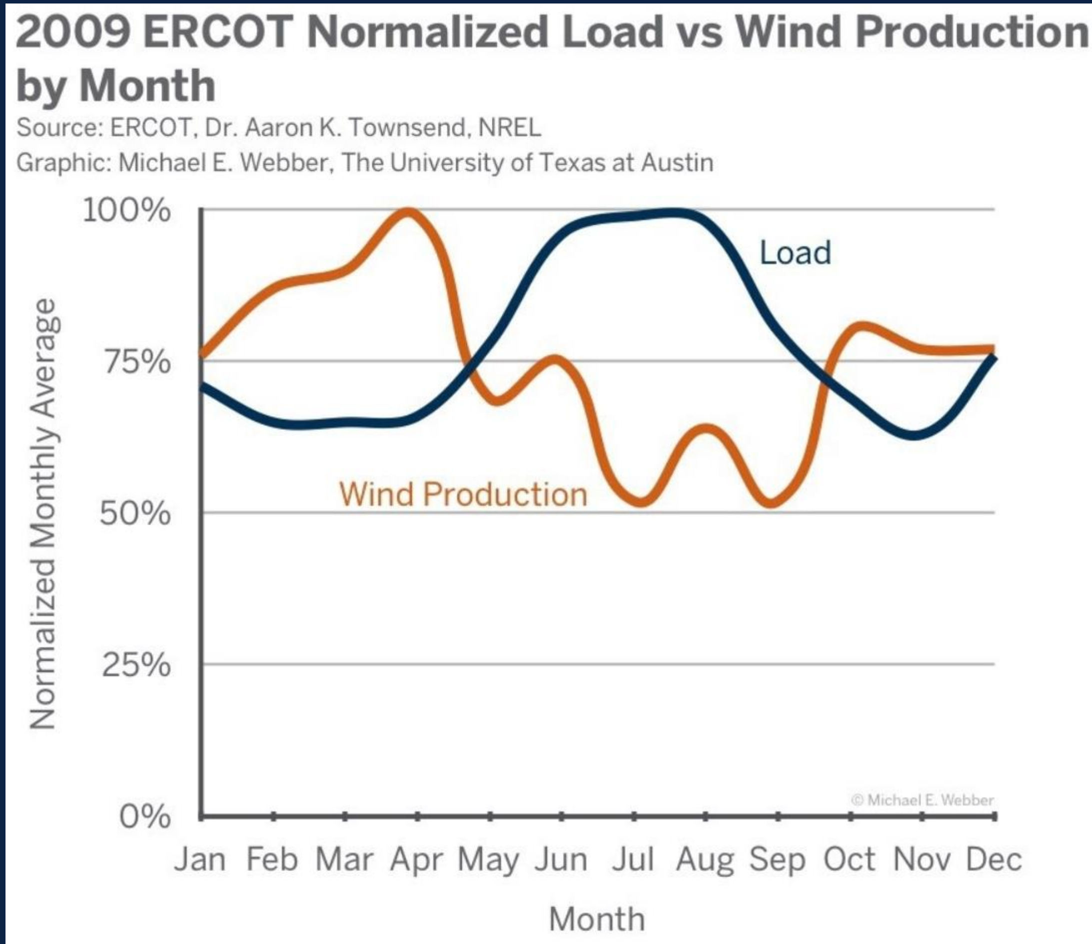


Figure 119: Continental wind production is lowest during the summer months, when demand for electricity is at its greatest.

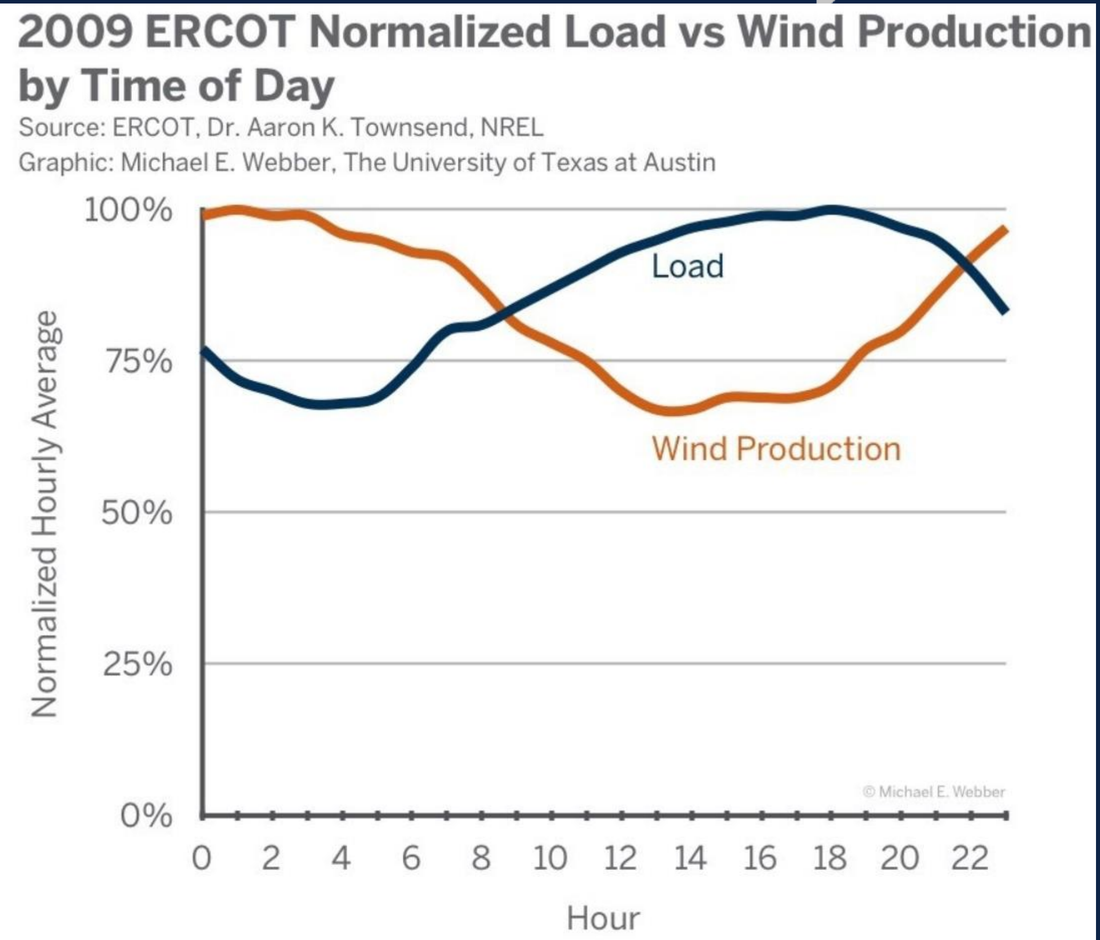


Figure 120: Continental wind is also out of alignment with demand during the day, with production being lowest during the times with the most demand.

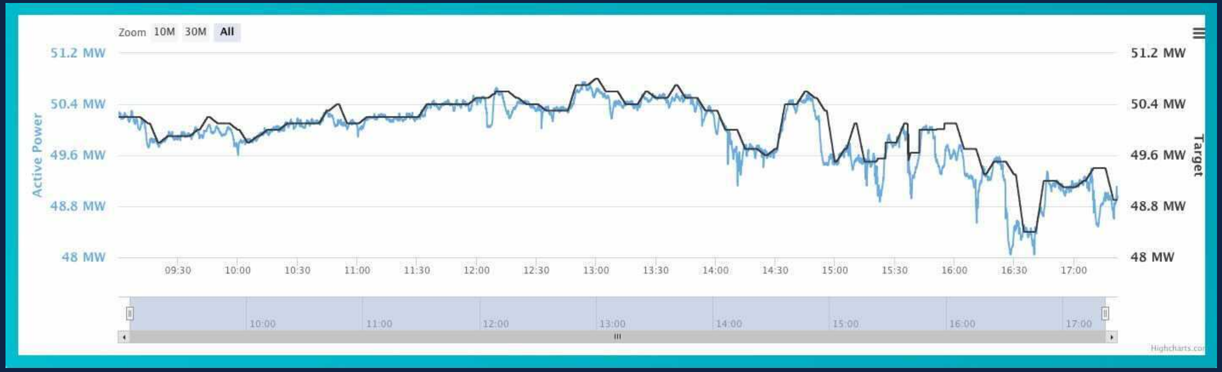
LANCIUM
BALANCING ENERGY

Frequency Stability

LANCIUM
BALANCING ENERGY

Frequency & Inertia

- As demand approaches supply
  - Frequency starts to drop
  - As frequency starts to drop current starts to rise on spinning generators.
  - As current starts to rise, so too does temperature and R
  - If it goes to far, generators “trip-off”.
  - This causes supply to drop – potentially causing a chain reaction.
  - To prevent collapse, load is shed by turning off load zones, i.e. blackouts.
- Immediate response to stabilize is “Primary Frequency Response”, must be done very rapidly. Typically within 15 seconds.
- Longer term response, e.g., within 15 minutes is called “Load Response” or LR. This is addressed by adding more generation or reducing load. Increases and decreases are controlled by the grid operator using computer generated “set points”.
- If you can vary your load rapidly, and in response to grid operator signals or price signals, you can be paid for it in the “ancillary markets”.
  - This can significantly reduce the price you pay for electricity.



LANCIUM
BALANCING ENERGY

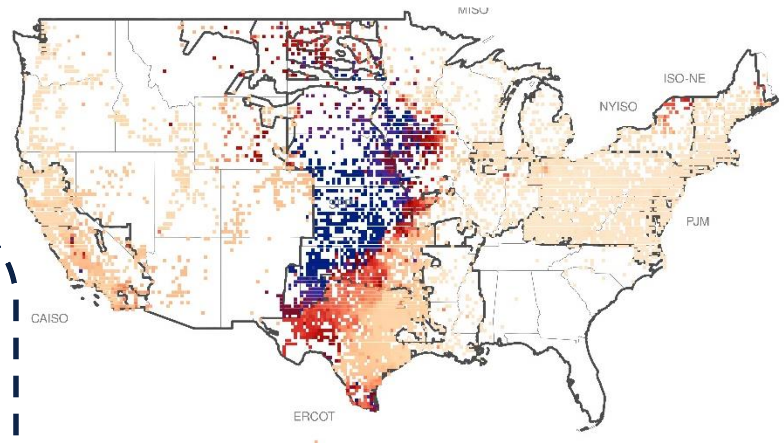
….Causing Increased Prevalence of Negative Power Prices…

- Depressed power prices can inhibit the buildout of new renewable generation

Frequency of negative real-time pricing

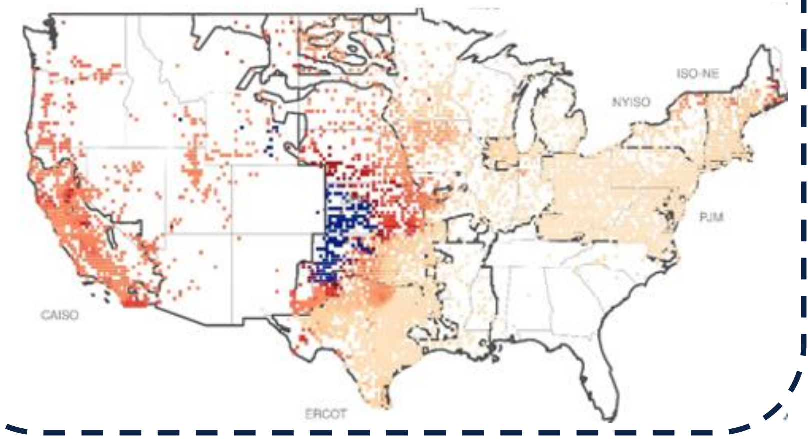
2013

2021



Lancium founded

2017



0% **10%** **20%**

Source: Lawrence Berkeley National Laboratory – Energy Markets & Policy – Frequency of negative locational marginal prices at nodes in the seven organized wholesale markets in the United States

Controllable Loads Are the Best Application of Grid Flexibility

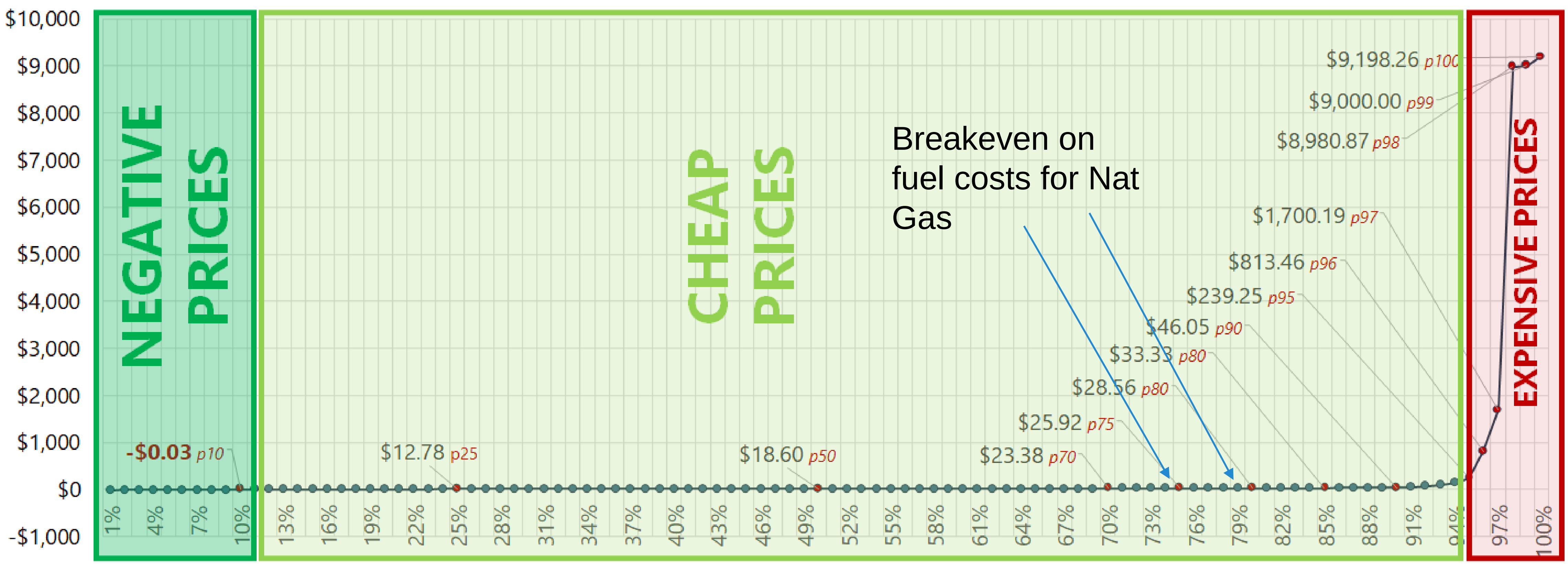
- Technological options and solutions for flexibility

		Flexibility			Economy	Quantity	Emissions
Resource designation	**Resource type**	**Start-up time**	**Run-time**	**Rampability**	**Start-up cost (USD/MWh)**	**Minimum load** (% of nominal power)	**Emissions intensity**
Baseload fossil generation	Generation	1–10 h	Continuous *(Subject to frequent outages)*	✗	55 - > 100	25–50%	High
Natural Gas Peaker	Generation	5 min	~ Days	✓	< 1 *(excludes degradation)*	20% per unit	High
Batteries	Storage	Seconds	< 4 h	✓	0 *(excludes degradation)*	0%	Varies
Load - Traditional demand response	Load	Minutes	Continuous	~	N/A	5–40%	N/A
LANCIUM	**Load**	**Seconds**	**Continuous**	✓	**0**	**0%**	**Low-Negative**

Source: International Renewable Energy Agency

2021 YTD REAL-TIME PRICE DISTRIBUTION

SmartResponse Value Proposition - Economic Dispatch in West Texas (LZ_West – 01/01 to 06/28)



YTD - 10% of all hours settled negative

Lithium is not the answer

At $100/KWh, one GWh of storage costs $100M

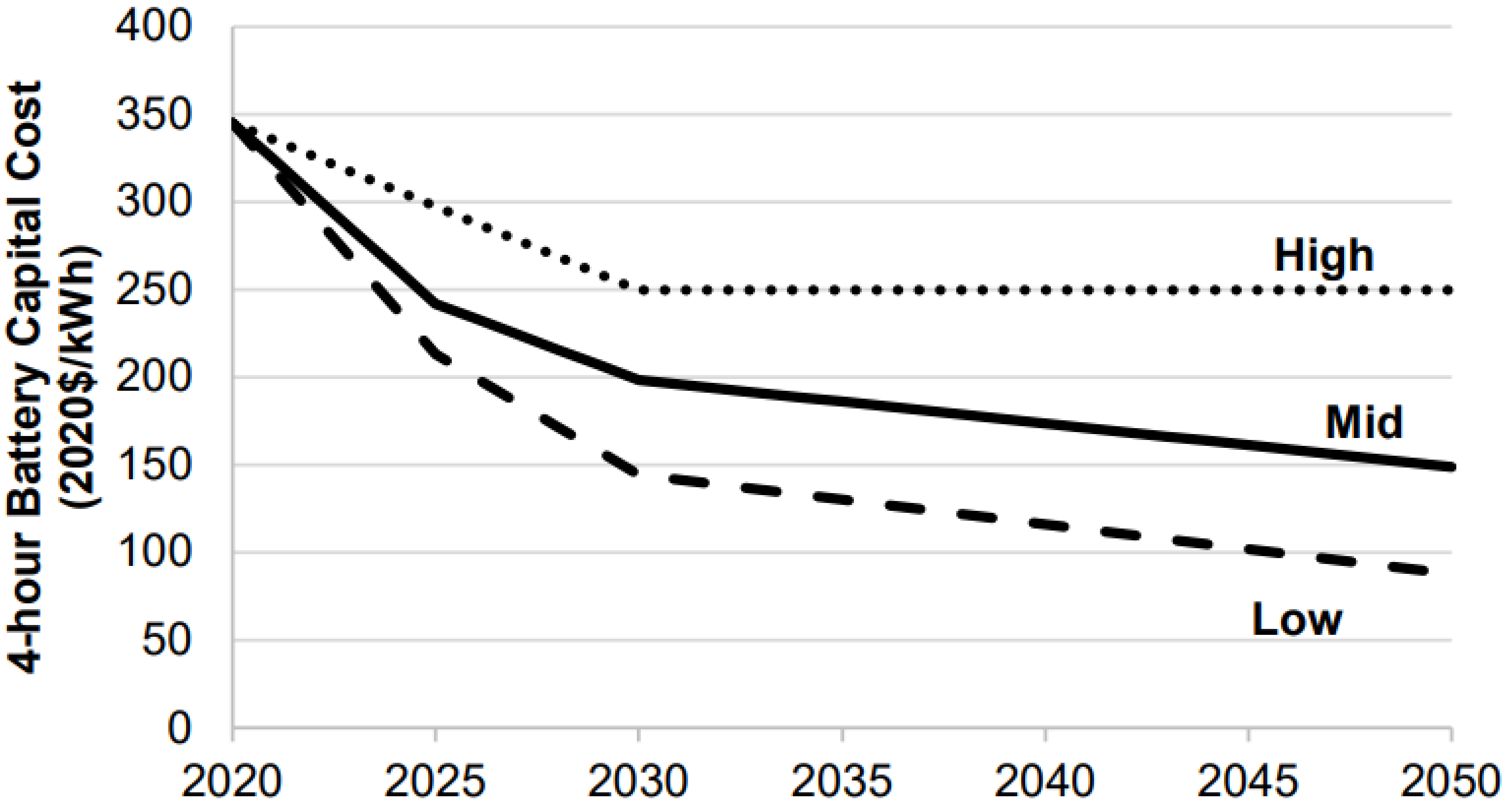
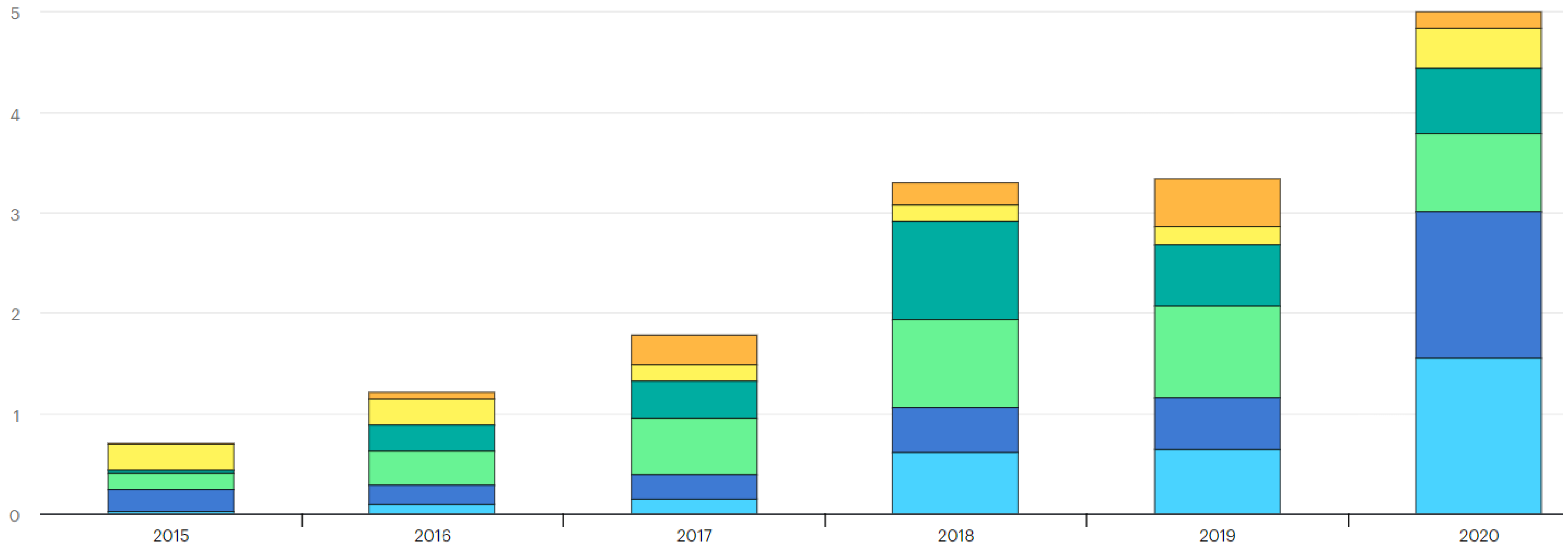


Figure ES-2. Battery cost projections for 4-hour lithium ion systems.

Cost Projections for Utility-Scale Battery Storage: 2021 Update Wesley Cole, A. Will Frazier, and Chad Augustine, National Renewable Energy Laboratory, NREL/TP-6A20-79236 June 2021. For non Li, see https://ambri.com/technology/

Energy storage (batteries) produced per year (GWh).



https://www.iea.org/reports/energy-storage 2022-09-07