Simulated AI Training Loads

WHITE PAPER

by Rich Scroggins



Introduction

The volatility of AI training workloads is causing difficulties for data center operators, particularly when running on generator sets. Data center operators need to understand how a generator set will perform when subjected to an AI training load profile, both on its own and when supported by a BESS.

This paper describes the unique qualities of AI training loads, some of the challenges those loads present for data center operation and grid stability, difficulties testing and modeling generator and BESS performance against these loads and performance of various generator/BESS hybrid system configurations and operating sequences to handle these workloads.

BACKGROUND

Servers cycle power as they are processing data, oscillating between peak and idle states. This is true of both conventional cloud servers as well as AI servers. The difference lies in the magnitude of the power drawn by the servers and the fact that cloud servers operate asynchronously while AI servers operate synchronously. In a cloud data hall there are hundreds of servers. At any point in time some will be at peak load and some will be at idle. In aggregate the peaks and idle states tend to cancel each other out and the load profile tends to be relatively flat. With AI workloads, the servers are operating synchronously, which means all are peaking at the same time and all are in an idle state at the same time. This is known as parallel processing and is the primary driver of the volatility of the AI workloads.

It is important to differentiate between AI training and AI inference. AI training refers to building the large language models (such as ChatGPT). This is what is going on in AI factories. AI inference refers to end user queries of the large language models. Inference workloads use parallel processing, like AI training but are nowhere near the scale of AI training workloads.

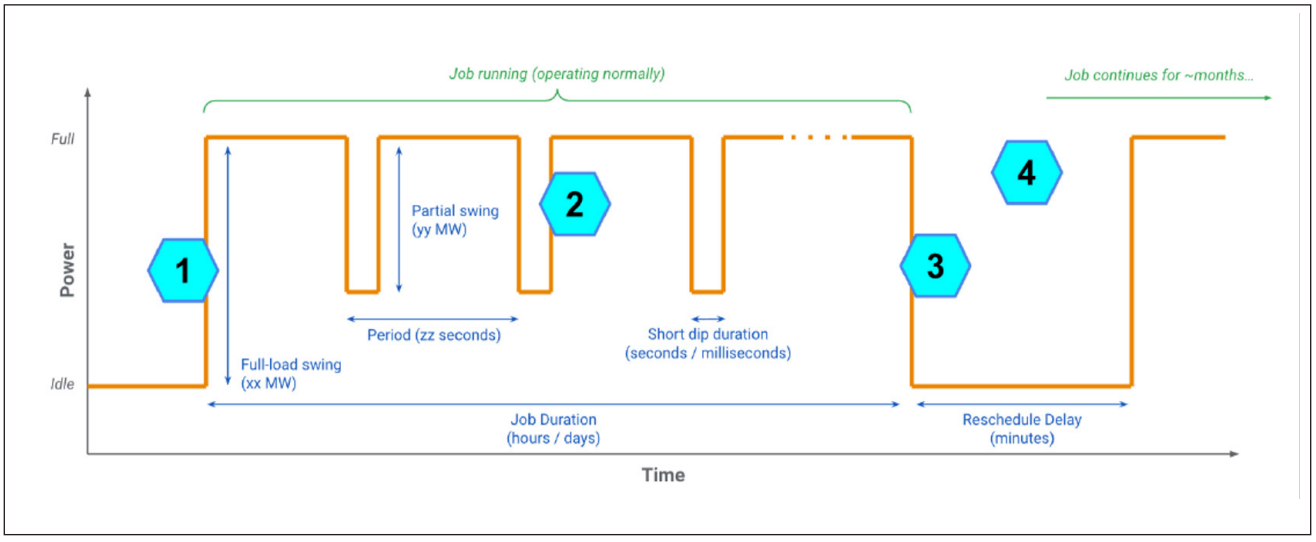


Figure 1. Google blog post – "Data Center Load Fluctuations"

Figure 1 is a simple illustration of a server executing a task. As the task begins the power level oscillates between a peak level and an active idle level. As the task is completed the power level reduces to a deep idle level until the next task is presented.

AI Inference also uses parallel processing however the power density and volatility of AI inference workloads is much less than the AI training load profile. That doesn't suggest that the challenges created by AI inference won't be significant. Unlike AI training workloads which will likely be concentrated in dedicated training data centers, AI inference workloads may be deployed in mixed-use colo data centers or small edge data centers, so their impact on the local electrical system must be understood and controlled. Figure 2 is a representation of training vs inference load profiles.

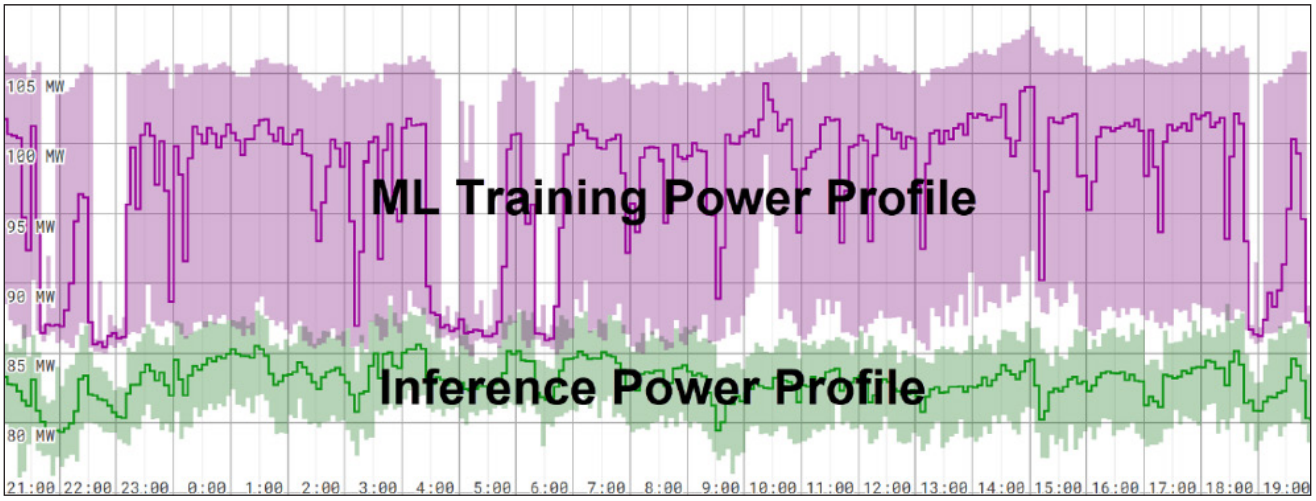


Figure 2. Google blog post – "Data Center Load Fluctuations"

AI load profile mitigation strategies

Data center operators and equipment suppliers are pursuing three methods to minimize power system transients imposed by AI training workloads. There are server software changes that can be used to smooth out the load profile, trading performance and energy efficiency for a reduction in volatility. Capacitors located at the server racks or integrated into the UPS absorb the subcyclic transients. BESS adjacent to generation and the utility connection will mitigate the effects of training workloads on generator sets and the power quality issues such as flicker that are imposed on the grid. In this paper we are going to be focusing on how BESS can be used to augment generator performance with AI training workloads.

Power system requirements with AI training workloads

AI training workloads are batched processes, which means they can be interrupted and restarted without impacting the final product. On site power systems won't have the same capability as a utility to handle volatile AI training loads so data center operators have a need to understand how volatile of a load profile a generator set or hybrid system can handle while maintaining suitable voltage and frequency characteristics. Power engineers need to define the maximum load transient the generator can support and maintain acceptable performance so data center engineers can curtail the load, so they don't exceed the maximum transient.

When powering AI training workloads it isn't necessary to maintain ISO 8528 G2 or G3 frequency and voltage regulation as the power is all rectified to dc. At a high level there are only a few basic requirements:

- The system must carry the load. The transient can't cause the generator to shut down
- Maximum voltage and frequency transients. Protective relays or rectifiers may trip off line if voltage or frequency transients exceed the setpoint.
- Maximum voltage and frequency ramp rate. UPS may reject the source if the voltage or frequency slew rates are too high.

Specfic parameters will vary by installation.

Simulated AI training load testing

In order to simulate AI training workloads, Cummins used a grid simulator in our Power Integration Center (PIC) lab. This grid simulator can be configured to operate as a source or a load. The simulator is operated in electronic load mode for this project. The simulator offers two advantages over a conventional load bank for AI load testing.

1. IT loads are constant power loads. Conventional load banks are constant impedance loads. The grid simulator is configured to create a constant current load which is a more accurate representation of an IT load.

(IT loads are fed by a switched mode power supply which actively controls the pulse width to maintain constant power, so if the source voltage drops the power supply will draw more current to sustain the power setpoint. A conventional load bank is a constant impedance load. If the source voltage drops the current and power drawn also drop. Generator controls have traditionally used a V/Hz function (if the frequency drops the voltage setpoint is reduced which reduces load on the engine) to assist generator sets in picking up large load. This doesn't work with active power loads because a reduction in voltage leads to an increase in current so load on the engine is not reduced. Turning off the V/Hz function when powering a conventional resistive load bank causes the generator set performance to more closely resemble performance with a constant power IT load. This is often done because constant power loads are not readily available.)

2. Conventional load banks aren't able to ramp loads. Load steps are simply switched on and off at most in one second increments. With the grid simulator entire load profiles can be recreated including load steps and ramps. These ramps can be as fast as 1ms from no load to full load or as slow as needed.

The grid simulator has a 500 kVA capacity, smaller than a typical generator set used in a data center, but it is expected that results will scale and be directionally correct. In this testing, 50 msec spikes were neglected. As these spikes are shorter than the relevant time constants of the AVR and governor control systems, the generator control will not respond to them. Further, it is expected that AI training workloads would have energy storage such as capacitors installed at the racks which would absorb these fast transients.

The load profiles used in the testing follow the general form of figure 3 on the following page.

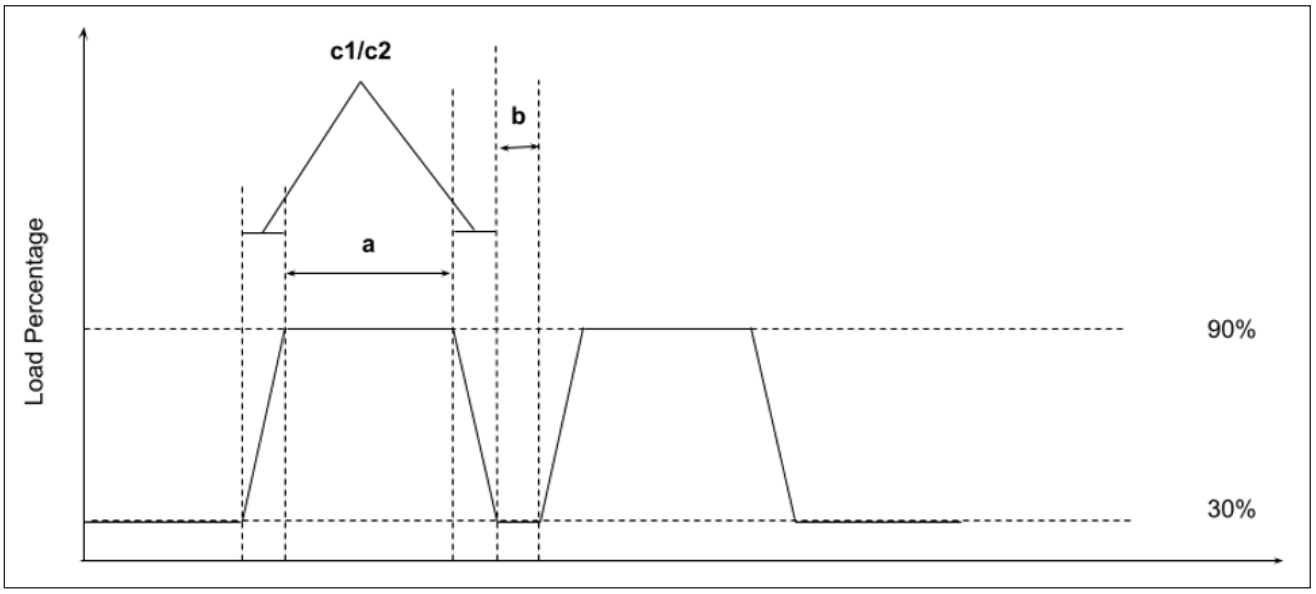


Figure 3. General load profile form

In general, test loads fluctuated between about 50% and 90% of generator set capacity, with load peaks lasting several seconds and troughs lasting about 1 second. Figure 4 displays a one line diagram of the system used for testing. One of the loads in the drawing is actually the grid simulator, configured as a constant current load. The one line shows two generator sets that can be paralleled. Only one of the generator sets is used in the testing described in this paper.

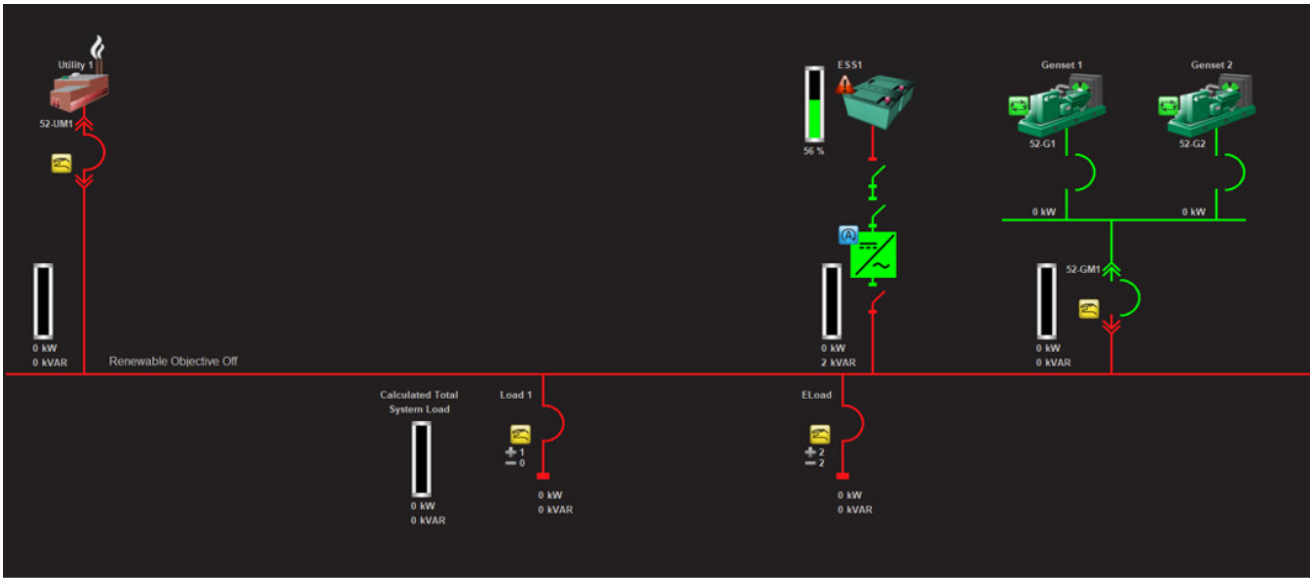


Figure 4. One line diagram for simulated AI load testing

Generator/BESS microgrid operation

With a BESS and generator set operating together to power AI training loads with minimal disruption the general goal will be to keep load on the generator set constant as much as possible and use the BESS to absorb all of the transients. There may be times when the generator set or BESS will need to support the loads independently or in parallel with the grid. In order to meet all of these requirements, the generator and BESS controls need to be able to support the following operating modes, either autonomously or in tandem with a microgrid controller.

Generator set operating modes

- **Isochronous (grid forming)** – Generator set control maintains constant voltage and frequency, either independently or paralleled with other generator sets. This is the standard mode of operation for generator sets running while not connected to a grid, either independently or paralleled with other generator sets in island operation.
- **Fixed power (grid following)** – Generator sets output real and reactive power based on external signals, either from a microgrid controller or other supervisory controls. This is the common mode of operation when generator sets are running in parallel with a grid, also known as load governing.

The BESS is capable of the following operating modes:

- **Grid forming isochronous** - Maintains nominal voltage and frequency.
- **Droop** – The BESS inverter maintains voltage and frequency based on a droop characteristic. In a droop system frequency is set based on charging and discharge rate in a linear relationship. As the BESS discharges real power, frequency decreases according to the droop characteristic. As the BESS charges frequency increases. Figure 5 illustrates this characteristic. The slope can be set to be more or less aggressive depending on the desired change in power corresponding to a deviation from nominal frequency. Reactive power (VARs) and voltage follow a similar characteristic.
- **Dynamic droop** – System operates in droop mode with microgrid controller dynamically adjusting the droop intercept point to maintain nominal voltage and frequency or real and reactive power
- **Fixed power (Grid following)** – System provides real and reactive power based on signals from the MGC

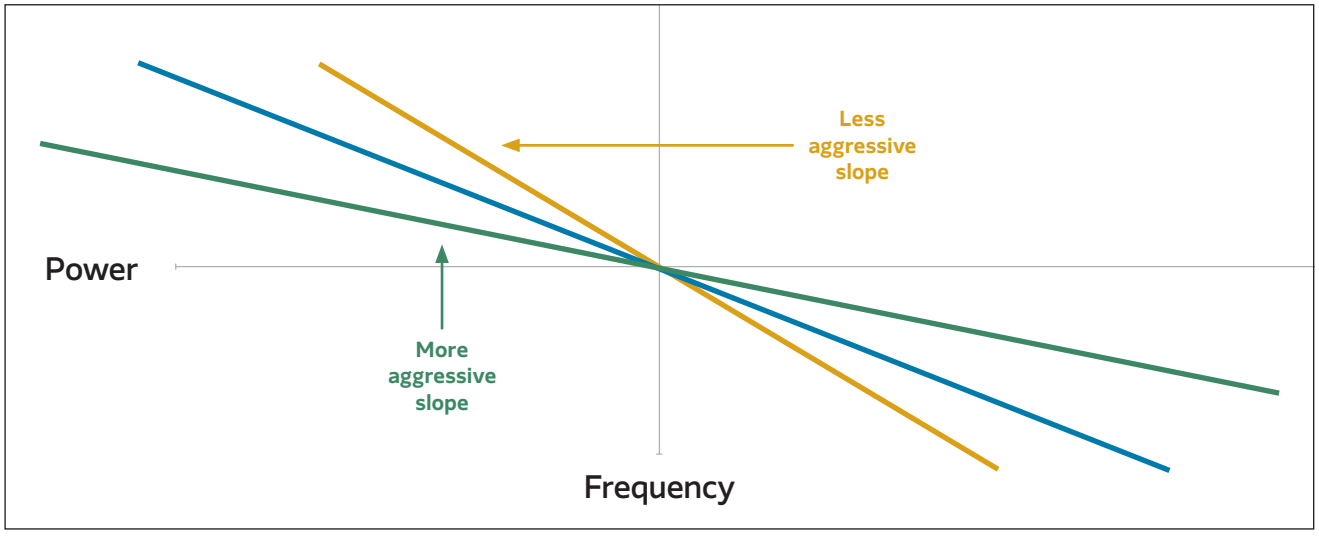


Figure 5. Droop characteristic

System configurations and operating modes

When combining generator sets with BESS to serve AI training loads there are several system operating modes to be considered. Where practical the aim should be to keep the load on the generator set constant and let the BESS absorb all of the transients. There will be circumstances where this will not be practical and tradeoffs will have to be evaluated.

Genset only

The Genset operates in isochronous mode. This is the standard operating mode for generator sets operating islanded. Even in systems that have BESS supporting generator sets there will likely be periods where the generator set needs to carry the AI loads.

An example scenario is one where the generator set is primarily used as backup with prime power coming from the grid, supported by BESS. On a utility outage a typical operating sequence would be to disconnect BESS, start the generator set (or paralleled generator sets), allow the gens to take load and stabilize, then sync the BESS to the gen.

The reasons for operating this way are that when the BESS is connected to the grid it will be operating in a grid following mode. When the grid is disconnected the BESS will have to transition to grid forming mode and many BESS systems do not make that transition seamlessly. A second reason for disconnecting BESS is that the BESS system may not be sized sufficiently to carry the entire load. In that scenario the BESS could become overloaded and cause breakers or protective relays to trip.

Figure 6 is a plot of a 275 kW generator set feeding the AI load. The load is oscillating between about 135 and 240 kW, or about 50% and 90% of capacity. We see voltage fluctuating between about 440V and 490 V and frequency fluctuations between 58 and 62 Hz. Whether this is a problem for the system will be due to some application specific factors: Are there underfrequency or undervoltage protective devices that will trip at these levels? If IT loads are being fed through a UPS, will this fluctuating voltage and frequency create problems? If a UPS is operating in bypass mode it may not be able to sync with the generator in order to return to double conversion mode due to the slew rates of the frequency or voltage.

Allowing the generator to run in isochronous mode in this condition requires no modification to standard settings. It is possible that certain settings could be adjusted to optimize genset-only performance with AI load. This is outside of the scope of this paper.

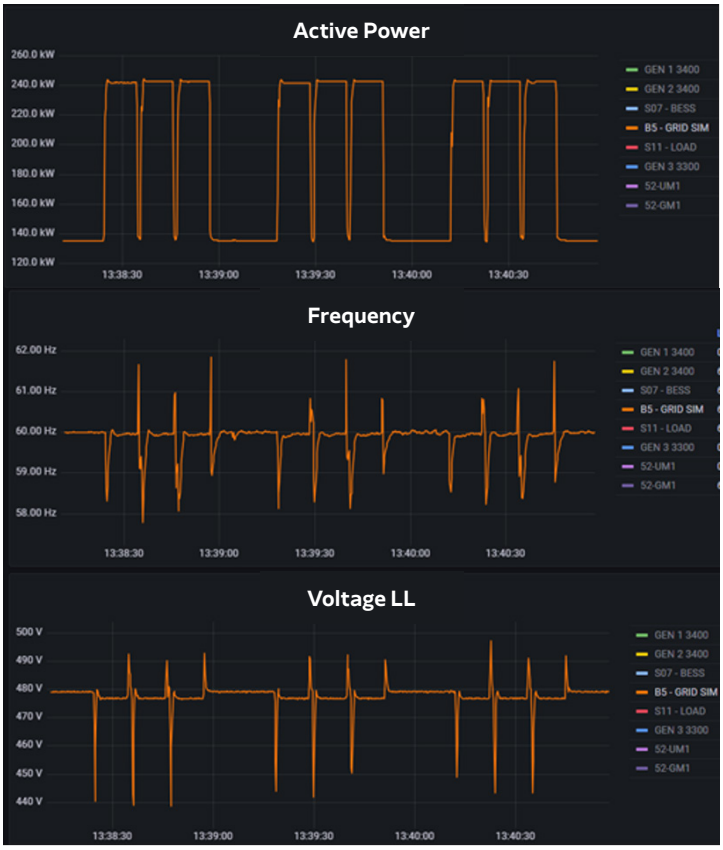


Figure 6. Simulated AI training load powered by generators only

BESS only

The scenario where a BESS would be the only asset powering the loads would be a demand response event. With some data centers operating where grids are constrained some utilities are mandating demand response support from large load operators. The BESS operates in grid forming mode with an aggressive droop slope. This means that the BESS will change frequency dramatically as the load fluctuates. The result is a frequency waveform that is a mirror image of the power waveform: When the power is at a peak, the frequency is at a trough.

See figure 7 to the right.



Figure 7. Simulated AI training load powered by BESS alone

In an application where only a single BESS is powering the load, operating the BESS in isochronous grid forming mode would probably be optimum. The purpose of this test is to characterize expected performance of a system with multiple BESS connected in parallel at medium voltage. For this reason droop operation was selected for this test

Hybrid systems, Genset + BESS

BESS IN GRID FORMING MODE, GENSET IN FIXED POWER MODE

The MGC sets nominal the voltage and frequency points for the BESS and sets the kW and kVAr (P and Q) points for the generator set. (Alternatively the kVAr point could be expressed as a power factor.) This mode should not be used if the generator is going to be online without the BESS. Figure 8 is a plot of this operating mode. A 275 kW generator set is powering AI loads fluctuating between 110 kW and 250 kW, about 40% to 90% of generator set capacity.

This operating mode yields excellent performance. With the yellow line representing genset output power we see the load on the genset remains constant. We see the BESS absorbing all the transients. Frequency variation is less than 0.3 Hz and voltage variation is less than 4 V, for a 50% load swing.

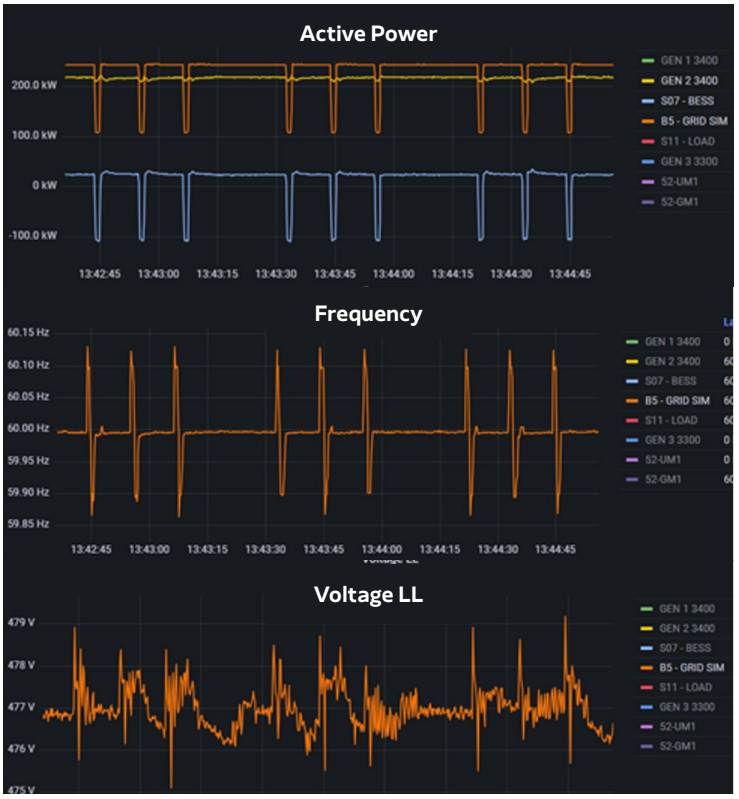


Figure 8. BESS in grid forming mode, genset in load govern (grid following) mode

GENSET IN ISOCHRONOUS, BESS IN DYNAMIC DROOP WITH AGGRESSIVE DROOP SLOPE

The MGC adjusts the BESS droop intercept point to keep the genset power constant. This mode should be used if there are situations where the genset will be online without the BESS, for example if the genset needs to establish the grid after an outage because the BESS doesn't have the capacity to support the entire load. This mode can also be used for adding BESS to a legacy system as it doesn't require any changes to existing generator control settings. In this mode the MGC adjusts droop intercept to keep genset power constant.

Figure 9 is a plot of a hybrid system using this configuration feeding a simulated load oscillating between 110 kW and 250 kW, a swing of 40%-90% of the generator sets 275 kW rating. The MGC controlled the BESS droop intercept point to maintain the generator set output at 10% of its rating. While the BESS absorbs the spikes the system fluctuates in this mode more than it does when the BESS is forming the grid. The generator set is not able to react to load changes as fast as the BESS does and the MGC doesn't actively control the BESS power setting fast enough to keep the generator power level constant. The frequency variation is about twice as high as what it was with the BESS in grid forming mode although it is likely acceptable for most systems.

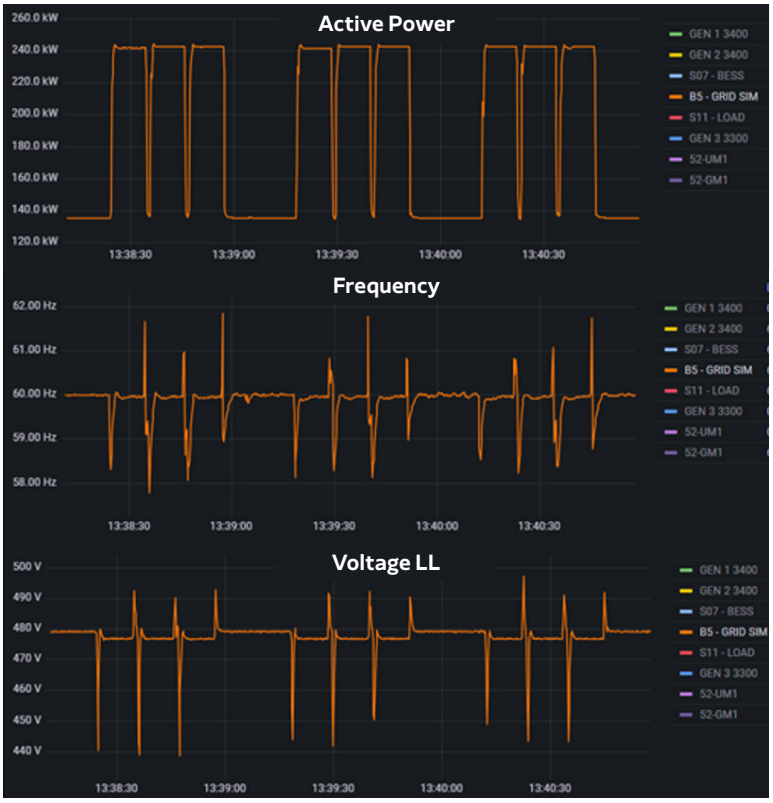


Figure 9. Gen in isochronous mode, BESS in dynamic droop

Conclusion

Data center operators and equipment suppliers are all evaluating how to best operate with volatile AI training loads. This is a considerable challenge as there is no real precedent for serving this type of load and each generation of AI server racks has different characteristics creating a moving target for the industry. For manufacturers of power generation equipment there are open questions regarding how much will rack adjacent energy storage (such as super capacitors) and system impedances (such as transformer windings) impact the load profile actually seen by the assets.

In this paper we've talked about the challenges of supporting these load profiles and also testing and modeling generator and BESS performance against them. We used a grid simulator to model the AI training loads using several different generator set and BESS configurations. We found that the goal of keeping load on the generator set constant and allowing the BESS to absorb the transients is best met by configuring the BESS in grid forming mode and running the generator set in load govern mode. There are however, circumstances where this configuration will not be practical. Operators will need to consider multiple options for ideal performance.

As AI load profiles, power system architectures and energy storage assets evolve operators and equipment suppliers will need to continue to evolve in their thinking about how to support these loads. The ability to model workloads and power system performance will be important for operators and suppliers in their ability to support and influence the data center industry.

About the author

Rich Scroggins
Technical Advisor



Rich Scroggins is a Technical Advisor in the Application Engineering group at Cummins. Rich has been with Cummins for 18 years in a variety of engineering and product management roles. Rich has led product development and application work with transfer switches, switchgear controls and networking and remote monitoring products and has developed and conducted seminars and sales and service training internationally on several products. Rich received his bachelors degree in electrical engineering from the University of Minnesota and an MBA from the University of St. Thomas.

Cummins Inc.
Box 3005
Columbus, IN 47202-3005
U.S.A.

cummins.com

Bulletin 6705740 Produced in U.S.A. 7/26
©2026 Cummins Inc.