



University of North Carolina at Charlotte (UNC Charlotte)

National Laboratory of Rockies (NLR)

National Rural Electric Cooperative Association (NRECA)

Integrated Data Center Modeling - Dynamic Modeling Approaches and Case Studies

September 2026

Md Hasnain Arifin

Sadique Faisal Khan

Michael Smith

Sukumar Kamalasadan

Yashodhan P. Agalgaonkar

Ravindra Singh

Arun Ravindran

DISCLAIMER

This report was prepared as an account of collaborative research. Neither the University of North Carolina at Charlotte (UNC Charlotte), the National Laboratory of Rockies (NLR), the National Rural Electric Cooperative Association (NRECA), nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the aforementioned institutions. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the participating organizations.

Abstract

The rapid growth of large-scale data centers, driven by artificial intelligence (AI), cloud computing, and high-performance computing workloads, has introduced new challenges for modern power systems. Unlike traditional loads, data center loads exhibit fast dynamic behavior, high power density, and nonlinear characteristics, which can significantly impact system stability, oscillatory behavior, and power quality. Consequently, traditional modeling approaches are insufficient for capturing the true impact of data centers on power system dynamics.

This report presents a comprehensive dynamic modeling framework for integrated data center power systems, capturing the interactions between grid infrastructure, co-located generation resources, power electronic interfaces, and server-level loads. Detailed models are developed for key components, including power supply units (PSUs), uninterruptible power supplies (UPS), IT loads, and AI-driven load profiles. Each component is represented using control-oriented dynamic models and validated under various operating conditions to ensure accurate system-level representation.

An integrated system model is constructed by coupling the data center architecture with a benchmark transmission system, enabling the study of grid–data center interactions under realistic operating scenarios. The framework supports analysis of both steady-state and transient behaviors, including voltage and frequency dynamics, oscillatory modes, and harmonic distortion.

A series of case studies are conducted to evaluate the impact of data center operation on power system performance. These include large load variations, source transitions, and disturbance scenarios. The results demonstrate that data center loads can significantly influence system dynamics, including shifts in oscillatory modes, reduction in damping characteristics, and increased harmonic content.

The proposed modeling and analysis framework provides a systematic approach for evaluating the dynamic impact of large-scale data centers on power systems. The findings highlight the importance of accurate load representation and coordinated control strategies in maintaining system stability and ensuring reliable operation in future power grids with high data center penetration.

Acknowledgments

This work was supported by UNC Charlotte. The authors—Md Hasnain Arifin, Sadique Faisal Khan, Michael Smith, Sukumar Kamalasadan, Yashodhan P. Agalgaonkar, Ravindra Singh, and Arun Ravindran—gratefully acknowledge the collaborative resources and technical expertise provided by the University of North Carolina at Charlotte (UNC Charlotte), the National Laboratory of Rockies (NLR), and the National Rural Electric Cooperative Association (NRECA).

Acronyms and Abbreviations

AC – Alternating Current
AI – Artificial Intelligence
BESS – Battery Energy Storage System
CCM – Continuous Conduction Mode
CPU – Central Processing Unit
DC – Direct Current
FRT – Fault Ride-Through
GFL – Grid-Following Mode
GFM – Grid-Forming Mode
GPU – Graphics Processing Unit
GT – Gas Turbine
HP – High-Priority
HV – High-Voltage
IBR – Inverter-Based Resource
IEEE – Institute of Electrical and Electronics Engineers
IT – Information Technology
LP – Low-Priority
LV – Low-Voltage
MOSFET – Metal–Oxide–Semiconductor Field-Effect Transistor
MV – Medium-Voltage
NERC – North American Electric Reliability Corporation
NPC – Neutral-Point-Clamped
PCC – Point of Common Coupling
PDU – Power Distribution Unit
PFC – Power Factor Correction
PID – Proportional–Integral–Derivative
PLL – Phase-Locked Loop
PSS – Power System Stabilizer
PSU – Power Supply Unit
PUE – Power Usage Effectiveness
PWM – Pulse Width Modulation
STG – Steam Turbine and Governor
THD – Total Harmonic Distortion
UPS – Uninterruptible Power Supply

Contents

Abstract	ii
Acknowledgments	iii
Acronyms and Abbreviations	iv
1.0 Introduction	1
1.1 Motivation: Data Center Impact on Power Systems	1
1.2 Need for Dynamic Modeling and Analysis	1
1.3 Objectives and Contributions	2
2.0 Computing Architecture of Modern AI Datacenters	3
2.1 Introduction	3
2.2 AI Compute Infrastructure	3
2.2.1 From Accelerator Servers to Rack-Scale AI Systems	3
2.2.2 Network and Logical Resource Architecture	4
2.3 AI Workloads and Resource Management	5
2.4 From AI Computation to IT Electrical Demand	6
2.5 Interface to the Data-Center Electrical System	9
2.6 Case Studies	9
2.6.1 Alibaba ASI: A Heterogeneous Production AI Infrastructure	9
2.6.2 Siemens/NVIDIA: A High-Density Greenfield Reference Architecture	10
2.7 Summary	10
3.0 Data Center Electrical Architecture	12
3.1 Introduction and Overview	12
3.2 Electrical Architecture and Power Flow Description	13
3.3 Architecture Representation and Modeling Approach	13
3.4 System Topology and Operating Modes	14
3.4.1 System Topology	14
3.4.2 Operating Modes	14
3.5 Component-Level Description	14
3.5.1 Utility Grid and Transmission System	15
3.5.2 Co-located Generation and Energy Storage	15
3.5.2.1 Gas Turbine Generator	15
3.5.2.2 Battery Energy Storage System	15

3.5.3	Uninterruptible Power Supply (UPS)	15
3.5.4	Power Distribution Unit (PDU)	15
3.5.5	Power Supply Unit (PSU)	15
3.5.6	IT Load	16
3.5.7	Cooling System Loads	16
3.6	Discussion and Summary	16
4.0	Data Center Modeling Approach with Grid and Co-Located Generation System Modeling with Protection System (Fault Ride Through) Included	17
4.1	Dynamic Modeling Framework for Data Center Electrical System	17
4.2	Modeling of Power Distribution Unit (PDU)	18
4.2.1	PDU Functional Representation	18
4.2.2	Electrical Modeling Approach	18
4.2.3	Protection and Switching Representation	18
4.3	Modeling of Power Supply Unit (PSU)	19
4.3.1	Modeling of Power Factor Correction (PFC) Converter	19
4.3.1.1	Topology	19
4.3.1.2	Filter Design	20
4.3.1.3	Control Architecture	22
4.3.1.4	Switching and PWM Implementation	23
4.3.1.5	Summary	24
4.3.2	Modeling of DC-DC Isolated Converter	24
4.3.2.1	Topology	24
4.3.2.2	Filter Design	26
4.3.2.3	Control Architecture	27
4.3.2.4	Switching and PWM Implementation	28
4.3.2.5	Summary	28
4.4	Modeling of IT Load	29
4.4.1	Overview of IT Load Modeling	29
4.4.2	Server Utilization Modeling	34
4.4.3	Server Power Consumption Model	34
4.4.4	Dynamic Behavior of IT Load	35
4.4.5	Coherent IT Load Representation	36
4.4.5.1	Implementation of IT Load in Simulation	36
4.4.6	Performance Evaluation of IT Load Model	37
4.4.6.1	AI Training Load Profile	37
4.4.6.2	Utilization-Based Load Profile	38
4.4.6.3	Discussion	38
4.5	Modeling of Uninterruptible Power Supply (UPS)	39
4.5.1	Topology	39
4.5.2	Modes of Operation	39
4.5.2.1	Bypass Mode (Eco Mode)	40
4.5.2.2	Double-Conversion Mode (Online Mode)	40
4.5.2.3	Battery Mode	40
4.5.2.4	Mode Transition	41

4.5.3	Filter Design	41
4.5.3.1	DC-Link Capacitor	41
4.5.3.2	Output Filter	41
4.5.4	Control Architecture	42
4.5.4.1	Rectifier Control (AC–DC Stage):	42
4.5.4.2	Inverter Control (DC–AC Stage):	43
4.5.5	Switching and PWM Implementation	45
4.5.6	Summary	45
4.6	Modeling of Cooling System	46
4.6.1	Cooling Load Representation Using Power Usage Effectiveness (PUE)	46
4.6.2	Cooling System Composition and Representative Load Breakdown	46
4.6.3	Cooling Load Modeling Approach	47
4.6.4	Electrical Interface and Voltage Level	48
4.7	Modeling of Colocated Power Plant Generation	48
4.7.1	Modeling of Gas Turbine (GT) Generator	48
4.7.2	Modeling of Battery Energy Storage System (BESS)	49
4.8	Protection System Design and Logic/Settings Considering Standards (IEEE 2800)	50
4.8.1	Fault Ride-Through (FRT) Philosophy	50
4.8.2	Voltage Ride-Through Requirements (IEEE 2800)	50
4.8.3	Frequency Ride-Through Requirements (IEEE 2800)	51
4.8.4	Protection Logic and Implementation	52
4.8.5	Post-Disturbance Recovery and Control Coordination	52
4.8.6	Summary	53
5.0	Study Framework for Impact Assessment of Data Center on Power System Dynamics	54
5.1	Test System Details	54
5.1.1	Overall Test System Configuration and Data Center Integration	54
5.2	Case Study Framework for Impact Assessment of Data Center on Power System Dynamics	57
5.2.1	Modal Characteristics and Harmonic Analysis under Different Load Conditions	57
5.2.1.1	Benchmark Loading Condition	57
5.2.1.2	Constant Data Center Load	58
5.2.1.3	Time Varying AI Training Load	58
5.2.1.4	Result Discussion	59
5.2.2	Modal Characteristics and Harmonic Analysis under Voltage Disturbance, and Grid–UPS Mode Transition	60
5.2.2.1	Test System & Event Details	60
5.2.2.2	Simulation Results	61
5.2.2.3	Result Discussions	63
6.0	Engineering Insights and Practical Implications	66
6.1	Introduction	66

6.2	Implications for Grid Stability and Planning	66
6.2.1	Interconnection and Transmission Planning	66
6.2.2	Dynamic Stability Assessment	66
6.2.3	Scenario-Based Planning and What-If Analysis	67
6.3	Challenges in Large Data Center Integration	67
6.3.1	Oscillatory Stability	67
6.3.2	Load Ramping and Energy Management	68
6.3.3	Disconnection, Source Transfer, and Reconnection	68
6.3.4	Ride-Through and Protection Coordination	68
6.4	Need for Advanced Control and Coordination	69
6.4.1	Coordinated Facility-Level Control	69
6.4.2	Oscillation Detection and Mitigation	69
6.4.3	Disturbance and Transition Coordination	69
6.5	Role of Data Centers in Future Power Systems	70
6.5.1	Flexible Computational Demand	70
6.5.2	Grid-Supportive Local Resources	70
6.5.3	Grid-Responsive Data Center Operation	70
6.6	Summary	70
7.0	Conclusions and Future Work	72
7.1	Summary of Key Findings	72
7.2	Contributions of the Study	72
7.3	Limitations	73
7.4	Future Research Directions	73
Appendix A	Detailed Modeling of Co-located Generation Components	75
A.1	Gas Turbine Generator Detailed Model	75
A.1	Synchronous Machine Model (dq-axis)	75
A.2	Steam Turbine and Governor (STG)	75
A.3	IEEE Type-1 Excitation System	76
A.4	Power System Stabilizer (Kundur PSS)	77
A.5	Model Parameters (from Simulink Masks)	77
A.2	Battery Energy Storage System Detailed Model	79
	References	81

List of Figures

2.1	Seven-day trace-derived GPU workload activity and illustrative H20-equivalent GPU-only electrical scenario. Panel (a) shows the reported aggregate GPU workload activity from the Alibaba trace. Panel (b) translates a capacity-consistent representation of the same workload into illustrative GPU electrical power using a 400 W peak-power assumption and an assumed 90 W idle power per GPU. The shared idle baseline is kept separate from workload-priority attribution, while the dynamic component is decomposed into high-priority (HP), low-priority (LP), and other activity. The MW values are illustrative and represent GPU electrical demand only; they do not represent measured Alibaba server-input, total IT, facility, or grid power.	7
2.2	Causal chain from AI workload activity to IT electrical demand. Workload behavior and scheduler decisions determine the active compute resources, while rack-level power conversion and energy storage can mediate the relationship between silicon power demand and the electrical demand observed by the facility.	8
3.1	Conventional Grid Connected Data Center Power Distribution System. . .	12
4.1	PSU Model.	19
4.2	PFC Converter Model.	20
4.3	Control architecture of the PFC converter.	22
4.4	DC-DC Isolated Converter Model.	25
4.5	Control architecture of the DC-DC isolated converter.	27
4.6	Physical representation and flow abstraction of the proposed IT load model.	32
4.7	Generalized IT load modeling framework.	33
4.8	Comparison of reference and modeled AI training load profiles.	37
4.9	Comparison of reference and modeled server load profiles.	38
4.10	UPS model showing rectifier, DC-link, battery, inverter, and output filter. .	39
4.11	Modes of operation of UPS.	40
4.12	Control architecture of the UPS Rectifier.	43
4.13	Control architecture of the UPS inverter.	44
4.14	Generalized control diagram for BESS Inverter.	49
5.1	Single Line Diagram of Kundur Two-Area System.	55
5.2	Kundur Two-Area System Connected with Data Center.	55
5.3	Single Line Diagram of Grid-Connected Data Center Model.	56

5.4	Grid Voltage with constant data center load.	61
5.5	Area-1 (Bus 7) Frequency with constant data center load.	62
5.6	Area-2 (Bus 9) Frequency with constant data center load.	62
5.7	Grid Voltage with AI Training Data Center Load.	62
5.8	Area-1 (Bus 7) Frequency with AI Training Data Center Load Profile. . . .	62
5.9	Area-2 (Bus 9) Frequency with AI Training Data Center Load Profile. . . .	63
A.1	Detailed control diagram for BESS inverter.	80
A.2	BESS outer loop.	80

List of Tables

4.1	Key parameters of the PFC converter	23
4.2	Key parameters of the DC–DC isolated converter	28
4.3	Key parameters of the UPS system	45
4.4	Cooling System Load Breakdown for 100 MW Data Center	47
4.5	Voltage Ride-Through Requirements	51
4.6	Frequency Ride-Through Requirements (Adapted from IEEE 2800 Table 3)	51
5.1	Identified oscillation modes for benchmark loading condition	58
5.2	Identified oscillation modes for constant data center loading condition . . .	58
5.3	Identified oscillation modes for time-varying AI training data center loading condition	59
5.4	Frequency dynamic characteristics under different data center load profiles and event scenarios	63
5.5	Modal characteristics and harmonic distortion under different data center load profiles and event scenarios	63
A.1	Synchronous machine parameters (per generator) used in the Kundur model implementation.	77
A.2	Steam turbine and governor (STG) parameters (per generator) used in the Kundur model implementation.	78
A.3	IEEE Type-1 excitation system parameters (per generator) used in the Kundur model implementation.	78

1.0 Introduction

1.1 Motivation: Data Center Impact on Power Systems

The rapid expansion of data centers has become a defining feature of modern power systems, driven by the increasing demand for cloud computing, artificial intelligence (AI), and large-scale data processing. These facilities are characterized by high power density, continuous operation, and rapidly varying load profiles, which distinguish them from traditional industrial and commercial loads.

Recent advancements in AI training and high-performance computing have further intensified the dynamic nature of data center loads. AI workloads, in particular, exhibit fast ramping characteristics and significant variability over short time scales. Such behavior introduces new challenges for power system operation, including increased stress on generation resources, voltage fluctuations, and potential degradation of system stability.

Unlike conventional loads that are often modeled as static or slowly varying, data center loads are predominantly interfaced through power electronic converters. These converters introduce nonlinearities, fast control dynamics, and interactions with the grid that can lead to oscillatory behavior and harmonic distortion. As a result, traditional modeling approaches are insufficient for capturing the true impact of data centers on power system dynamics.

Furthermore, modern data centers are increasingly integrated with co-located generation resources such as gas turbines (GT) and battery energy storage systems (BESS). These configurations provide operational flexibility and resilience but also introduce additional dynamic interactions that must be carefully analyzed.

1.2 Need for Dynamic Modeling and Analysis

Accurate dynamic modeling of data center power systems is essential for understanding their impact on grid stability and performance. Traditional steady-state analysis methods are inadequate for capturing fast transient phenomena, control interactions, and oscillatory behavior introduced by converter-based loads.

Electromechanical and electromagnetic dynamic studies have long been used to assess system stability and oscillations in power systems [1]. However, the increasing penetration of converter-dominated loads requires enhanced modeling frameworks that incorporate detailed representations of power electronic devices and their control systems.

Dynamic modeling enables the analysis of key phenomena such as:

- Voltage and frequency response under disturbances
- Oscillatory interactions between grid and load dynamics
- Stability margins and damping characteristics
- Harmonic distortion and power quality issues

In the context of data centers, these analyses are particularly important due to the fast and coordinated behavior of large aggregated loads. Accurate representation of PSU, UPS, and IT load dynamics is necessary to capture system-level responses under realistic operating conditions.

Moreover, the integration of co-located generation resources introduces additional control layers and dynamic interactions. These systems can support grid stability but may also introduce new modes of oscillation if not properly coordinated.

1.3 Objectives and Contributions

The primary objective of this work is to develop a comprehensive dynamic modeling framework for integrated data center power systems and to evaluate their impact on grid performance. The key contributions offered in this report are summarized as follows:

- Development of detailed dynamic models for key data center components, including power supply units (PSU), uninterruptible power supplies (UPS), and IT loads.
- Representation of AI-driven load profiles capturing fast ramping and variability characteristics.
- Integration of data center models with a benchmark transmission system based on the Kundur two-area model [2]
- Inclusion of co-located generation resources such as gas turbines (GT) and battery energy storage systems (BESS) within the modeling framework.
- Validation of component-level models under various operating conditions to ensure accurate dynamic representation.
- Evaluation of system-level performance through case studies involving load variations, source transitions, and disturbance scenarios.
- Analysis of key dynamic phenomena, including voltage and frequency response, oscillatory behavior, damping characteristics, and harmonic distortion.

Through these contributions, the report provides a systematic approach for analyzing the dynamic interactions between large-scale data centers and power systems, with a focus on stability and power quality.

2.0 Computing Architecture of Modern AI Datacenters

2.1 Introduction

Modern AI data centers differ from conventional enterprise and cloud data centers not only in scale, but also in the way computation is organized. Large AI workloads increasingly execute on accelerator-centric systems in which GPUs or other specialized accelerators dominate both computational throughput and electrical demand. These accelerators are interconnected within servers, across racks, and across larger scale-out fabrics, while software schedulers determine which resources become active, where workloads are placed, and when they execute. As a result, the electrical load presented by an AI data center is not determined by installed hardware alone. It emerges from the interaction of workload characteristics, scheduler decisions, network topology, compute architecture, and local power-management mechanisms.

This distinction is important in the context of power-system studies. From the perspective of the electrical system, a data center ultimately appears as a time-varying load. However, the magnitude, temporal variability, spatial concentration, and flexibility of that load are shaped upstream by the computing system. A distributed training job may activate hundreds or thousands of accelerators concurrently, an online inference service may maintain a persistent but variable load, and an offline inference workload may be delayed or preempted with relatively little impact on user-facing performance. Likewise, the same number of nominally available accelerators may not be interchangeable if they differ in model, host configuration, or network locality.

Therefore, the purpose of this chapter is to describe the computing architecture that generates the IT electrical demand subsequently supplied by the data-center electrical system. The focus is on AI workloads, accelerator servers and rack-scale systems, network and logical resource organization, scheduling, and the mapping from computation to rack and facility IT power. The detailed electrical path from rack input through power supplies, power distribution, uninterruptible power systems, medium- and low-voltage distribution, generation, energy storage, and grid interconnection is addressed in the following chapter.

2.2 AI Compute Infrastructure

2.2.1 From Accelerator Servers to Rack-Scale AI Systems

The basic building block of an AI data center is the accelerator-enabled compute node. A typical node combines one or more host CPUs with several GPUs or other accelerators, local high-bandwidth accelerator memory, system memory, network interfaces, local storage, and power-conversion hardware. In current production fleets, these nodes are often heterogeneous. The Alibaba AI workload traces, for example, reveal servers with 1, 2, 8, and 16 accelerators, multiple accelerator models and generations, and a range of CPU capacities [3]. This type of environment is representative of a shared hyperscale infras-

structure that has evolved over time rather than a greenfield deployment built around a single hardware generation.

The architectural trend in new AI systems, however, is toward much tighter integration at rack scale. Rather than treating the rack as a passive enclosure for independent servers, modern systems increasingly treat the rack itself as a computational unit. Rack-scale systems combine accelerator trays, CPU resources, high-bandwidth interconnect, switching, power conversion, and cooling into a coordinated platform. NVIDIA’s GB200 NVL72 is an example of this approach. In such systems, many accelerators participate in a common high-bandwidth communication domain, reducing the distinction between an individual server and the larger rack-level compute fabric.

This trend matters because the power density of the rack has risen rapidly. The Siemens AI data-center blueprint uses GB200 NVL72 racks with a nominal power level of approximately 127 kW [4]. More recent Vera Rubin NVL72 designs move toward roughly 190–230 kW per rack [5]. At these densities, internal power delivery becomes an architectural issue rather than merely a facility-design detail. Low-voltage distribution that was practical for earlier servers can require very large currents when scaled to hundreds of kilowatts within a rack.

The transition toward higher-voltage internal distribution illustrates this shift. For rack demands above roughly 200 kW, maintaining a 48–51 V DC backbone would require currents on the order of several kiloamperes, increasing conductor size, thermal losses, and mechanical complexity. Emerging rack designs therefore move toward an approximately 800 V DC internal backbone, with higher-voltage distribution through the rack and localized step-down conversion close to the compute trays. Conceptually, the internal rack power path becomes:

facility AC $\rightarrow$ high-voltage rack DC $\rightarrow$ localized conversion $\rightarrow$ GPU/CPU point of load.

This architecture reduces distribution current and allows power conversion to be placed closer to the silicon.

The move toward rack-scale AI systems also changes short-timescale power behavior. Workload phases can cause rapid variations in accelerator activity, particularly in large training and inference systems. Newer rack designs therefore incorporate greater local energy storage and closed-loop power smoothing so that the electrical demand seen at the rack input is less volatile than the instantaneous power requested by the silicon. At sufficiently short time scales,

$$P_{\text{silicon}}(t) \neq P_{\text{rack input}}(t), \quad (2.0.1)$$

because the rack-level power subsystem can buffer and shape transient demand. This distinction becomes important when developing dynamic data-center load models for power-system studies.

2.2.2 Network and Logical Resource Architecture

The computing capacity of an AI data center cannot be understood independently of its network. Large training and inference workloads frequently rely on communica-

tion among many accelerators, and the communication cost depends strongly on locality. Within a node or rack-scale system, accelerators may communicate through specialized high-bandwidth interconnects such as NVLink and NVSwitch. Across servers and racks, communication is carried by scale-out fabrics such as InfiniBand or high-performance Ethernet. At still larger scale, these fabrics are embedded in the broader data-center network.

This hierarchy means that physically available accelerators are not always computationally equivalent. A workload may require a set of GPUs that can communicate within a favorable network domain, while another set of otherwise idle accelerators may be unusable because it would require communication across a slower or more congested portion of the network. For this reason, schedulers increasingly incorporate topology into placement decisions.

The Alibaba traces provide a useful operational example. They expose both logical cluster identifiers and access-switch locality identifiers. The observed data do not support a simple physical hierarchy in which every cluster contains a fixed set of access switches and every access switch contains a fixed set of servers. Instead, the evidence is better interpreted as overlapping logical views of the same server fleet. A server may simultaneously belong to a scheduling or administrative domain and to a network-locality domain. This distinction is important because the scheduler’s resource topology need not correspond one-to-one with the physical building, rack, or electrical topology.

Thus, a modern AI data center should be understood as both a physical compute fabric and a set of software-visible resource domains. The physical fabric contains accelerator hardware and network links. The logical fabric contains clusters, resource pools, locality domains, quotas, priorities, and other abstractions used to control access to that hardware. These two views interact continuously through the scheduler.

2.3 AI Workloads and Resource Management

AI data centers commonly serve several classes of workload with very different computational and temporal characteristics. Model training often involves large distributed jobs that require many accelerators concurrently and can be highly sensitive to communication locality. Online inference tends to emphasize latency, availability, and continuous service. Offline inference is often throughput-oriented and can be more tolerant of delay or interruption. Development and experimentation workloads are generally smaller and more irregular. Production traces such as Alibaba’s show that these workload classes coexist within the same shared infrastructure rather than operating in isolation.

The scheduler translates these workload requirements into physical activity. A job typically arrives with a set of constraints, including requested accelerator model, accelerator count, CPU capacity, memory demand, and possibly topology or locality requirements. The scheduler must identify a feasible placement that satisfies these requirements while respecting tenant quotas, resource reservations, and service priorities. In heterogeneous fleets, this can partition the available hardware into several partially substitutable capacity pools. An idle accelerator of the wrong model, or an accelerator attached to a host lacking sufficient CPU capacity, may not be useful to the arriving job.

Placement also creates fragmentation. A small job assigned to a large multi-GPU

server may leave residual resources that cannot satisfy a later whole-node request. Similarly, sufficient free accelerators may exist globally while being distributed across network domains that do not satisfy the locality requirements of a communication-intensive workload. As a consequence, installed accelerator capacity and usable accelerator capacity are not the same quantity.

Priority and preemption introduce another layer of control. Large shared infrastructures often divide workloads into guaranteed or high-priority classes and lower-priority, best-effort, or spot classes. Lower-priority workloads can be suspended, terminated, or migrated when higher-priority work arrives. These mechanisms were introduced primarily to improve utilization and service guarantees, but from a power-system perspective they also create different degrees of computational flexibility. A persistent online inference service is relatively inflexible, while a low-priority batch workload may be delayed or preempted with much less impact.

Scheduling therefore acts as the control layer between software demand and physical electrical activity. In a conventional scheduler, the primary objectives are typically performance, utilization, fairness, and service quality. In a grid-aware environment, the same control layer could additionally incorporate power, energy, or external grid constraints. The architecture does not imply that such coordination is universally deployed today, but it establishes the mechanism through which computational flexibility could be translated into electrical flexibility.

2.4 From AI Computation to IT Electrical Demand

The central relationship for power-system analysis is the causal chain from workload activity to the aggregate electrical demand of the IT equipment. Figure 2.2 summarizes this relationship.

The Alibaba trace provides an empirical illustration of the resulting temporal behavior. Figure 2.1 shows a seven-day trace-derived GPU workload activity profile together with an illustrative H20-equivalent GPU-only electrical scenario. The upper panel reports the workload activity directly from the trace, while the lower panel maps that activity to GPU electrical demand using a 400 W peak-power assumption and an assumed 90 W idle power per GPU. The resulting MW profile is intended to illustrate the temporal shape and priority composition of workload-driven GPU demand; it is not a measurement of Alibaba’s server-input, facility, or grid power.

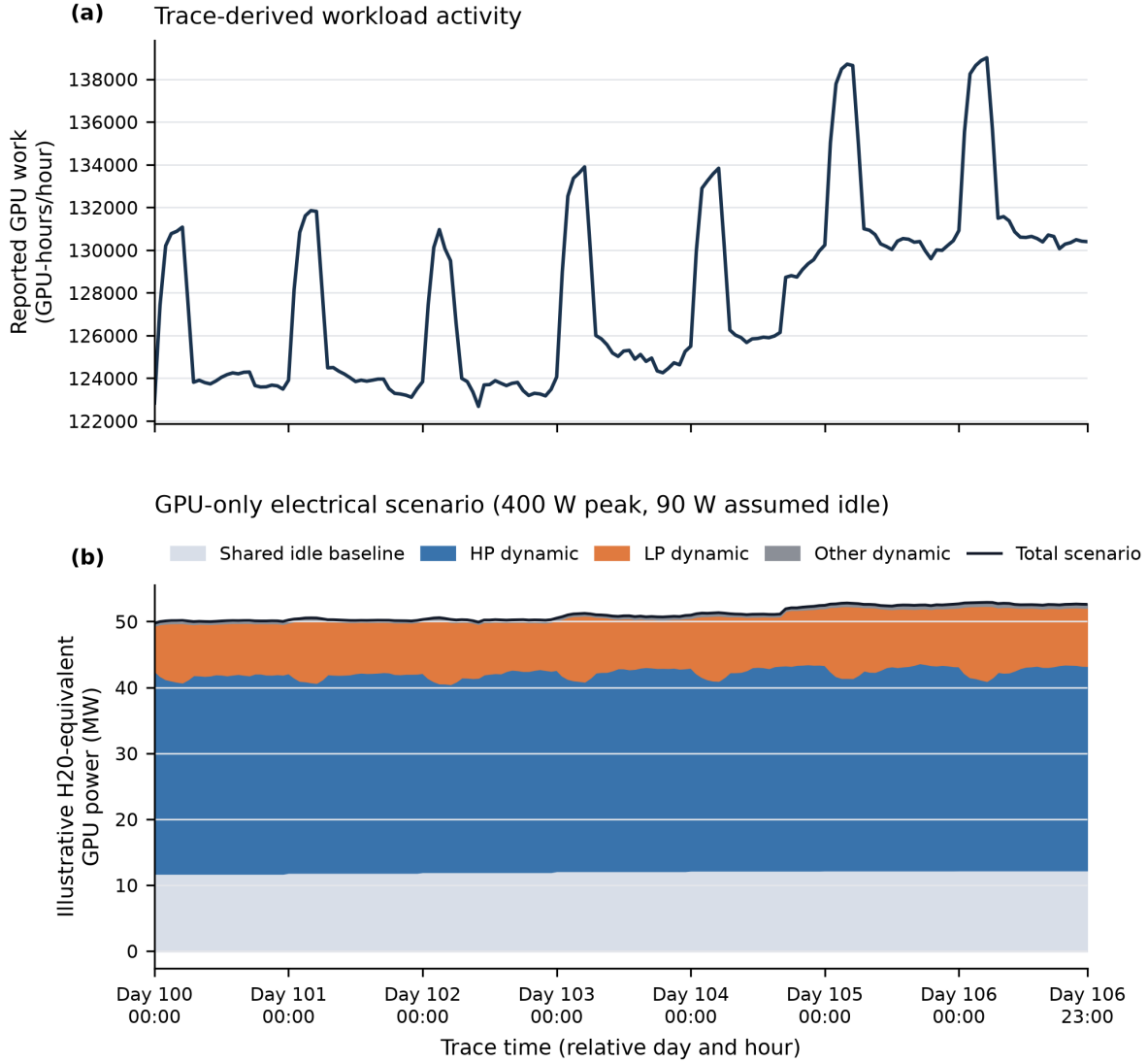


Figure 2.1: Seven-day trace-derived GPU workload activity and illustrative H2O-equivalent GPU-only electrical scenario. Panel (a) shows the reported aggregate GPU workload activity from the Alibaba trace. Panel (b) translates a capacity-consistent representation of the same workload into illustrative GPU electrical power using a 400 W peak-power assumption and an assumed 90 W idle power per GPU. The shared idle baseline is kept separate from workload-priority attribution, while the dynamic component is decomposed into high-priority (HP), low-priority (LP), and other activity. The MW values are illustrative and represent GPU electrical demand only; they do not represent measured Alibaba server-input, total IT, facility, or grid power.

Workloads generate resource requests and transition through execution phases. The scheduler selects physical resources and determines placement, priority, preemption, and migration. Those decisions determine which GPUs, CPUs, memory devices, and network interfaces are active. Device-level activity produces silicon power demand, which is then modified by local power conversion, energy storage, and rack-level power smoothing before

appearing as rack input power. Aggregate rack demand becomes the IT load presented to the downstream electrical architecture.

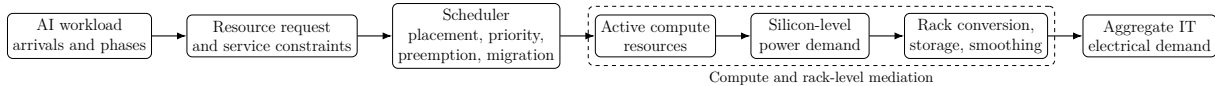


Figure 2.2: Causal chain from AI workload activity to IT electrical demand. Workload behavior and scheduler decisions determine the active compute resources, while rack-level power conversion and energy storage can mediate the relationship between silicon power demand and the electrical demand observed by the facility.

The resulting electrical demand has three characteristics of particular importance to the grid. The first is magnitude. Aggregate IT power depends on the number and type of active accelerators, the degree of CPU and network activity, the utilization level of those devices, and the configuration of the servers or rack-scale systems in which they operate. Installed capacity therefore provides only an upper bound and should not be equated with actual demand.

The second characteristic is temporal behavior. Job arrivals, job completions, workload phase transitions, preemption, migration, and power-management actions can all create changes in load. Some changes are gradual, while others can be highly synchronized. For example, a large distributed job may activate many accelerators at nearly the same time. At sub-second time scales, rack-level energy storage and active power smoothing may reduce the magnitude of the transient seen by the facility. At longer time scales, scheduler behavior becomes increasingly important because jobs may be delayed, consolidated, migrated, or preempted.

The third characteristic is spatial concentration. Scheduling determines where within the compute fabric the load occurs. A topology-aware scheduler may intentionally place many accelerators within the same locality domain, causing substantial power concentration in a particular rack or group of racks even if the total facility load is unchanged. This matters for electrical distribution because local feeders, busways, cooling systems, and other infrastructure respond to the spatial distribution of IT demand as well as to the facility total.

These mechanisms operate over multiple time scales. Sub-second behavior is dominated by accelerator activity, local conversion, and rack-level power smoothing. Seconds-to-minutes behavior can reflect power management, preemption, and changes in workload concurrency. Minutes-to-hours behavior is increasingly influenced by scheduling, queueing, and job deferral. Over still longer intervals, workload relocation, maintenance, and capacity planning may alter the spatial or temporal distribution of demand. The precise boundaries between these time scales depend on workload and system design, but the hierarchy is useful when determining which computing mechanisms should be represented in a power-system model.

The same architecture also determines which portions of the IT load may be flexible. Computational flexibility can arise from workload deferral, preemption, migration, concurrency control, and power capping. However, computational flexibility should not be equated with unconstrained electrical controllability. Service deadlines, checkpoint

overhead, data locality, communication topology, latency requirements, and hardware restrictions limit the degree to which a workload can be shifted or curtailed. Grid-oriented models should therefore distinguish between critical and potentially flexible compute demand rather than treating the entire data center as a uniform controllable load.

2.5 Interface to the Data-Center Electrical System

The computing architecture described in this chapter produces the IT electrical load that must be supplied by the downstream data-center power system. At an abstract level, that load may be written as

$$P_{IT}(t) = f(W(t), S(t), H, N, C(t)), \quad (2.0.2)$$

where $W(t)$ represents workload activity, $S(t)$ represents scheduler and placement decisions, H represents the installed compute hardware, N represents the relevant network and locality structure, and $C(t)$ represents device- and rack-level power-control behavior. This expression is not intended as a detailed analytical model, but rather as a statement of the principal mechanisms that determine IT demand.

The electrical architecture begins at the point where this IT load is supplied and conditioned. The following chapter describes the power supplies, power distribution units, uninterruptible power systems, low- and medium-voltage distribution, co-located generation, energy storage, and grid interconnection that deliver power to the computing system. In this sense, the two chapters form a continuous modeling chain: computing architecture determines the characteristics of the IT load, while electrical architecture determines how that load interacts with the facility power system and the external grid.

2.6 Case Studies

2.6.1 Alibaba ASI: A Heterogeneous Production AI Infrastructure

The Alibaba AI infrastructure provides a useful production example of a large shared accelerator fleet. The available traces indicate a heterogeneous environment with multiple accelerator models, several accelerator counts per server, varying CPU capacities, and recurring server resource configurations. The fleet changes over time through server additions, removals, and changes in reported resource attributes. This makes it representative of an operational hyperscale system that has evolved incrementally rather than a greenfield facility built around a single hardware generation.

The workload mix is similarly heterogeneous. Training, online inference, offline inference, and development workloads coexist within the same shared infrastructure. Jobs differ in duration, accelerator count, priority, and resource requirements. The scheduler must therefore balance hardware compatibility, CPU/GPU coupling, locality, fragmentation, and service priority. The paper associated with the trace further describes topology-aware placement, consolidation and migration, and the use of lower-priority or spot-like capacity to increase utilization [6].

The logical organization exposed by the trace is also informative. Servers are associated with cluster identifiers and, where available, access-switch locality identifiers. These

should be interpreted conservatively as scheduler-visible resource and network-locality domains rather than as a complete reconstruction of the physical data-center topology. The observed data do not establish how many buildings, racks, power domains, or physical access switches are represented, and some access-switch identifiers overlap more than one released cluster. For power-system modeling, the main lesson is that workload placement is constrained by several overlapping resource dimensions that are not captured by accelerator count alone.

2.6.2 Siemens/NVIDIA: A High-Density Greenfield Reference Architecture

The Siemens 100 MW AI data-center blueprint provides a complementary view. Whereas the Alibaba trace reveals the operational structure of a heterogeneous production fleet, the Siemens design illustrates a standardized greenfield architecture built around rack-scale AI systems. The blueprint uses NVIDIA GB200 NVL72 systems with a nominal rack demand of approximately 127 kW and organizes them into modular high-density infrastructure blocks with dedicated liquid cooling, network support, and redundant electrical distribution.

From the perspective of this chapter, the importance of the Siemens design is not the detailed electrical topology, which is treated in the following chapter, but the physical instantiation of rack-scale compute. The blueprint illustrates the consequences of treating the rack as a computational system: compute density, cooling requirements, network organization, and power delivery must be coordinated as part of a common architecture. Direct-to-chip liquid cooling is not an optional enhancement at these densities but an integral part of the compute platform.

The evolution from GB200-class systems toward Vera Rubin-class NVL72 systems pushes this trend further. Rack demand approaching 200 kW or more motivates a transition toward higher-voltage DC distribution inside the rack, localized conversion near the compute trays, larger rack-level energy storage, and more active power smoothing. Thus, the architecture of the computing rack itself increasingly contains mechanisms that were once considered only at the facility level. This evolution is particularly relevant for grid studies because it means that fast computational power transients may be partially shaped before they reach the external data-center electrical system.

Taken together, the Alibaba and Siemens/NVIDIA cases illustrate two complementary forms of modern AI infrastructure. The first is a heterogeneous shared production fleet shaped by years of operational evolution and complex scheduling. The second is a highly standardized, high-density rack-scale architecture designed for new AI facilities. Both are important for power-system analysis. The former emphasizes workload diversity, scheduling, and operational flexibility, while the latter emphasizes the electrical and thermal consequences of rapidly increasing rack-scale compute density.

2.7 Summary

Modern AI data centers should be viewed as software-controlled, accelerator-centric electrical loads. Their demand is generated by a chain that begins with workload arrivals and service requirements, passes through scheduling and topology-aware placement, acti-

vates specific compute and network resources, and is then shaped by device- and rack-level power-management mechanisms before appearing as facility IT load.

This architecture has direct implications for power-system modeling. Installed accelerator capacity is not equivalent to active electrical demand, network topology influences the spatial distribution of load, scheduler decisions shape temporal behavior, and newer rack-scale systems can modify short-timescale transients through local power conversion and energy storage. At the same time, workload priority, preemption, and deferral create potential flexibility that may be relevant to future grid-interactive operation.

The next chapter begins at this compute-to-power boundary and develops the electrical architecture that supplies, conditions, distributes, and supports the IT load described here.

3.0 Data Center Electrical Architecture

3.1 Introduction and Overview

Modern data centers are characterized by high power density, precise reliability requirements, and extensive use of power electronic interfaces. The electrical architecture is therefore designed to ensure uninterrupted power delivery, high power quality, and the ability to accommodate rapid and large variations in load demand [7,8].

Unlike conventional load centers, data center power systems consist of multiple interconnected subsystems, including utility supply, local generation, energy storage, power conditioning units, and distributed IT loads. These subsystems are typically organized hierarchically, with successive voltage levels and multiple stages of power conversion. Such architectures enable isolation of sensitive loads from grid disturbances while allowing flexible operation and integration with emerging large-scale loads [9–11].

The electrical architecture considered in this work is illustrated in Fig. 5.3. The system represents a grid-connected data center with co-located generation and energy storage, designed to capture both steady-state operation and dynamic interactions with the grid.

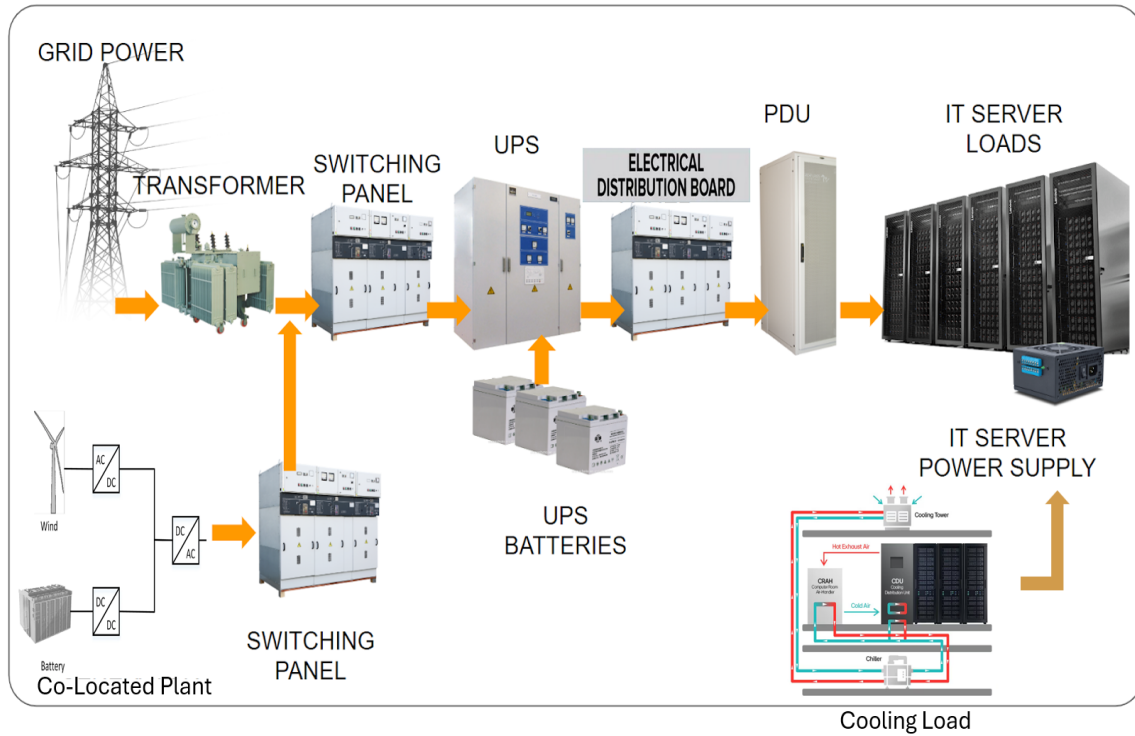


Figure 3.1: Conventional Grid Connected Data Center Power Distribution System.

3.2 Electrical Architecture and Power Flow Description

Data center electrical architectures are generally structured across multiple voltage levels, including high-voltage (HV), medium-voltage (MV), and low-voltage (LV) stages. Power is supplied from the transmission grid at HV levels, stepped down to MV for distribution, and further converted to LV for utilization within IT and auxiliary systems.

In the system considered in this work, bulk power is supplied from a 230 kV utility grid, which is stepped down to a 13.8 kV MV bus through transmission and transformer stages. This MV level serves as the primary point of common coupling (PCC) for the data center.

A co-located generation plant is connected at the MV level, consisting of a gas turbine (GT) generator and a battery energy storage system (BESS). The integration of such local resources enhances system resiliency, enables partial islanded operation, and provides fast dynamic support during disturbances [12, 13].

The MV bus is further stepped down to a 480 V LV distribution level, supplying both the IT infrastructure and auxiliary loads, such as cooling systems. At this level, power is conditioned by an uninterruptible power supply (UPS) before being delivered to IT equipment. Power flow in the system follows a hierarchical path:

- Utility grid and co-located plant supply power to the MV bus,
- MV bus feeds LV distribution through transformers,
- UPS system conditions and regulates power,
- PDUs distribute power to racks,
- PSUs convert power to low-voltage DC for IT loads.

This layered architecture enables both electrical isolation and controlled interaction between different subsystems.

3.3 Architecture Representation and Modeling Approach

To enable system-level dynamic analysis, the data center electrical architecture is represented using a modular modeling framework. Each subsystem is modeled independently and then interconnected through electrical interfaces. This approach is consistent with modular data center modeling, in which individual components are modeled, preserving overall system dynamics.

The architecture is therefore categorized into the following major subsystems:

- Utility grid and transmission system,
- Co-located generation and energy storage,
- Uninterruptible power system (UPS),
- Power distribution network (PDU),

- End-use loads (PSU and IT load),
- Auxiliary loads (cooling systems).

Each subsystem is modeled based on its electrical characteristics, control behavior, and interaction with adjacent components. This modular representation allows scalable modeling and facilitates integration with higher-level grid studies.

3.4 System Topology and Operating Modes

3.4.1 System Topology

The electrical system follows a AC/DC topology, with AC power used for bulk transmission and distribution, while DC subsystems are employed at the UPS and IT load interfaces.

At the LV level, the UPS employs a double-conversion topology along with a bypass and battery mode, comprising a rectifier, a battery, a DC link, and an inverter system. The DC link, typically operating at 800 V, interfaces with the battery system and serves as an energy buffer.

Downstream of the UPS, power is distributed in AC form and converted locally at the rack level through PSUs to a regulated low DC voltage required by IT equipment.

3.4.2 Operating Modes

To ensure an uninterrupted power supply at any grid conditions, the system operates in multiple modes depending on grid conditions:

- **Eco (Bypass) Mode:** Under stable conditions, the UPS operates in bypass mode to improve efficiency, allowing AC power to feed the load directly [14].
- **Double Conversion Mode:** Power is supplied from the grid and a co-located plant. The UPS performs continuous AC–DC–AC conversion, ensuring voltage regulation and isolation from upstream disturbances.
- **Battery-Supported Mode:** During grid disturbances such as voltage sags or outages, the DC link is supported by the battery system. The inverter continues to supply regulated AC power to the load, ensuring uninterrupted operation.

These operating modes enable the data center to maintain high reliability while optimizing efficiency and dynamic performance.

3.5 Component-Level Description

The overall system comprises several key components, each contributing to the functionality and dynamic behavior of the data center electrical architecture.

3.5.1 Utility Grid and Transmission System

The utility grid serves as the primary power source using the Kundur two-area system to capture inter-area dynamics and system-level interactions [1,2]. Transmission lines and transformers connect the grid to the data center’s MV bus.

3.5.2 Co-located Generation and Energy Storage

To enhance system reliability, operational flexibility, and dynamic performance, modern data centers increasingly incorporate co-located generation and energy storage systems. These resources provide local power support at the medium-voltage level, reducing dependency on the utility grid and enabling improved response to disturbances.

The integration of co-located generation allows the data center to operate under both grid-connected and partially islanded conditions. In particular, fast-acting energy storage systems complement conventional generation by providing rapid transient support, while mechanical generation sources contribute to sustained power delivery. Together, these resources play a critical role in maintaining system stability under dynamic loading conditions associated with data center operation.

3.5.2.1 Gas Turbine Generator

The gas turbine generator provides local generation support at the MV level. It contributes to a steady-state power supply and enhances system stability during disturbances.

3.5.2.2 Battery Energy Storage System

The BESS provides fast dynamic support, including frequency regulation, transient stabilization, and load balancing. Its fast response complements the slower dynamics of the gas turbine.

3.5.3 Uninterruptible Power Supply (UPS)

The UPS system ensures continuous power delivery to IT loads. The double-conversion topology isolates the load from grid disturbances and maintains voltage and frequency stability. The DC link serves as an energy buffer and interfaces with the battery system [15].

3.5.4 Power Distribution Unit (PDU)

The PDU distributes conditioned AC power from the UPS to multiple downstream feeders. It provides protection and isolation functions.

3.5.5 Power Supply Unit (PSU)

PSUs convert AC power into a regulated DC voltage for IT equipment. These units include power-factor correction and DC–DC conversion stages, which contribute to non-linear load characteristics and harmonic interactions.

3.5.6 IT Load

IT loads exhibit highly dynamic and nonlinear behavior driven by computational workload and server utilization. These loads introduce rapid power variations and can significantly impact system stability and power quality [10, 16].

3.5.7 Cooling System Loads

Cooling systems are essential for maintaining the thermal operating limits of IT equipment by removing the heat generated during computation. These systems typically consist of chillers, cooling towers, pumps, and air handling units, which together regulate temperature and airflow within the data center [8, 17].

From an electrical perspective, cooling systems account for a significant portion of the total data center load and are directly supplied by the AC distribution network. Their power consumption is primarily driven by the thermal load, which is closely related to the IT power demand [7].

In operation, cooling systems respond to changes in IT load by adjusting cooling capacity through control mechanisms such as variable speed drives and temperature regulation loops. As a result, cooling load variations are inherently coupled with IT load dynamics, introducing both steady-state and transient power variations that must be considered in system-level analysis.

3.6 Discussion and Summary

The electrical architecture presented in this work reflects the structure of modern large-scale data centers, combining conventional AC distribution with power electronic subsystems. The modular representation adopted enables scalable modeling and facilitates integration with power system studies. Furthermore, the inclusion of co-located generation and energy storage highlights the growing trend toward flexible, grid-interactive data center operations. The next chapter discusses the data center modeling approach with grid and co-located generation system modeling, including the protection system (fault ride through).

4.0 Data Center Modeling Approach with Grid and Co-Located Generation System Modeling with Protection System (Fault Ride Through) Included

4.1 Dynamic Modeling Framework for Data Center Electrical System

This chapter focuses on the dynamic modeling of key data center electrical subsystems to accurately represent their behavior under varying operating conditions and their interaction with the grid. The modeling is developed with an emphasis on capturing the electrical and control dynamics of individual components that influence the overall system response.

The objective is to establish detailed and physically representative models of the data center subsystems, particularly those interfaced through power electronic converters and control loops. These components govern the dynamic characteristics of the system, including voltage regulation, power flow behavior, and transient response under disturbances [9, 10, 18].

The modeling is carried out in a modular manner, where each subsystem is developed independently based on its role within the power delivery architecture. This approach allows clear representation of component-level dynamics while maintaining flexibility for system-level integration in subsequent studies.

The modeling sequence follows the power-delivery path within the data center. It begins with the Power Distribution Unit (PDU), which represents the intermediate AC distribution and protection interface. This is followed by the Power Supply Unit (PSU), including the power factor correction (PFC) stage and DC–DC conversion, where converter dynamics and control behavior are modeled in detail.

Subsequently, the IT load is modeled to capture its dynamic and nonlinear characteristics driven by computational demand. The Uninterruptible Power Supply (UPS) is then developed, including its topology, operating modes, filtering elements, and control architecture of both rectifier and inverter stages. The modeling further includes a simplified representation of cooling system loads, with their contributions modeled as functions of IT load demand, allowing the combined IT and cooling effects to be represented through an aggregated load consumption.

In addition, a brief representation of co-located resources, such as gas turbine generation and battery energy storage systems, is included to reflect their interaction with the data center electrical system. The chapter also incorporates fault-ride-through (FRT) modeling, ensuring that the system's response to grid disturbances is appropriately represented.

The interconnection of these subsystems and their integration with the external power system are addressed in a subsequent chapter.

4.2 Modeling of Power Distribution Unit (PDU)

The Power Distribution Unit (PDU) serves as the intermediate distribution stage between the centralized Uninterruptible Power Supply (UPS) and the downstream rack-level Power Supply Units (PSUs) within the data center electrical architecture. In the data center model, the PDU distributes regulated 480 V AC power from the UPS output to multiple feeders supplying IT racks and auxiliary loads, such as cooling equipment, while providing protection and selective isolation capability.

4.2.1 PDU Functional Representation

In this study, the PDU is modeled as a low-voltage AC distribution subsystem rather than a power-electronic conversion stage. Its primary functions include:

- Distribution of 480 V AC power to multiple downstream branches,
- Incorporation of circuit breakers (CBs) for feeder-level protection and isolation,
- Aggregation of downstream rack and auxiliary load demands.

4.2.2 Electrical Modeling Approach

Each PDU feeder is modeled as an independent branch supplying an aggregated load representing one or more IT racks. The total active and reactive power drawn at the PDU output is given by:

$$P_{\text{PDU}} = \sum_{i=1}^{N_f} P_i, \quad Q_{\text{PDU}} = \sum_{i=1}^{N_f} Q_i, \quad (4.0.1)$$

where N_f is the number of outgoing feeders, and P_i , Q_i denote the active and reactive power demands of the i -th feeder.

The PDU input voltage is assumed to be regulated by the upstream UPS. Therefore, voltage dynamics at the PDU terminals are primarily influenced by upstream grid disturbances and UPS control actions rather than internal PDU behavior. Line losses and the PDU's internal impedance are neglected.

4.2.3 Protection and Switching Representation

Each feeder branch includes a circuit breaker modeled as an ideal controllable switch. The breaker logic enables:

- Selective isolation of individual IT racks or load groups,
- Disconnection under abnormal operating conditions,
- Simulation of maintenance and fault scenarios.

Overall, the proposed PDU modeling framework provides a physically consistent representation suitable for analyzing data center dynamics, grid interactions, and protection scenarios.

4.3 Modeling of Power Supply Unit (PSU)

The Power Supply Unit (PSU) converts AC input power into a regulated DC output required by IT loads. A typical PSU consists of a front-end Power Factor Correction (PFC) stage followed by a DC–DC conversion stage. The PFC stage rectifies AC power to DC, ensuring high power quality by regulating the DC-link voltage and shaping the input current, while the DC–DC stage provides voltage step-down, electrical isolation, and maintains the desired output voltage by DC voltage regulation.

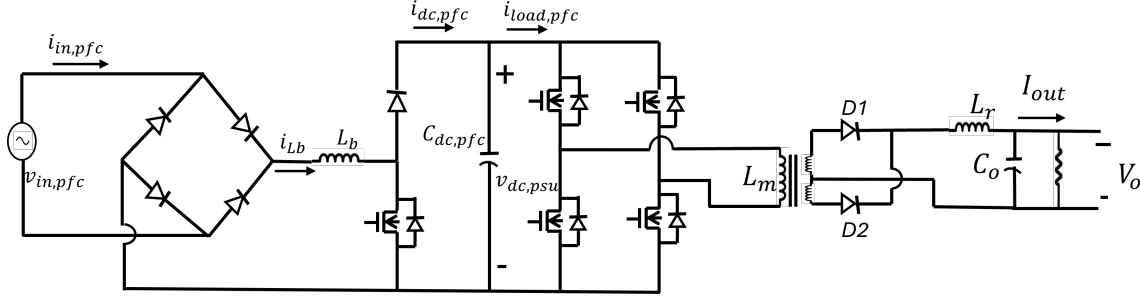


Figure 4.1: PSU Model.

As illustrated in Fig. 4.1, the PFC stage interfaces with the AC supply, performs rectification, and regulates the DC-link voltage while shaping the input current to achieve near-unity power factor. The regulated DC-link voltage is then processed by the DC–DC isolated converter, which provides voltage scaling, galvanic isolation, and a stable low-voltage DC output suitable for rack-level IT equipment.

The modeling of each stage, including topology, filter design, control architecture, and switching, is presented in detail in the following subsections.

4.3.1 Modeling of Power Factor Correction (PFC) Converter

The PFC converter forms the front-end stage of the PSU and is responsible for converting the input AC power into a regulated DC-link voltage. In addition to voltage conversion, the PFC stage ensures near-unity power factor and low total harmonic distortion (THD) by shaping the input current to follow the supply voltage waveform.

This functionality plays a critical role in improving power quality and minimizing harmonic injection into the grid, while providing a stable DC-link voltage for the downstream DC–DC conversion stage.

4.3.1.1 Topology

The implemented PFC converter consists of a single-phase diode-bridge rectifier followed by a controlled-switching stage for current shaping and DC-link voltage regulation. The input AC voltage is first rectified using a diode bridge to produce a unipolar DC waveform.

This rectified voltage is then processed by an inductor-based switching stage comprising an inductor L , a MOSFET switch, and a diode. The switching action regulates the DC-link voltage while shaping the input current to follow the supply voltage waveform, thereby improving power factor and reducing harmonic distortion. The rectified voltage

can be expressed as:

$$v_{\text{dc,rect}}(t) = |v_s(t)|. \quad (4.0.2)$$

where:

- $v_{\text{dc,rect}}(t)$ is the rectified DC voltage,
- $v_s(t)$ is the input AC voltage,
- t represents time.

Under steady-state conditions, the DC-link voltage is related to the duty ratio of the switching device as:

$$v_{\text{dc}} = \frac{v_{\text{dc,rect}}}{1 - D}. \quad (4.0.3)$$

where:

- v_{dc} is the DC-link voltage,
- $v_{\text{dc,rect}}$ is the rectified input voltage,
- D is the duty ratio of the switching device.

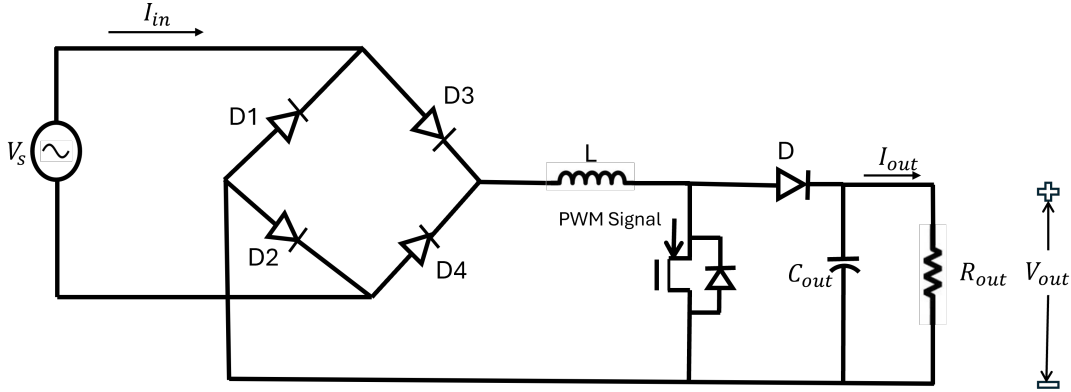


Figure 4.2: PFC Converter Model.

As shown in Fig. 4.2, the implemented model includes the rectifier, inductor-based switching stage, DC-link capacitor, and measurement blocks. The switching stage interfaces the rectified AC input with the DC-link, ensuring both voltage regulation and input current shaping.

4.3.1.2 Filter Design

The PFC converter employs an input inductive filter and an output DC-link capacitive filter, forming an effective L - C filtering structure. The inductor L is connected in series with the rectified input, while the capacitor C_{dc} is placed across the DC-link.

This filtering stage shapes the input current, suppresses high-frequency switching harmonics, and stabilizes the DC-link voltage, thereby ensuring improved power quality and reliable operation of the downstream conversion stages.

Inductor Dynamics: The inductor governs input current shaping and ensures continuous conduction mode (CCM) operation. Its voltage-current relationship is given by:

$$v_L = L \frac{di_L}{dt}. \quad (4.0.4)$$

where:

- v_L is the inductor voltage,
- L is the inductance,
- i_L is the inductor current.

During switching operation:

$$v_L = L \frac{di_L}{dt} = \begin{cases} v_{\text{dc,rect}}, & (\text{switch ON}) \\ v_{\text{dc,rect}} - v_{\text{dc}}, & (\text{switch OFF}) \end{cases} \quad (4.0.5)$$

where:

- v_{dc} is the DC-link voltage,
- $v_{\text{dc,rect}}$ is the rectified input voltage.

This relationship highlights that the inductor controls the rate of change of current, enabling current shaping and reduction of harmonic distortion.

DC-Link Capacitor: The DC-link capacitor smooths the output voltage and maintains energy balance:

$$C_{\text{dc}} \frac{dV_{\text{dc}}}{dt} = i_L - i_o. \quad (4.0.6)$$

where:

- C_{dc} is the DC-link capacitance,
- V_{dc} is the DC-link voltage,
- i_L is the inductor current,
- i_o is the load current.

A sufficiently large capacitance reduces voltage ripple and provides energy buffering during transient conditions.

The selection of L and C_{dc} is based on allowable current and voltage ripple limits, as well as the desired dynamic response of the converter.

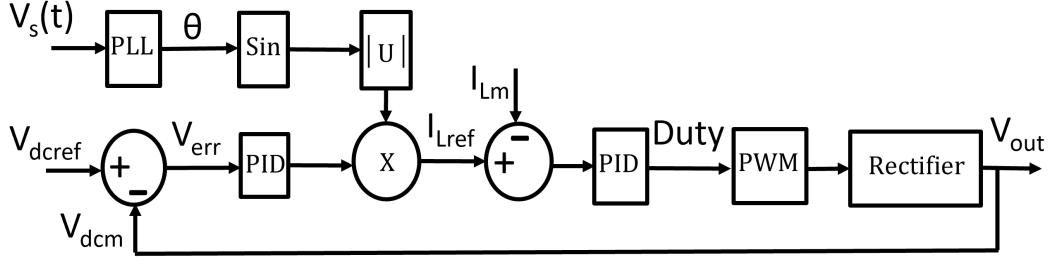


Figure 4.3: Control architecture of the PFC converter.

4.3.1.3 Control Architecture

The PFC converter employs a cascaded dual-loop control structure consisting of an outer voltage loop and an inner current loop. This structure regulates the DC-link voltage while ensuring that the input current follows the input AC voltage waveform, thereby achieving near-unity power factor operation.

PLL and Reference Current Generation: A phase-locked loop (PLL) is used to synchronize with the input AC voltage $v_s(t)$ and extract its phase angle θ . This phase information is used to generate a sinusoidal signal that is in phase with the input voltage. The outer voltage control loop determines the amplitude of the reference current, which is multiplied by a sinusoidal signal to produce the reference inductor current. This ensures that the input current remains sinusoidal and aligned with the input voltage.

Outer Voltage Loop: The outer control loop regulates the DC-link voltage by comparing the measured DC-link voltage with its reference value. The resulting error is processed through a proportional–integral–derivative (PID) controller to generate the reference inductor current. This reference current represents the required input current magnitude needed to maintain the DC-link voltage under varying load conditions and determines the amount of active power drawn from the input source.

Inner Current Control Loop: The inner control loop regulates the inductor current to track the reference current generated by the outer loop. The difference between the reference and measured inductor currents is processed through a PID controller to generate a control signal that determines the PWM duty cycle. By continuously adjusting the duty ratio, the controller ensures fast and accurate current tracking.

In continuous conduction mode (CCM), the inductor current i_L closely represents the input current. Therefore, regulating the inductor current effectively achieves input current shaping.

Power Factor Operation: Since the reference inductor current is proportional to the instantaneous input voltage waveform, the input current remains in phase with the input voltage. This results in near-unity power factor operation with reduced reactive power and harmonic distortion.

Control Operation: Overall, the cascaded control structure enables the DC-link voltage to be regulated through the outer loop, while the inner loop enforces input current shaping. Variations in load or input conditions are compensated by dynamically adjusting the duty ratio, ensuring stable operation, improved power quality, and fast dynamic response.

4.3.1.4 Switching and PWM Implementation

The control signal generated by the inner current control loop is used to determine the switching duty ratio of the converter. The duty ratio is expressed as:

$$D(t) = \frac{v^*(t)}{v_{dc}}, \quad (4.0.7)$$

where:

- $D(t)$ is the duty ratio applied to the switching device,
- $v^*(t)$ is the control voltage generated by the inner current controller,
- v_{dc} is the DC-link voltage.

This duty ratio represents a normalized control signal that determines the ON and OFF durations of the switching device.

The duty command is applied to a PWM generator operating at a fixed switching frequency of 10 kHz, which produces the gate signals for the MOSFET.

Through this switching action, the converter modulates the inductor voltage and regulates the energy transfer from the AC side to the DC-link. As a result, the input current is shaped to follow the input voltage waveform, ensuring near-unity power factor and reduced harmonic distortion.

Variations in load or input conditions are compensated by dynamically adjusting the duty ratio, allowing the converter to maintain a stable DC-link voltage.

The selected switching frequency provides a trade-off between switching losses and passive component size, enabling efficient operation while maintaining fast dynamic response. The key parameters used in modeling the PFC converter are summarized in Table 4.1.

Table 4.1: Key parameters of the PFC converter

Component / Parameter	Symbol	Value	Description / Remarks
Input AC Voltage	v_s	480 V	Input AC supply to the PFC stage
DC-link Voltage	V_{dc}	480 V	Regulated DC-link output
Boost Inductor	L	0.002 H	Input current shaping and energy transfer
DC-link Capacitor	C_{dc}	0.045 F	DC-link voltage stabilization and energy buffering
Switching Frequency	f_s	10 kHz	PWM switching frequency

4.3.1.5 Summary

The developed PFC converter model accurately captures the electrical behavior and control dynamics of the front-end AC–DC conversion stage of the PSU. It ensures a regulated DC-link voltage while shaping the input current to follow the supply voltage waveform, thereby improving power quality and reducing harmonic injection into the grid under varying operating conditions. The key characteristics are summarized as follows:

- Regulated DC-link voltage,
- Sinusoidal input current,
- Unity power factor,
- Low THD,
- Fast dynamic response.

4.3.2 Modeling of DC-DC Isolated Converter

The DC–DC isolated converter forms the second stage of the PSU and processes the regulated DC-link voltage from the PFC stage to produce a low-voltage DC output (48 V) suitable for rack-level IT equipment. In addition to voltage step-down, this stage provides isolation, tight voltage regulation, and fast transient response, ensuring reliable operation under dynamic IT load conditions.

4.3.2.1 Topology

The DC–DC stage is implemented as a full-bridge isolated converter employing a high-frequency transformer for voltage scaling and isolation. This stage processes the regulated DC-link voltage from the PFC converter and converts it to a low-voltage DC output suitable for rack-level IT equipment.

The primary-side full-bridge switching network generates a high-frequency alternating voltage, which is applied to the transformer. The transformer provides isolation and scales the voltage according to the turns ratio. The secondary-side voltage is then rectified and filtered to produce a regulated DC output. The primary-side voltage can be expressed as:

$$v_p(t) = d(t) V_{\text{in}}, \quad (4.0.8)$$

where:

- $v_p(t)$ is the primary-side voltage of the transformer,
- $d(t)$ is the duty ratio generated by the control loop,
- V_{in} is the input DC-link voltage from the PFC stage,
- t represents time.

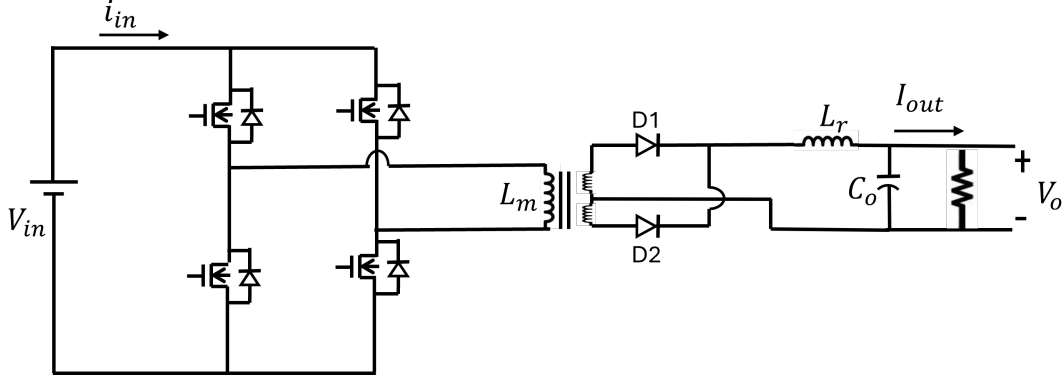


Figure 4.4: DC-DC Isolated Converter Model.

The transformer voltage relationship is given by:

$$v_s(t) = \frac{1}{n} v_p(t), \quad (4.0.9)$$

where:

- $v_s(t)$ is the secondary-side voltage of the transformer,
- $v_p(t)$ is the primary-side voltage,
- $n = \frac{N_p}{N_s}$ is the transformer turns ratio,
- N_p and N_s are the number of primary and secondary turns, respectively,
- t represents time.

After rectification and filtering, the output voltage is:

$$V_{\text{out}} = \frac{d V_{\text{in}}}{n} - I_L R_{\text{on}}, \quad (4.0.10)$$

where:

- V_{out} is the DC output voltage,
- d is the steady-state duty ratio,
- V_{in} is the input DC-link voltage,
- n is the transformer turns ratio,
- I_L is the output inductor current,
- R_{on} represents equivalent conduction resistance of switches and transformer windings, which accounts for conduction losses in switches and transformer windings.

The implemented model, shown in Fig. 4.4, includes the primary-side full-bridge converter, high-frequency transformer, secondary-side rectification stage, output filter, and voltage control loop. The converter interfaces with the PFC stage at the DC-link, while regulating the output to the required low-voltage level for IT loads.

The use of high-frequency switching enables reduced transformer size and improved dynamic performance, while the isolation ensures safe and reliable operation in data center power systems.

4.3.2.2 Filter Design

The DC–DC converter employs a secondary-side L – C output filter to smooth the rectified high-frequency waveform and provide a stable DC output. The inductor is connected in series with the rectified secondary output, while the capacitor is placed across the load, forming a low-pass filter.

This filtering stage attenuates switching ripple, regulates output current, and ensures a stable voltage supply to the IT load.

Inductor Dynamics: The output inductor governs current smoothing and limits the rate of current change. Its dynamics are given by:

$$v_L = L \frac{di_L}{dt} = v_s - V_{\text{out}}, \quad (4.0.11)$$

where:

- v_L is the voltage across the output inductor,
- L is the inductance of the output filter inductor,
- i_L is the inductor current,
- $\frac{di_L}{dt}$ is the rate of change of inductor current,
- v_s is the rectified secondary-side voltage,
- V_{out} is the output voltage.

This relationship ensures controlled current variation and reduces high-frequency ripple in the output current.

Output Capacitor: The output capacitor maintains voltage stability and provides energy buffering between the converter and the load. Its dynamics are expressed as:

$$C \frac{dV_{\text{out}}}{dt} = i_L - i_{\text{load}}. \quad (4.0.12)$$

where:

- C is the output filter capacitance,
- V_{out} is the output voltage,

- $\frac{dV_{out}}{dt}$ is the rate of change of output voltage,
- i_L is the inductor current,
- i_{load} is the load current drawn by the IT load.

The capacitor absorbs switching ripple and supports the load during transient conditions, thereby maintaining a regulated DC output voltage.

The selection of L and C is based on the allowable current and voltage ripple limits, as well as the desired transient response of the converter.

4.3.2.3 Control Architecture

The DC–DC converter employs a closed-loop control strategy to regulate the output voltage under varying load conditions. A proportional–integral–derivative (PID) controller is used to ensure accurate tracking of the reference voltage.

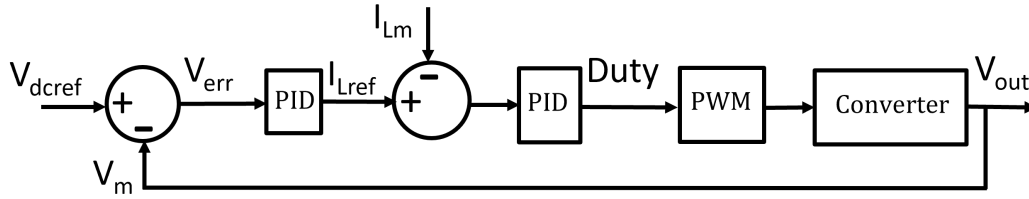


Figure 4.5: Control architecture of the DC–DC isolated converter.

As illustrated in Fig. 4.5, the output voltage is continuously measured and compared with a reference value, producing a voltage error. This error is processed through a PI controller to produce a reference inductor current, $i_{L,ref}$, which represents the required inductor current to maintain the desired output voltage.

The reference inductor current is then compared with the measured inductor current i_L to form a current error. This error is processed by a second PI controller to generate the control signal that determines the PWM generator’s switching duty ratio.

Through coordinated regulation of the output voltage and inductor current, the converter achieves stable output-voltage control while ensuring appropriate current dynamics within the power stage.

Control Operation: An increase in load demand results in a temporary drop in the output voltage, which increases the voltage error and consequently raises the reference inductor current. The current control action then increases the duty ratio, allowing more energy to be transferred through the transformer and restoring the output voltage to its reference value.

The duty ratio directly determines the effective voltage applied to the transformer primary, thereby controlling the transferred power and output voltage. By continuously adjusting the duty ratio, the controller ensures stable operation under dynamic IT load conditions.

The controller bandwidth is selected to be sufficiently higher than the IT load’s dominant dynamics to ensure fast transient response while maintaining stability.

4.3.2.4 Switching and PWM Implementation

The duty-related control command obtained from the voltage control loop is applied to a PWM generator to produce gating signals for the full-bridge switches. The PWM generator operates at a fixed switching frequency of 10 kHz.

$$d(t) = \frac{v^*(t)}{V_{\text{in}}}, \quad (4.0.13)$$

where:

- $d(t)$ is the duty ratio applied to the PWM generator,
- $v^*(t)$ is the control voltage output from the controller,
- V_{in} is the DC-link input voltage,
- t represents time.

The switching action produces a high-frequency square-wave voltage at the transformer primary. This voltage is transferred through the transformer, rectified on the secondary side, and filtered to obtain a regulated DC output. The selected switching frequency provides a balance between switching losses and passive component size, enabling compact design and fast dynamic response. The key parameters of the DC–DC isolated converter are summarized in Table 4.2.

Table 4.2: Key parameters of the DC–DC isolated converter

Component / Parameter	Symbol	Value	Description / Remarks
Input DC Voltage	V_{in}	480 V	DC-link input from PFC stage
Output Voltage	V_{out}	48 V	Regulated DC output for IT loads
Transformer Turns Ratio	n	10	Voltage scaling and isolation
Output Inductor	L	1.62 μH	Output current smoothing
Output Capacitor	C	0.271 mF	Output voltage stabilization
Switching Frequency	f_s	10 kHz	Full-bridge switching frequency

4.3.2.5 Summary

The developed DC–DC isolated converter model accurately represents the electrical behavior and control dynamics of the second-stage PSU, enabling reliable voltage conversion, isolation, and robust performance under dynamic IT load conditions. The key characteristics are summarized as follows:

- Voltage step-down,
- Isolation,
- Low ripple,
- Fast transient response.

4.4 Modeling of IT Load

The IT load represents the computational demand of a data center and is primarily associated with servers, storage systems, and networking equipment. Unlike conventional static loads, its power consumption is closely related to computational workload and can vary significantly with server utilization, workload scheduling, and application characteristics. Therefore, an appropriate IT load model should represent both the magnitude and temporal variation of computational demand while remaining scalable for system-level power-system studies.

In this work, a generalized IT load modeling framework is developed to establish a connection between data-center workload characteristics and the detailed electrical model. Depending on data availability, the framework can utilize measured server or rack power, historical or forecasted workload profiles, or direct server-utilization measurements. These inputs are aggregated and normalized, when required, to obtain a representative workload trajectory. The resulting profile is subsequently translated into server utilization, server power consumption, aggregate IT demand, and finally an equivalent electrical load connected to the power supply unit (PSU).

The proposed framework therefore provides a common modeling interface for different data-center sizes, types, and workload characteristics. The physical interpretation of the modeling process and the generalized mathematical workflow are illustrated in Fig. 4.6 and Fig. 4.7, respectively. The individual modeling stages are described in the following subsections.

4.4.1 Overview of IT Load Modeling

The operating characteristics of IT loads depend strongly on the purpose, scale, computing architecture, and workload of the data center. Data centers may generally be categorized according to their application and ownership structure as hyperscale, colocation, wholesale colocation, enterprise, telecom, and other specialized facilities. Hyperscale facilities typically contain large populations of interconnected servers and support cloud computing, large-scale storage, high-performance computing, and increasingly artificial intelligence (AI) applications. Enterprise data centers primarily support the computing requirements of a particular organization, whereas colocation and wholesale colocation facilities provide space, power, and cooling infrastructure for multiple customers or large-scale tenants. Telecom data centers primarily support communication, content-delivery, and network services.

These facility types can differ substantially in server population, computing density, workload composition, resource allocation, and operating scale. Consequently, their aggregate IT power profiles may exhibit different magnitudes and temporal characteristics. Highly coordinated computational workloads, such as AI training, may produce relatively rapid changes in aggregate server activity, whereas diversified workloads distributed among many independent applications may produce comparatively smoother demand. A generalized IT load model should therefore be capable of accommodating different workload profiles without requiring a separate electrical model for every data-center application.

Data centers are also commonly classified according to infrastructure availability us-

ing Tier I–IV classifications. The tier classification primarily describes the redundancy, maintainability, and fault-tolerance of the supporting electrical and cooling infrastructure rather than the computational workload itself. Increasing tier levels introduce additional capacity components, distribution paths, and redundancy to improve service continuity. Accordingly, data-center type and workload characteristics primarily determine the computational demand, whereas tier classification influences how that demand is supplied and maintained during normal operation, maintenance, equipment outages, and other operating conditions.

Considering these variations, directly developing a separate server-level model for every data-center type, tier, capacity, and workload is impractical for system-level studies. Instead, the proposed framework uses a common modeling procedure in which measured, historical, forecasted, or representative workload data are translated into a generalized IT load representation. The overall IT load modeling procedure can be summarized as follows:

1. Characterize the data center based on its type, tier, capacity, and workload.
2. Obtain measured, historical, forecasted, or representative workload/load data.
3. Aggregate and normalize the available data, when required.
4. Map the processed workload profile to the aggregate server utilization $U(t)$.
5. Convert $U(t)$ into the corresponding IT power demand $P_{IT}(t)$.
6. Translate $P_{IT}(t)$ into the equivalent electrical load representation for simulation.

Depending on the available dataset, measurements or representative profiles may be obtained at the individual server, rack, cluster, or aggregated IT-load level. When multiple load profiles are available, they are first combined to obtain an aggregate IT demand:

$$P_{IT,agg}(t) = \sum_{i=1}^{N_g} P_i(t) \quad (4.0.14)$$

where $P_i(t)$ represents the measured or estimated power consumption of the i -th server, rack, cluster, or workload group, and N_g is the number of aggregated groups.

Since the absolute power level can vary substantially among facilities, the aggregated profile may be normalized before it is applied to the generalized model. A normalized profile is expressed as

$$P_{IT,n}(t) = \frac{P_{IT,agg}(t) - P_{IT,min}}{P_{IT,max} - P_{IT,min}} \quad (4.0.15)$$

where $P_{IT,min}$ and $P_{IT,max}$ denote the minimum and maximum power levels associated with the selected profile or operating range. The resulting profile satisfies

$$0 \leq P_{IT,n}(t) \leq 1 \quad (4.0.16)$$

and preserves the temporal characteristics of the original workload while removing its dependence on absolute facility capacity.

Within the utilization-based modeling approach adopted in this work, the normalized workload profile is mapped to an aggregate server utilization trajectory $U(t)$. If server-utilization measurements are directly available, the normalization and mapping stage may be bypassed and the measured utilization may be applied directly.

Figure 4.6 illustrates the physical interpretation of the proposed modeling framework. The measured or prescribed server-utilization and workload information are processed through the mathematical model to determine individual and aggregate IT power. The resulting power demand is then translated into a time-varying resistance connected to the 48-V DC bus, thereby establishing a direct link between the computational workload representation and its physical electrical implementation.

The complete generalized modeling workflow is shown in Fig. 4.7. The upper stages identify the data-center characteristics and available workload information, while the subsequent stages perform aggregation, normalization, utilization modeling, power calculation, coherent load aggregation, and physical electrical translation.

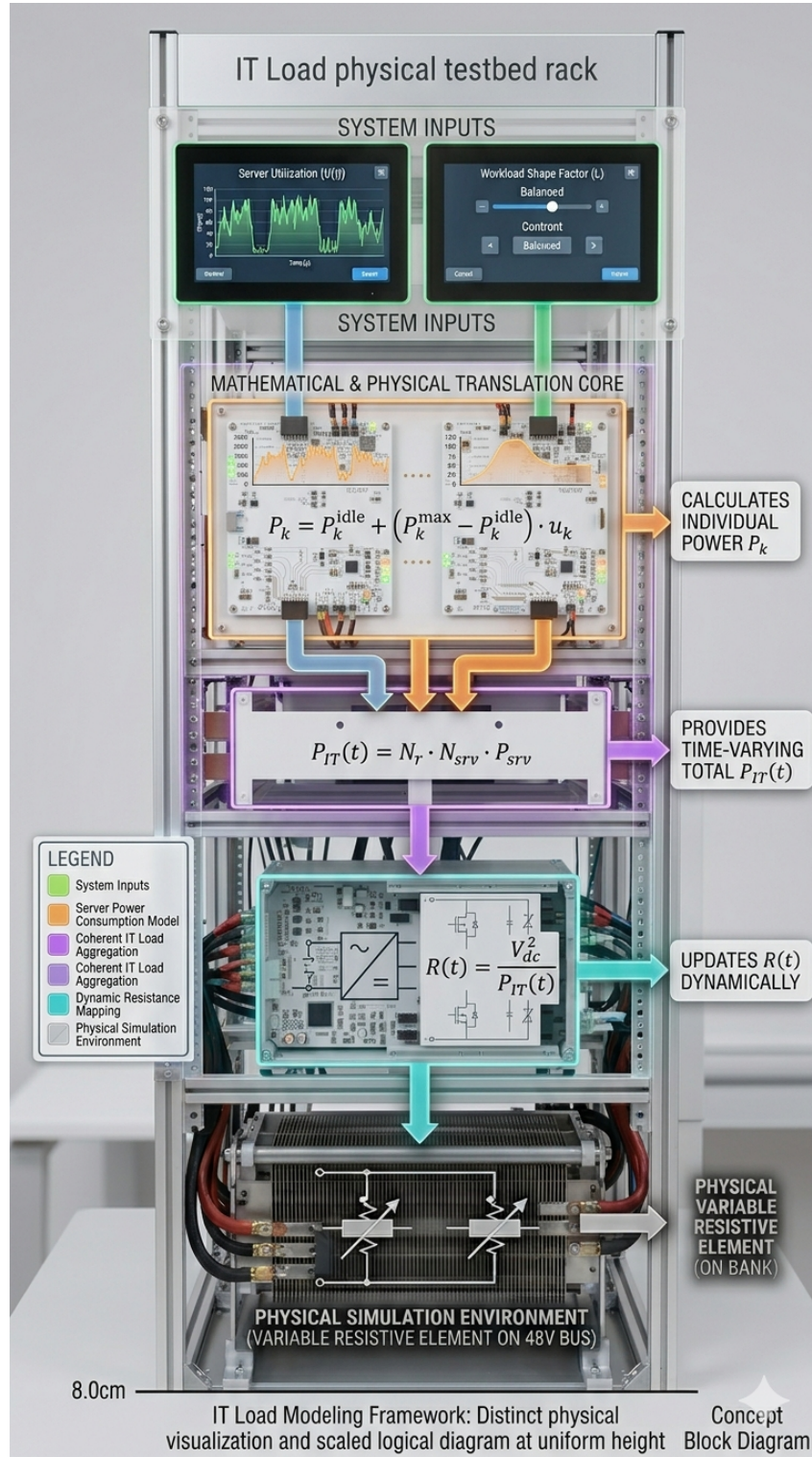


Figure 4.6: Physical representation and flow abstraction of the proposed IT load model.

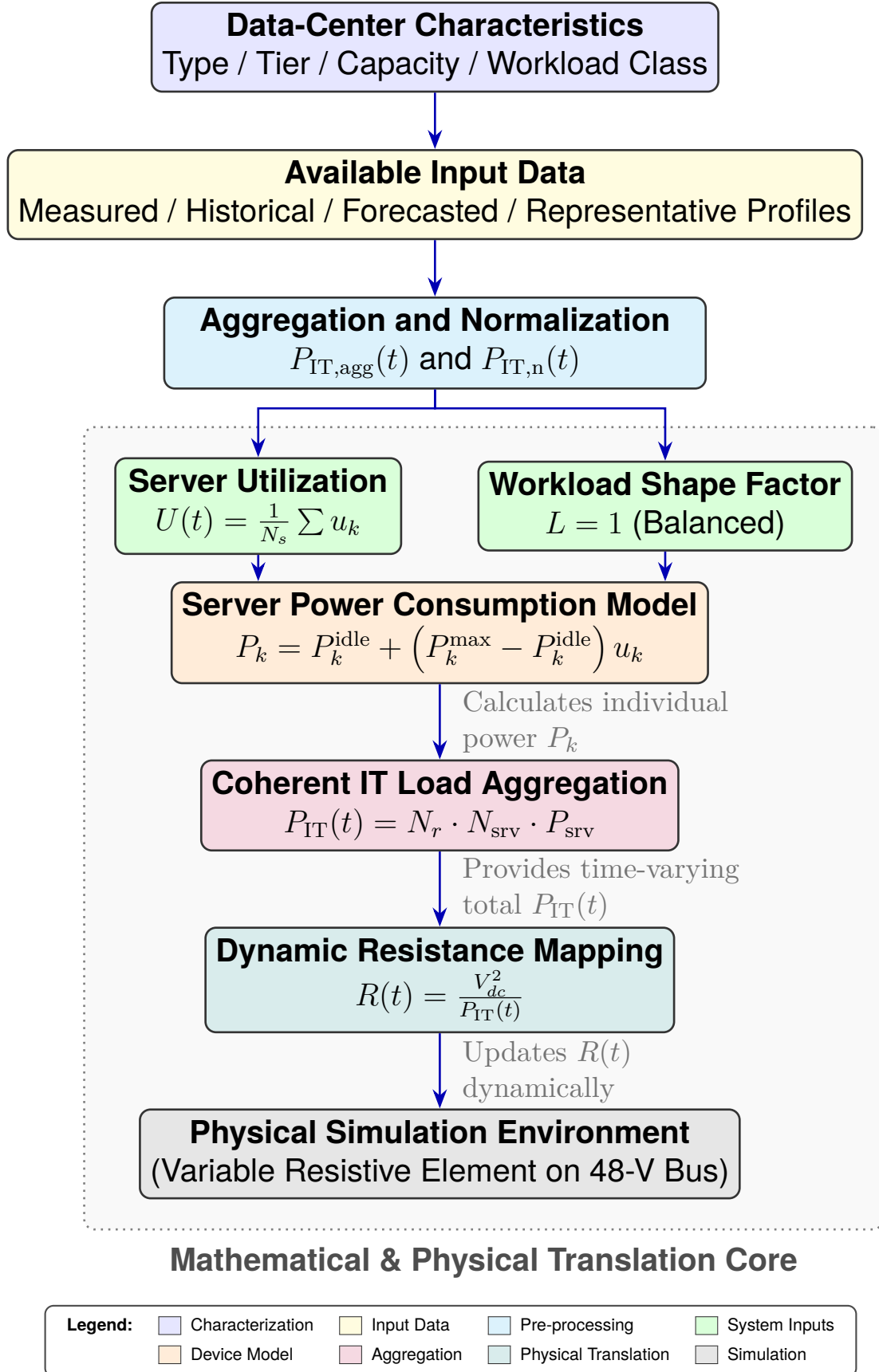


Figure 4.7: Generalized IT load modeling framework.

4.4.2 Server Utilization Modeling

The generalized workload representation developed in the preceding subsection is translated into server utilization for determining the corresponding IT power consumption. Let the data center consist of N_s homogeneous or equivalently aggregated servers. The utilization of the k -th server is denoted by u_k , where

$$0 \leq u_k \leq 1. \quad (4.0.17)$$

The average utilization of the data-center server population is defined as

$$U = \frac{1}{N_s} \sum_{k=1}^{N_s} u_k \quad (4.0.18)$$

where U may be obtained directly from server-utilization measurements or derived from the normalized workload profile described in Section 4.4.1. This aggregate utilization captures the combined effect of computational workload, task scheduling, and resource allocation across the server cluster [19].

In addition to the average utilization level, the distribution of computational tasks among the available servers can affect their individual operating points. To represent different workload distribution strategies, such as load balancing and workload consolidation, a utilization shaping factor L is introduced. The effective server utilization is expressed as

$$u_k = \frac{U}{L(1 - U) + U} \quad (4.0.19)$$

where

- $L = 1$ corresponds to *perfect load balancing*, where the workload is uniformly distributed across the available servers, and
- $L \rightarrow 0$ represents *maximum consolidation*.

The parameter L therefore provides a mechanism for representing differences in workload-management strategies without changing the underlying IT load model. In this work, $L = 1$ is adopted to represent uniform workload distribution across the server population.

4.4.3 Server Power Consumption Model

Once the server utilization is determined, the corresponding electrical power demand is calculated using a utilization-dependent server power model. Servers consume a nonzero amount of power under idle conditions, while their power consumption generally increases with computational utilization. Over the operating range considered in this work, this relationship is represented using an approximately linear model [10, 19].

The power consumption of the k -th server is expressed as

$$P_k = P_k^{\text{idle}} + (P_k^{\text{max}} - P_k^{\text{idle}}) u_k \quad (4.0.20)$$

where

- P_k^{idle} is the idle power consumption of the server, and
- P_k^{max} is the server power consumption at full utilization.

The total IT load power is obtained by aggregating the power consumption of all servers:

$$P_{\text{IT}} = \sum_{k=1}^{N_s} P_k \quad (4.0.21)$$

For a large-scale data center represented using homogeneous or equivalent servers, the formulation can be simplified as

$$P_{\text{IT}} = N_s \left[P^{\text{idle}} + \left(P^{\text{max}} - P^{\text{idle}} \right) U \right]. \quad (4.0.22)$$

This formulation separates the temporal characteristics of the computational workload, represented by $U(t)$, from the physical scale of the modeled data center, represented by N_s , P^{idle} , and P^{max} . Consequently, workload profiles obtained from different facilities or datasets can be mapped to an appropriate data-center capacity while preserving their characteristic temporal variations.

4.4.4 Dynamic Behavior of IT Load

The utilization and server power models provide a generalized mechanism for introducing time-varying computational demand into the electrical simulation. Unlike conventional loads that are frequently represented using static or slowly varying characteristics, data-center IT loads can exhibit rapid changes associated with computational activity, task scheduling, and workload transitions. These characteristics are particularly important for large-scale high-performance computing and AI workloads, for which coordinated changes in server activity can produce significant variations in aggregate power demand [9].

Accordingly, the dynamic IT load is represented as a time-varying function of server utilization:

$$P_{\text{IT}}(t) = f(U(t)). \quad (4.0.23)$$

Using the server power relationship in (4.0.22), the dynamic aggregate power can be expressed as

$$P_{\text{IT}}(t) = N_s \left[P^{\text{idle}} + \left(P^{\text{max}} - P^{\text{idle}} \right) U(t) \right]. \quad (4.0.24)$$

Here, $U(t)$ can represent a directly measured utilization trajectory or a utilization profile derived from the aggregation and normalization of measured, historical, forecasted, or representative workload data.

This formulation enables different temporal workload characteristics to be incorporated without changing the underlying electrical model. Different profiles associated with enterprise computing, diversified colocation loads, cloud computing, or rapidly varying AI workloads can therefore be represented by applying the corresponding $U(t)$ trajectory to the same generalized server-power formulation.

4.4.5 Coherent IT Load Representation

Although the server-level formulation provides a direct relationship between computational utilization and electrical power, explicitly representing every individual server becomes computationally impractical for large data centers containing thousands of servers. Therefore, the individual server loads are aggregated into a coherent equivalent representation suitable for system-level simulation.

The coherent IT load is expressed as

$$P_{\text{IT,tot}} = N_r \cdot N_{\text{srv}} \cdot P_{\text{srv}} \quad (4.0.25)$$

where

- N_r is the number of equivalent racks,
- N_{srv} is the number of servers per rack, and
- P_{srv} is the power consumption of a representative server.

It is noted that $P_{\text{IT,tot}}$ is equivalent to the aggregated IT load P_{IT} defined in (4.0.22), with

$$N_s = N_r N_{\text{srv}}. \quad (4.0.26)$$

The coherent formulation therefore preserves the aggregate power consumption and temporal workload characteristics while avoiding explicit electrical representation of every individual server. This reduces computational complexity and makes the model suitable for integration with the detailed data-center power-delivery system.

The formulation also provides a convenient scaling mechanism for representing data centers of different capacities. Once a workload profile has been mapped to $U(t)$, the modeled capacity can be adjusted through the equivalent number of racks and servers and their corresponding power ratings. Thus, the workload profile determines the temporal variation of IT demand, while the coherent rack/server representation determines its aggregate power scale.

4.4.5.1 Implementation of IT Load in Simulation

The preceding modeling stages convert data-center workload information into the time-varying aggregate IT power demand $P_{\text{IT}}(t)$. To interface this computational load representation with the detailed electrical model, the IT load is implemented as a variable-resistive element connected to the PSU DC output at the 48-V bus, as illustrated in Fig. 4.6 and Fig. 4.7.

For a specified IT power demand $P_{\text{IT}}(t)$, the corresponding equivalent resistance is calculated using the fundamental power-voltage relationship:

$$R(t) = \frac{V_{dc}^2}{P_{\text{IT}}(t)} \quad (4.0.27)$$

where

- V_{dc} is the PSU output/DC-bus voltage, and

- $P_{IT}(t)$ is the time-varying aggregate IT load power obtained from the utilization-based model.

The resistance $R(t)$ is continuously updated according to the prescribed IT-load trajectory. As a result, the effective electrical load seen by the PSU varies with the computational workload, ensuring that the instantaneous power consumption follows the desired $P_{IT}(t)$ profile.

This implementation establishes the complete connection between the computational workload representation and the detailed electrical model. Measured or representative workload data are first aggregated and normalized, translated into server utilization, converted into individual and aggregate server power, and finally implemented as a time-varying electrical load at the PSU output.

The proposed approach therefore provides a common electrical interface for representing different data-center capacities and workload profiles while maintaining compatibility with the detailed PSU and overall data-center model. The resulting implementation provides the basis for evaluating the developed IT load model against representative workload profiles in the following subsection.

4.4.6 Performance Evaluation of IT Load Model

The developed IT load model is validated by comparing its ability to reproduce realistic data center load profiles reported in [9, 19]. The objective is to demonstrate that the proposed modeling approach, with variable-resistive-load implementation, can accurately reproduce time-varying IT load behavior. From the corresponding time-varying power demand $P_{IT}(t)$, this power is translated into an equivalent resistance using (4.0.27), allowing the load to be physically realized as a time-varying electrical load within the PSU–IT load circuit.

4.4.6.1 AI Training Load Profile

Fig. 4.8 compares a AI data center load profile during training [9] with the corresponding load profile generated by the proposed implementation.

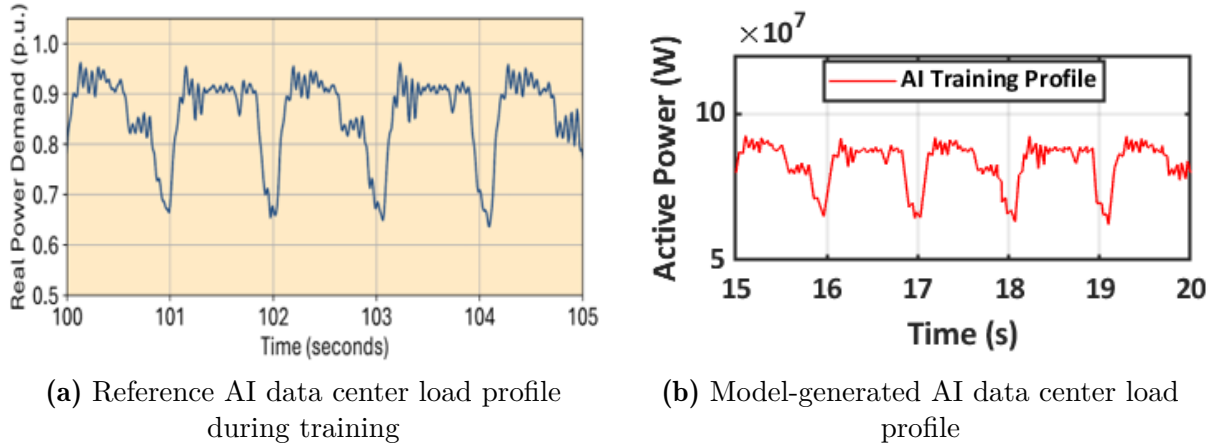


Figure 4.8: Comparison of reference and modeled AI training load profiles.

The proposed implementation successfully reproduces the characteristic features of the reference load profiles through dynamic resistance realization. This demonstrates that the model with the dynamic resistance realization effectively captures the characteristics of high-performance computing loads.

4.4.6.2 Utilization-Based Load Profile

Fig. 4.9 presents the relationship between server utilization and power consumption [19], along with the corresponding load profile generated by the model.

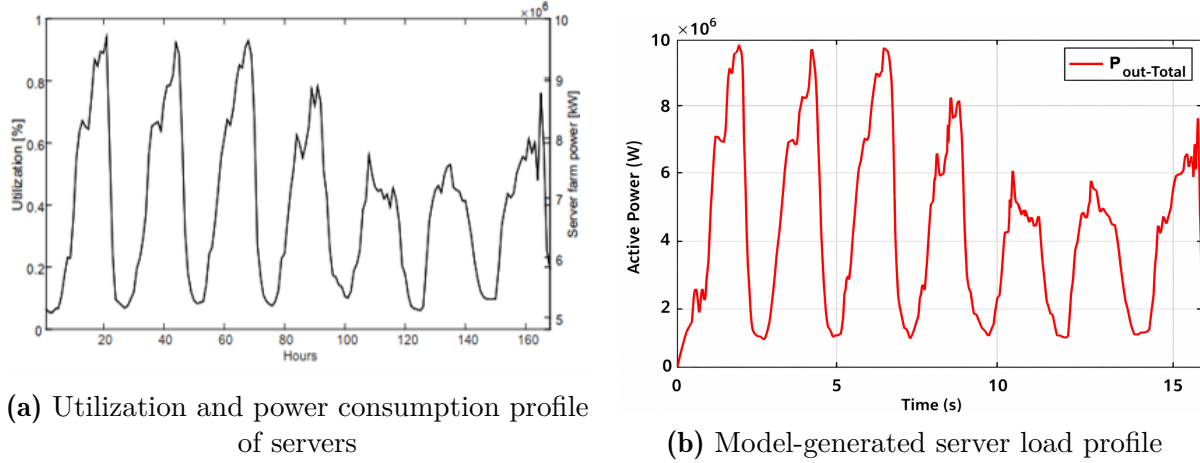


Figure 4.9: Comparison of reference and modeled server load profiles.

The results confirm that the proposed model accurately translates utilization-based power variations into corresponding power demand profiles. The generated load closely follows the expected temporal variations, validating the correctness of the utilization-to-power mapping through the model.

4.4.6.3 Discussion

Overall, the validation results demonstrate that:

- The proposed framework is capable of representing a wide range of IT load behaviors by accommodating time-varying power demand profiles
- The approach enables integration of different types of workload-driven IT load characteristics within a unified modeling and simulation framework
- The utilization-based formulation, combined with the power-to-resistance mapping, enables a physically consistent realization of IT load behavior within the PSU-load circuit
- The proposed approach provides a flexible, scalable interface for incorporating additional subsystems, such as cooling loads, particularly when cooling demand is modeled as a function of IT load power.

4.5 Modeling of Uninterruptible Power Supply (UPS)

The Uninterruptible Power Supply (UPS) ensures continuous and high-quality power delivery to critical IT loads under normal and disturbed operating conditions. The UPS operates in multiple modes, including bypass (eco mode), double-conversion mode, and battery mode.

From a modeling perspective, the UPS system consists of three main subsystems:

- AC–DC conversion stage (rectifier),
- DC energy storage system (battery),
- DC–AC conversion stage (inverter).

4.5.1 Topology

The UPS topology follows a double-conversion architecture in which the input AC power is first converted to DC by a controlled rectifier and then inverted back to AC using a voltage-source inverter.

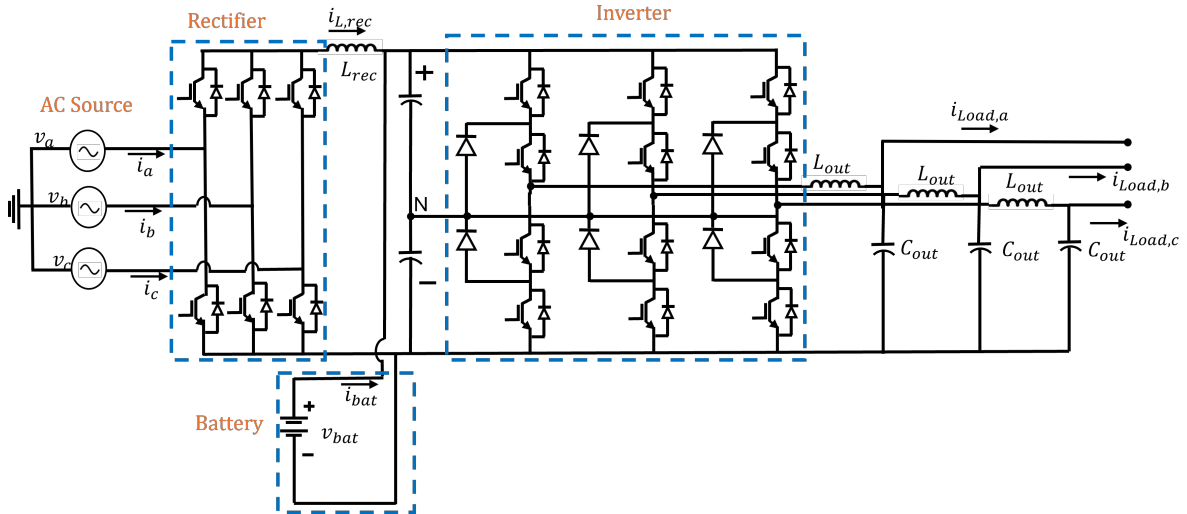


Figure 4.10: UPS model showing rectifier, DC-link, battery, inverter, and output filter.

As shown in Fig. 4.10, the rectifier stage converts the input AC voltage into a regulated DC-link voltage. The DC-link is interfaced with a battery system that provides energy storage and backup capability. The inverter stage converts the DC-link voltage into a controlled three-phase AC output.

The inverter output is interfaced with the load through an output filter, which smooths the converter switching effects and ensures a high-quality sinusoidal voltage supplied to the load.

4.5.2 Modes of Operation

The UPS operates in multiple modes depending on grid conditions and system requirements. These modes determine the power flow path and control objectives of the rectifier and inverter stages.

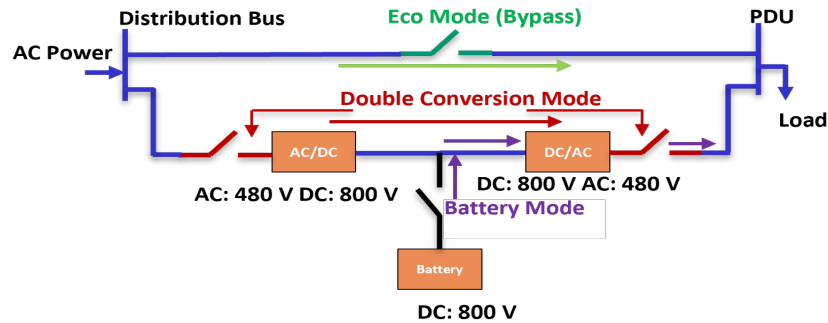


Figure 4.11: Modes of operation of UPS.

4.5.2.1 Bypass Mode (Eco Mode)

In bypass or eco mode, the load is supplied directly from the utility grid via a bypass path, while the UPS remains in standby.

- High efficiency due to reduced conversion losses,
- Limited power conditioning capability,
- Load is exposed to grid disturbances.

4.5.2.2 Double-Conversion Mode (Online Mode)

In double-conversion mode, the input AC power is first converted to DC through the rectifier and then inverted back to AC to supply the load.

- Complete isolation from grid disturbances,
- Continuous operation of both rectifier and inverter stages.

In this mode, the DC-link voltage is regulated by the rectifier, while the inverter controls the output voltage supplied to the load.

4.5.2.3 Battery Mode

In battery mode, the grid supply is unavailable or outside acceptable limits. The DC-link is maintained by the battery, and the inverter continues supplying power to the load.

- Ensures uninterrupted power supply,
- Inverter operates as a standalone (grid-forming) voltage source,
- Battery provides the required energy to support the load.

4.5.2.4 Mode Transition

The transition between operating modes is governed by system conditions such as voltage deviation, frequency limits, and fault detection. The control system ensures smooth transitions to avoid disturbances in the load voltage.

In the developed model, mode selection is implemented through control logic that switches between:

- Rectifier-controlled DC-link regulation (double-conversion mode),
- Inverter-controlled voltage generation (battery mode),
- Bypass path (direct grid supply).

4.5.3 Filter Design

The UPS employs filtering elements on both the DC-link and inverter output to ensure stable operation and improved power quality.

4.5.3.1 DC-Link Capacitor

The DC-link capacitor stabilizes the DC voltage and balances the instantaneous power difference between the rectifier and inverter stages:

$$C_{dc} \frac{dV_{dc}}{dt} = i_{rec} - i_{inv}, \quad (4.0.28)$$

where:

- C_{dc} is the DC-link capacitance,
- V_{dc} is the DC-link voltage,
- i_{rec} is the rectifier current,
- i_{inv} is the inverter input current.

The DC-link capacitor acts as an energy buffer, maintaining voltage stability during load transients and mode transitions.

4.5.3.2 Output Filter

As shown in Fig. 4.10, the inverter output is interfaced with the load through a three-phase L – C filter. This representation depicts a balanced three-phase filtering structure, with identical components in each phase to ensure uniform attenuation of switching ripple and balanced voltage conditions at the load terminals.

The filter attenuates high-frequency switching components and ensures a smooth three-phase voltage at the load.

Inductor Dynamics: The output inductors limit the rate of current change and suppress high-frequency components. For a balanced three-phase system, the dynamics can be expressed in vector form as:

$$v_{L,abc} = L_{\text{out}} \frac{di_{abc}}{dt} = v_{\text{inv},abc} - v_{\text{o},abc}, \quad (4.0.29)$$

where:

- L_{out} is the per-phase inductance,
- i_{abc} is the three-phase inductor current vector,
- $v_{\text{inv},abc}$ is the inverter output voltage,
- $v_{\text{o},abc}$ is the load-side voltage.

Capacitor Dynamics: The output capacitors provide voltage smoothing and support load dynamics. The capacitor behavior for the three-phase system is given by:

$$C_{\text{out}} \frac{dv_{\text{o},abc}}{dt} = i_{abc} - i_{\text{load},abc}, \quad (4.0.30)$$

where:

- C_{out} is the per-phase capacitance,
- $v_{\text{o},abc}$ is the three-phase load voltage,
- i_{abc} is the inductor current vector,
- $i_{\text{load},abc}$ is the load current.

The combined action of L_{out} and C_{out} forms a balanced three-phase filter that attenuates switching harmonics and maintains a high-quality sinusoidal voltage at the load.

4.5.4 Control Architecture

The UPS control architecture consists of independent control strategies for the rectifier and inverter stages.

4.5.4.1 Rectifier Control (AC–DC Stage):

The rectifier employs a cascaded voltage–current control structure to regulate the DC-link voltage while ensuring controlled input current dynamics. The control architecture consists of two hierarchical loops: an outer voltage regulation loop and an inner current control loop.

As illustrated in Fig. 4.12, the rectifier control is organized into two main functional stages: voltage regulation and current control. In the control architecture:

- $V_{\text{dc,ref}}$ represents the desired DC-link voltage,
- $V_{\text{dc,m}}$ represents the measured DC-link voltage,

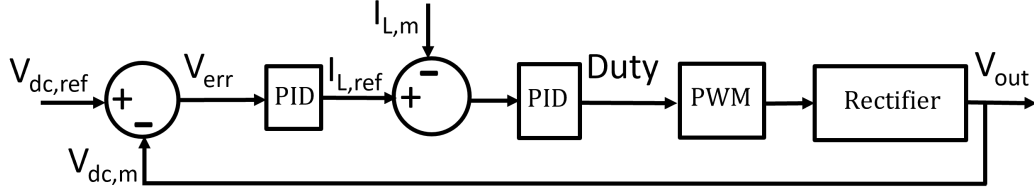


Figure 4.12: Control architecture of the UPS Rectifier.

- $I_{L,ref}$ is the reference inductor current,
- $I_{L,m}$ is the measured inductor current,
- the PID controllers regulate voltage and current loops,
- D represents the duty ratio used for PWM generation.

Voltage Regulation: The DC-link voltage is regulated by comparing the reference voltage $V_{dc,ref}$ with the measured voltage $V_{dc,m}$. The resulting voltage error is processed through a proportional–integral–derivative (PID) controller to generate the reference inductor current $I_{L,ref}$.

This reference current represents the amount of current required from the AC side to maintain the desired DC-link voltage under varying load conditions.

Current Control: The reference inductor current $I_{L,ref}$ is compared with the measured inductor current $I_{L,m}$ to form a current error. This error is processed through a second PI controller to generate the duty ratio D .

The duty ratio determines the switching action of the rectifier through the PWM block, enabling the inductor current to track its reference accurately.

Control Operation: Any change in load demand results in a corresponding deviation in the DC-link voltage, which alters the voltage error and consequently adjusts the reference inductor current $I_{L,ref}$. The current control loop responds by modifying the duty ratio, thereby regulating the power drawn from the AC source to restore the DC-link voltage to its reference value.

Through this cascaded control structure, the rectifier maintains a stable DC-link voltage while ensuring smooth and controlled input current dynamics.

4.5.4.2 Inverter Control (DC–AC Stage):

The UPS inverter is modeled as a voltage-controlled source that generates a balanced three-phase AC output using a three-level neutral-point-clamped (NPC) converter. The control objective is to synthesize a sinusoidal output voltage with desired magnitude, frequency, and phase, while maintaining stability under different operating modes.

As illustrated in Fig. 4.13, the inverter control is organized into three main functional stages: voltage regulation, phase generation, and reference voltage synthesis. In the control architecture:

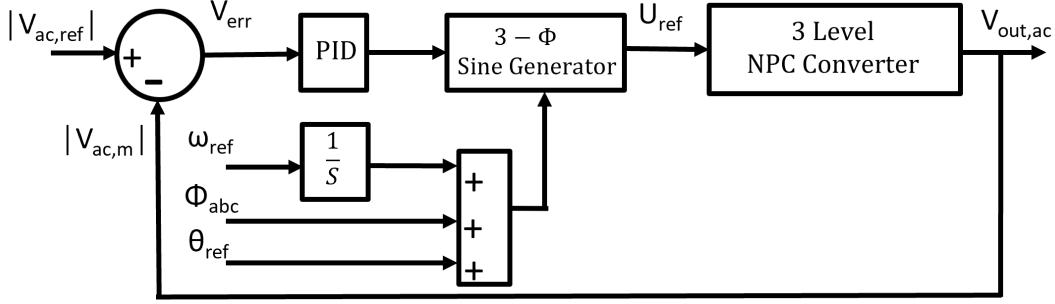


Figure 4.13: Control architecture of the UPS inverter.

- $|V_{ac,ref}|$ represents the desired output voltage magnitude,
- θ_{ref} represents the reference phase angle,
- ϕ_{abc} defines the three-phase displacement,
- the sinusoidal block generates unit three-phase waveforms,
- the PI controller regulates the output voltage magnitude,
- U_{ref} represents the reference voltage signal tracked by the NPC inverter using an average model.

Voltage Regulation: The output AC voltage magnitude is regulated by comparing the reference voltage magnitude $|V_{ac,ref}|$ with the measured output voltage magnitude $|V_{ac,m}|$. The resulting voltage error is processed through a proportional–integral–derivative (PID) controller to generate a normalized voltage command signal. This signal determines the amplitude of the inverter reference voltage.

Phase Processing and Three-Phase Construction: The reference angular frequency ω_{ref} is used to obtain the phase angle θ_{ref} , which defines the instantaneous position of the output voltage waveform.

The three-phase structure is established by incorporating the phase displacement term ϕ_{abc} , which represents the fixed phase shifts of a balanced three-phase system:

$$\phi_{abc} = \{0^\circ, -120^\circ, +120^\circ\}. \quad (4.0.31)$$

By combining θ_{ref} with ϕ_{abc} , three phase-shifted angle signals are obtained, corresponding to the three phases of the inverter output.

Reference Voltage Formation: The phase signals are passed through a sinusoidal function block to generate normalized three-phase waveforms. These waveforms are scaled by the voltage command signal obtained from the voltage control loop, producing the reference voltage signal U_{ref} .

Inverter Operation: The reference voltage U_{ref} is applied to the three-level NPC inverter, which operates using an average model representation. The inverter's output voltage is generated from the reference signal. The inverter converts the DC-link voltage into a controlled three-phase AC output, supplying the load with the required voltage magnitude and frequency. This control structure allows the inverter to operate as a grid-forming voltage source during battery or islanded operation, while maintaining stable and balanced three-phase output under varying load conditions.

4.5.5 Switching and PWM Implementation

The switching implementation of the UPS system is directly governed by the control signals generated in the rectifier and inverter control architectures. The rectifier and inverter stages employ different modulation approaches based on their respective control objectives and modeling levels.

- **Rectifier Switching:** In the rectifier stage, the duty ratio D generated by the inner current control loop is applied to a PWM generator operating at a fixed switching frequency. This PWM signal determines the gating of the rectifier switching devices, thereby controlling the inductor current and regulating the DC-link voltage. The switching action ensures that the input current follows its reference while maintaining the desired DC-link voltage under varying load conditions.
- **Inverter Implementation (Average Model):** In the inverter stage, the control architecture generates a three-phase reference voltage signal U_{ref} , which represents the desired output voltage waveform in terms of magnitude, phase, and frequency.

Instead of explicitly generating switching pulses, the inverter is modeled using an average-value representation. In this approach, the converter directly tracks the reference voltage U_{ref} to produce the corresponding AC output voltage. The key parameters of the UPS system used in this study are summarized in Table 4.3.

Table 4.3: Key parameters of the UPS system

Component / Parameter	Symbol	Value	Description / Remarks
Input AC Voltage	v_s	480 V	Utility grid input
DC-link Voltage	V_{dc}	800 V	Regulated DC-link voltage
Output Voltage	V_{out}	480 V	Three-phase AC output to load
Battery Voltage	V_{bat}	800 V	DC energy storage system
Switching Frequency (Rectifier)	f_s	10 kHz	PWM switching frequency

4.5.6 Summary

The developed UPS model captures the dynamic interaction between the rectifier, DC-link, battery, and inverter stages. It ensures:

- Regulated DC-link voltage,

- Seamless transition between operating modes,
- Stable and sinusoidal AC output voltage,
- Fast dynamic response under load variations.

4.6 Modeling of Cooling System

Cooling systems constitute a significant portion of the total facility demand in large-scale data centers. However, in the present work, the cooling system is not modeled as a separate detailed dynamic subsystem. Instead, its electrical impact is incorporated through its dependency on IT load demand, so that the combined effect of IT and cooling consumption is represented within the overall aggregated data center load.

4.6.1 Cooling Load Representation Using Power Usage Effectiveness (PUE)

The cooling demand is estimated using the Power Usage Effectiveness (PUE) metric, which relates the total facility power consumption to the IT load. PUE is defined as:

$$\text{PUE} = \frac{P_{\text{total}}}{P_{\text{IT}}} \quad (4.0.32)$$

For modern hyperscale data centers, PUE values typically range between 1.2 and 1.4. In this study, a representative value of 1.3 is adopted to characterize a contemporary, energy-efficient facility. Accordingly, the total facility load is expressed as:

$$P_{\text{total}} = \text{PUE} \times P_{\text{IT}} \quad (4.0.33)$$

The non-IT portion, primarily attributed to cooling and associated auxiliary infrastructure, is therefore estimated as:

$$P_{\text{cooling}} = (\text{PUE} - 1) \times P_{\text{IT}} \quad (4.0.34)$$

Under this representation, the cooling demand varies proportionally with the IT load. This provides a practical way to capture the dominant influence of cooling on total data center demand without introducing separate thermo-mechanical subsystem models.

4.6.2 Cooling System Composition and Representative Load Breakdown

Although the cooling system is not explicitly modeled as an independent subsystem in the present simulation framework, a representative breakdown of typical data center cooling loads is provided for context. This breakdown reflects common industry practice and identifies the major contributors to cooling demand.

The values listed in Table 4.4 are intended for illustrative and engineering reference purposes only; they are not implemented as separate component-level models in the present study.

Table 4.4: Cooling System Load Breakdown for 100 MW Data Center

Cooling Component	Rated Power (MW)	Cooling Load Share (%)	Load Type	Representative Description
Chillers	15.0	50	Motor-driven load	Major cooling production unit
Pumps	6.0	20	Motor-driven load	Cooling fluid circulation
Cooling tower fans	4.5	15	Fan load	Heat rejection support
CRAC/CRAH fans	4.5	15	Fan load	Air handling and rack-level cooling
Total Cooling Load	30.0	100	–	–

In the implemented model, these components are not represented individually; instead, their combined effect is incorporated into the aggregated load through the PUE-based formulation described above.

4.6.3 Cooling Load Modeling Approach

From an electrical modeling perspective, the cooling load is represented indirectly as a load component dependent on IT demand. Thus, the cooling contribution is embedded within the overall data center load rather than introduced as a separate network-connected subsystem.

This approach allows the combined IT and cooling demand to be represented using a unified load model with adjustable power or equivalent resistance, depending on the study objective and implementation stage. As the IT load varies with workload conditions, the cooling demand varies proportionally through the adopted PUE relationship, thereby reflecting the first-order coupling between computation demand and supporting thermal infrastructure.

Such an aggregated representation is appropriate for system-level dynamic simulations in which the primary interest lies in the electrical interaction between the data center and the grid, rather than in detailed thermal or motor-drive dynamics of individual cooling equipment.

4.6.4 Electrical Interface and Voltage Level

In practical operation, cooling equipment is supplied from the low-voltage AC distribution system within the data center. In the present model, the aggregated cooling contribution is included within the total load supplied from the low-voltage AC side downstream of the PDU.

Therefore, the implemented electrical demand seen by the upstream network represents the combined effect of IT and cooling power consumption. This abstraction preserves the impact of cooling on total data center loading while maintaining computational efficiency and consistency with the overall component-level modeling framework.

4.7 Modeling of Colocated Power Plant Generation

4.7.1 Modeling of Gas Turbine (GT) Generator

The gas turbine (GT) generator provides local generation support at the medium-voltage (MV) level and contributes to both steady-state power supply and dynamic system response. In the developed model, each GT unit is represented using a detailed synchronous generator integrated with standard control subsystems, including a steam turbine-governor (STG), an excitation system, and a power system stabilizer (PSS).

The synchronous machine block represents the electromechanical energy conversion process, where mechanical input power from the turbine is converted into electrical output power delivered to the grid. The rotor dynamics capture the inertia-driven response of the generator, which plays a key role in system frequency behavior and oscillation characteristics.

The STG system provides primary frequency control. It adjusts the mechanical input power based on speed deviation through a droop mechanism, ensuring that the generator responds to frequency variations and contributes to system stability. The turbine stages capture the inherent delays and dynamics associated with steam energy conversion.

The excitation system regulates the terminal voltage of the generator by controlling the field voltage. An IEEE Type-1 excitation model is employed, where the voltage regulator compares the measured terminal voltage with a reference value and adjusts the field excitation accordingly. This control loop is critical for maintaining voltage stability and reactive power support.

To enhance damping of low-frequency electromechanical oscillations, a Kundur-type power system stabilizer (PSS) is included. The PSS introduces a stabilizing signal into the excitation system based on speed or power deviations, improving system damping and reducing oscillatory behavior.

Overall, the GT generator model captures the key electromechanical and control dynamics required for system-level stability and oscillation studies while maintaining compatibility with the benchmark network framework.

Detailed mathematical formulations, control block structures, and parameter settings of the synchronous machine, turbine-governor, excitation system, and PSS are provided in Appendix A.1.

4.7.2 Modeling of Battery Energy Storage System (BESS)

The battery energy storage system (BESS) provides fast dynamic support at the MV level and plays a critical role in enhancing system stability during disturbances. The BESS is modeled as a grid-interfaced converter system consisting of a DC energy source (battery) and a power electronic inverter.

The inverter is implemented as a three-level neutral-point-clamped (NPC) converter using a reference-voltage-controlled average model.

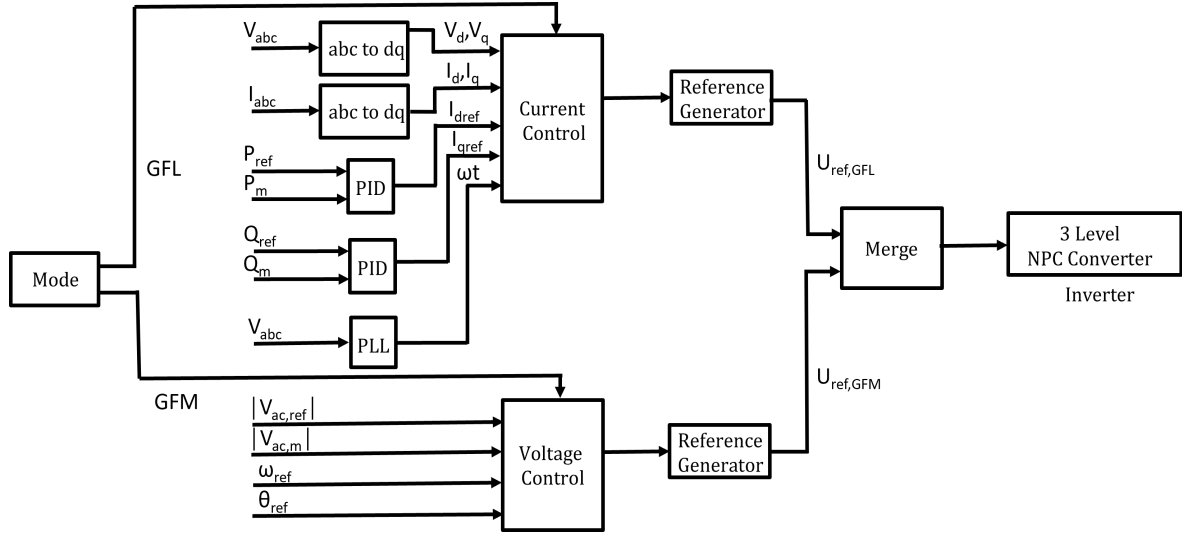


Figure 4.14: Generalized control diagram for BESS Inverter.

The overall control structure of the BESS inverter is illustrated in Fig. 4.14. The inverter operates primarily in grid-following (GFL) mode, where synchronization with the grid is achieved through a phase-locked loop (PLL). The PLL provides the reference angle needed to synchronize the inverter output with the grid voltage.

In GFL operation, the control system is structured around inner current control loops and outer control objectives. The inner control loop regulates the inverter current to track reference values, enabling precise control of active and reactive power exchange with the grid. The outer control loop generates these current references based on system requirements such as power injection or DC-link regulation, as shown in Fig. A.2.

For transient and disturbance studies, the control is implemented in the dq reference frame. Grid-side voltages and currents are transformed into dq components, allowing decoupled control of active and reactive power. The controller generates reference signals in the dq frame, which are then converted back to three-phase quantities to produce the inverter reference voltage, as illustrated in Fig. A.1.

Although the primary mode of operation in this study is grid-following, the modeling framework is compatible with grid-forming (GFM) control strategies, where the inverter can regulate voltage and frequency directly. This flexibility allows the BESS to support both grid-following and grid-forming functionalities depending on system requirements.

Overall, the BESS model captures the essential converter control dynamics and fast response characteristics required for studying its interaction with the grid and its impact

on system stability.

Detailed control equations, transformation relationships, and controller implementations are provided in Appendix A.2.

4.8 Protection System Design and Logic/Settings Considering Standards (IEEE 2800)

The protection system of the grid-connected data center is designed to ensure reliable operation under both normal and disturbed grid conditions while complying with modern interconnection standards such as IEEE 1547 and IEEE 2800. In converter-dominated systems, conventional protection schemes based on immediate tripping are no longer sufficient, as large-scale disconnection during disturbances can adversely affect grid stability. Therefore, fault ride-through (FRT) capability is incorporated as a fundamental requirement in the protection design.

4.8.1 Fault Ride-Through (FRT) Philosophy

Fault ride-through refers to the capability of a resource to remain connected and operational during transient voltage and frequency disturbances for a specified duration. Instead of tripping instantaneously, the system is required to tolerate disturbances within predefined voltage and frequency envelopes.

In the developed model, protection and switching elements are implemented at key interfaces, particularly at the point of common coupling (PCC) and within the medium-voltage distribution network, to enable FRT functionality and controlled system response during disturbances.

The FRT logic is structured based on three operational regions defined in IEEE 2800:

- **Continuous Operation Region:** Normal operating range where the system must remain connected indefinitely.
- **Mandatory Operation Region:** Disturbance region where the system must ride through for a minimum specified duration.
- **Permissive Operation Region:** Extreme conditions where operation is allowed but tripping is permitted.

The combined area of these regions defines the must-ride-through zone, whereas conditions outside this envelope correspond to the may-trip zone.

4.8.2 Voltage Ride-Through Requirements (IEEE 2800)

The voltage ride-through characteristics implemented in this work are based on IEEE 2800 requirements for inverter-based resources (IBRs). For the considered data center configuration (converter-dominated load with BESS/UPS), the criteria corresponding to “all other IBR types” are adopted.

As observed from Table 4.5, the system is required to remain connected even under severe voltage depressions down to 0.25 p.u. for short durations. This ensures that

Table 4.5: Voltage Ride-Through Requirements

Voltage (p.u.)	Operation Region	Minimum Ride-Through Time (s)
> 1.20	May Trip	N/A
> 1.10	Mandatory	1.0
> 1.05	Continuous	1800
$0.90 \leq V \leq 1.05$	Continuous	Continuous
< 0.90	Mandatory	6.0
< 0.70	Mandatory	3.0
< 0.50	Mandatory	1.20
< 0.25	Mandatory	0.32
< 0.10	Permissive	0.32

large-scale tripping of converter-based loads is avoided during faults, thereby maintaining system stability.

The voltage is evaluated at the high side of the main transformer using RMS measurements, and the ride-through duration is accumulated over a specified time window.

4.8.3 Frequency Ride-Through Requirements (IEEE 2800)

In addition to voltage disturbances, the system must also tolerate frequency deviations. The implemented frequency ride-through characteristics are summarized in Table 4.6.

Table 4.6: Frequency Ride-Through Requirements (Adapted from IEEE 2800 Table 3)

System Frequency (Hz)	Minimum Ride-Through Time (s)
> 61.8	May Trip
> 61.2	299
$58.8 \leq f \leq 61.2$	Continuous
< 58.8	299
< 57.0	May Trip

The frequency ride-through profile can be interpreted in terms of three distinct operational regions:

1. **Continuous Operation Region ($58.8 \leq f \leq 61.2$ Hz):** This represents the normal operating band of the power system. Within this range, the data center and associated converter-based resources (e.g., BESS and UPS) are required to remain connected indefinitely without any restriction. No protective action is initiated, and normal control operation is maintained. This ensures stable operation under typical grid frequency variations.
2. **Mandatory Ride-Through Region ($57.0 \leq f < 58.8$ Hz and $61.2 < f \leq 61.8$ Hz):**

In this region, the system experiences moderate under-frequency or over-frequency conditions. The resource must remain connected for a minimum duration of 299 seconds (approximately 5 minutes), even though the frequency deviates from its nominal value. This requirement prevents premature disconnection during system disturbances such as generation-load imbalance or transient events. During this interval, control systems may adjust active power output (if applicable) to support grid recovery.

3. **Permissive Tripping Region ($f < 57.0 \text{ Hz}$ and $f > 61.8 \text{ Hz}$):** These frequency levels correspond to severe system conditions. In this region, continued operation is not mandatory, and protective tripping is permitted to prevent equipment damage or instability. However, depending on system design and control capability, the resource may still attempt to ride through if operational limits are not exceeded.

Overall, the frequency ride-through requirement ensures that the system does not disconnect unnecessarily during transient frequency excursions, thereby contributing to grid stability. By enforcing sustained operation in the mandatory region and allowing controlled tripping only under extreme conditions, the protection scheme aligns with IEEE 2800 objectives for converter-dominated power systems.

4.8.4 Protection Logic and Implementation

The protection system integrates FRT requirements into its decision-making logic rather than relying solely on instantaneous threshold-based tripping. The implemented logic follows a time-dependent structure:

1. Measure voltage and frequency at the PCC using filtered RMS quantities.
2. Identify the corresponding operating region (continuous, mandatory, or permissive).
3. Start a timer when the system enters a disturbance region.
4. Compare the elapsed duration with the minimum ride-through requirement.
5. Initiate tripping only if the disturbance duration exceeds the allowable limit.

To ensure compliance with IEEE 2800:

- Instantaneous tripping based on sub-cycle measurements is avoided.
- Duration-based accumulation is implemented for varying disturbances.
- Multiple disturbances within a short interval are treated cumulatively.

4.8.5 Post-Disturbance Recovery and Control Coordination

Following fault clearance, the system is required to restore normal operation in a controlled manner. In accordance with IEEE 2800, the real power output is restored to its pre-disturbance value within a specified time (typically within 1 second) after voltage recovery.

Additionally, converter controls (UPS and BESS) are coordinated to:

- Provide voltage support during faults,
- Maintain DC-link stability,
- Avoid control instability during transitions,
- Ensure smooth reconnection to normal operating conditions.

4.8.6 Summary

The implemented protection system integrates modern FRT requirements into the overall control and protection architecture of the data center. By adhering to IEEE 2800 ride-through criteria, the system ensures:

- Enhanced grid support during disturbances,
- Prevention of unnecessary tripping,
- Improved system stability and resilience,
- Compatibility with converter-dominated power system operation.

5.0 Study Framework for Impact Assessment of Data Center on Power System Dynamics

5.1 Test System Details

This chapter presents simulation-based studies conducted on a comprehensive grid-connected data center test system. The test system integrates a detailed data center power architecture with a benchmark high-voltage transmission network, enabling assessment of dynamic interactions, stability behavior, and power-quality performance. The configuration of the test system is derived from the data center AC power architecture described in Section 3.1. The following subsection describes the overall test system configuration and its integration with the utility grid and co-located power resources.

5.1.1 Overall Test System Configuration and Data Center Integration

The test system integrates a detailed grid-connected data center model with the Kundur two-area, four-machine benchmark power system. The overall configuration consists of three interconnected layers:

- The high-voltage (HV) transmission network.
- The medium-voltage (MV) interconnection and co-located power resources.
- The low-voltage (LV) data center electrical system.

The HV transmission network is represented using the Kundur two-area, four-machine benchmark system, which provides a well-established platform for studying inter-area electromechanical dynamics and system-level stability. The single-line diagram of the original Kundur system is shown in Fig. 5.1. The system consists of two interconnected areas linked through parallel 230 kV transmission lines forming the primary power-transfer corridor. Each area contains two synchronous generators rated at 900 MVA and connected to the transmission network through step-up transformers.

Under nominal operating conditions, approximately 400 MW of power is transferred between the two areas, producing the characteristic low-frequency inter-area oscillation mode in the range of approximately 0.6–0.8 Hz. This characteristic makes the Kundur system particularly suitable for evaluating the influence of large converter-interfaced data center loads on oscillatory behavior and damping performance.

For data center integration, the Kundur system serves as the upstream utility grid. The data center is connected in Area 1 at the transmission level, as illustrated in Fig. 5.2. To accommodate the data center without altering the benchmark operating point, a portion of the original aggregated load in Area 1 is replaced by the explicit data center load while preserving the overall system loading condition. This maintains the steady-state power-flow condition and generator operating points of the benchmark system, allowing the dynamic effects introduced by the data center to be evaluated without major changes in the original system loading.

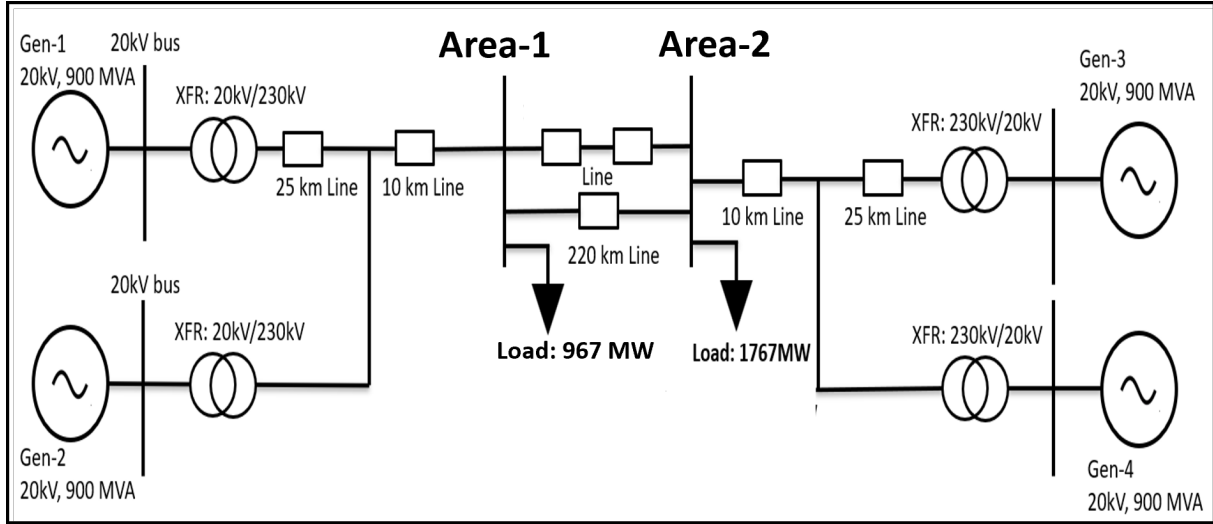


Figure 5.1: Single Line Diagram of Kundur Two-Area System.

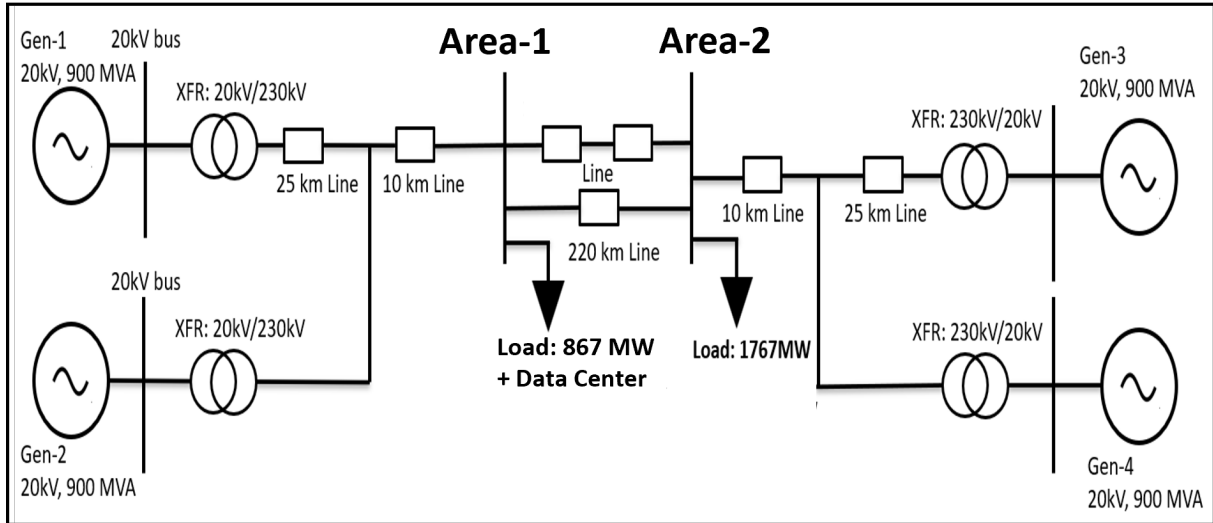


Figure 5.2: Kundur Two-Area System Connected with Data Center.

The data center is interfaced with the utility grid through a hierarchical voltage structure consisting of:

- HV level: 230 kV transmission network (Kundur system),
- MV level: 13.8 kV distribution bus (PCC),
- LV level: 480 V AC distribution within the data center, and
- DC level: approximately 800 V DC link within the UPS.

Power from the 230 kV transmission network is stepped down to the 13.8 kV MV bus through transformer stages. The 13.8 kV bus serves as the point of common coupling (PCC) between the utility grid and the data center, where voltage magnitude, frequency,

and active/reactive power exchange are monitored. Co-located resources, including a gas turbine (GT) generator and a battery energy storage system (BESS), are connected at the MV level to provide local generation and dynamic support.

The detailed internal electrical configuration of the grid-connected data center is shown in Fig. 5.3.

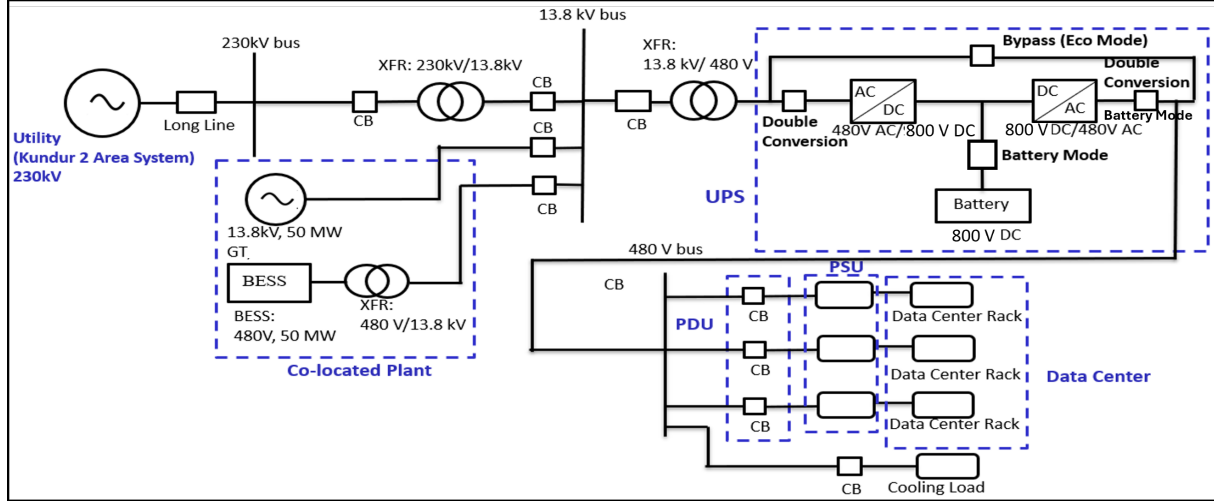


Figure 5.3: Single Line Diagram of Grid-Connected Data Center Model.

From the 13.8 kV MV bus, power is stepped down to 480 V AC, forming the primary low-voltage distribution level within the data center. The Uninterruptible Power Supply (UPS) interfaces the upstream AC system with the downstream data center loads and incorporates an internal DC link operating at approximately 800 V. The UPS supports bypass, double-conversion, and battery-supported operating modes, as described in Section 3.5.

Downstream of the UPS, conditioned 480 V AC power is distributed through Power Distribution Units (PDUs) to the rack-level Power Supply Units (PSUs), which convert the AC supply to the regulated low-voltage DC required by the IT equipment. Cooling systems are supplied from the AC distribution network and are incorporated into the overall data center load according to the modeling approach described in Section 3.6. This configuration therefore represents the complete electrical pathway from the transmission network through the MV and LV distribution stages to the converter-interfaced IT loads and cooling systems.

Protection and switching elements are included at key interfaces, particularly at the PCC and within the MV distribution network, to enable fault ride-through (FRT) functionality and controlled operation under grid disturbances. These elements also facilitate transitions among the different data center operating conditions considered in the subsequent case studies.

All four generators in the Kundur system are represented using detailed synchronous machine models with associated excitation systems, turbine-governor dynamics, and power system stabilizers. Transmission lines are represented using distributed-parameter models suitable for electromechanical transient analysis, while transformers include their

corresponding impedance and tap settings. This modeling framework preserves the fundamental dynamic characteristics of the original Kundur system while enabling investigation of data center–grid interactions.

The complete test system is implemented in MATLAB/Simulink using a discrete-time simulation framework with a fixed simulation time step of $50 \mu\text{s}$. The selected time step provides compatibility with the converter and control dynamics represented within the detailed data center model while enabling system-level transient and oscillatory studies.

Overall, the integrated test system combines the established dynamic characteristics of the Kundur two-area transmission network with the detailed data center architecture developed in this work. The resulting framework enables systematic evaluation of data center load dynamics, grid stability, oscillatory behavior, power-quality performance, and control interactions under various operating and disturbance conditions. This configuration serves as the foundation for the case studies and discussions presented in the subsequent sections of this chapter.

5.2 Case Study Framework for Impact Assessment of Data Center on Power System Dynamics

5.2.1 Modal Characteristics and Harmonic Analysis under Different Load Conditions

This section presents the simulation outcomes of the oscillation-mode identification study conducted on the Kundur two-area benchmark system under multiple operating conditions and disturbance events. The primary objective is to evaluate how different loading condition and also data center load influence the dominant oscillatory behavior. Oscillation modes are extracted using the sub-space identification based workflow, and the resulting frequencies and damping ratios are summarized in the tables below.

5.2.1.1 Benchmark Loading Condition

To validate the implemented dynamic model and the oscillation analysis framework, a benchmark loading condition based on the well-established Kundur two-area system is adopted as a reference case. This test system is widely used in power system stability studies for evaluating inter-area oscillatory behavior [20].

Under the nominal loading condition, a small disturbance is introduced in the system to excite its natural electromechanical modes, resulting in a ring-down response. The system dynamics are captured using generator rotor angle (δ) and rotor speed deviation ($\Delta\omega$) signals, which are commonly used for oscillation analysis. These time-domain responses are then processed to extract modal information using the adopted identification method.

Modal analysis is performed on the resulting ring-down signals to identify the dominant oscillation modes, including their frequencies and damping ratios. The extracted modal characteristics are summarized in Table 5.1.

As observed, the dominant inter-area mode appears around 0.65 Hz with relatively low damping, which is consistent with the classical Kundur two-area benchmark system. In addition, two well-damped local modes are identified in the range of 1.1–1.2 Hz, corre-

Table 5.1: Identified oscillation modes for benchmark loading condition

Case	Mode (Hz)	Damping (%)	THD (%)
Benchmark Loading	0.65	2.13	0.7
	1.12	11.50	
	1.17	9.08	

sponding to oscillations within individual areas.

The close agreement of the identified modal characteristics with reported benchmark results confirms the correctness of the system modeling and provides a reliable baseline for further analysis under modified loading conditions.

5.2.1.2 Constant Data Center Load

To evaluate the impact of a steady data center demand on system dynamics, a constant loading condition is considered in which a fixed 100 MW data center load is connected to the system. This scenario represents a baseline operating condition where the IT load remains time-invariant, without any temporal variability or workload-driven fluctuations.

The data center load is modeled as an aggregated constant power demand connected at the selected bus (Area-1), representing the combined consumption of IT equipment and associated infrastructure. To maintain consistency with the benchmark operating point, this 100 MW data center load is introduced in place of the original 100 MW AC load in Area-1, ensuring that the overall system power flow and loading level remain unchanged.

To excite the system dynamics, a small disturbance is introduced, resulting in a ring-down response. The system behavior is monitored using generator rotor angle (δ) and rotor speed deviation ($\Delta\omega$) signals, which are commonly used for electromechanical oscillation analysis. These time-domain responses are processed using the adopted modal identification framework to extract the dominant oscillatory modes.

The identified modal characteristics corresponding to this loading condition are summarized in Table 5.2.

Table 5.2: Identified oscillation modes for constant data center loading condition

Case	Mode (Hz)	Damping (%)	THD (%)
Constant IT Load with event	0.64	2.0	2.2
	0.90	1.12	
	0.83	7.3	

5.2.1.3 Time Varying AI Training Load

To evaluate the impact of realistic workload variations on system dynamics, a time-varying data center loading condition is considered based on an AI training load profile. This profile captures the temporal variations in power demand associated with high-performance computing workloads, as illustrated in Fig. 4.8b.

The data center load is modeled as a dynamic aggregated demand connected at the selected bus (Area-1), representing the combined behavior of IT equipment under varying computational intensity. Similar to the constant loading case, the time-varying AI training load replaces the original 100 MW AC load in Area-1, ensuring that the overall system loading level remain consistent.

To excite the system dynamics, a small disturbance is introduced, resulting in a ring-down response under time-varying loading conditions. The system behavior is monitored using generator rotor angle (δ) and rotor speed deviation ($\Delta\omega$) signals. These time-domain responses are processed using the adopted modal identification framework to extract the dominant oscillatory modes. The identified modal characteristics corresponding to this loading condition are summarized in Table 5.3.

Table 5.3: Identified oscillation modes for time-varying AI training data center loading condition

Case	Mode (Hz)	Damping (%)	THD (%)
AI profile (Fig. 4.8b) of IT load	0.64	0.79	2
	0.995	0.74	
	1.15	9.58	
	1.85	2.86	

5.2.1.4 Result Discussion

The modal characteristics obtained under constant and time-varying data center loading conditions provide important insights into the impact of data center operational behavior on power system dynamics.

- Shift in dominant oscillation modes: For both constant and time-varying data center loading conditions, the dominant inter-area oscillation mode remains around 0.64–0.65 Hz, indicating that the fundamental inter-area dynamics of the system are preserved. However, the higher-frequency local modes exhibit noticeable shifts under both loading conditions compared to the benchmark case, suggesting altered local electromechanical interactions due to data center integration.
- Excitation of new oscillation modes under time-varying load: Unlike the benchmark and constant loading cases, which exhibit three dominant modes, the time-varying AI training load results in the appearance of an additional oscillation mode, leading to four identified modes. This indicates that the repetitive and dynamic nature of the AI load profile introduces additional excitation into the system, giving rise to new oscillatory components.
- Reduction in damping characteristics: A reduction in damping is observed for the dominant inter-area mode under both loading conditions compared to the benchmark system. For the ~ 0.64 Hz mode, the damping decreases from 2.13% (benchmark) to 2.0% (constant load) and further to 0.79% under time-varying AI load,

corresponding to an approximate 63% reduction relative to the benchmark case. In addition, other modes also exhibit noticeable reductions in damping under time-varying conditions, indicating a broader degradation of system damping and reduced ability to attenuate oscillations.

- **Increase in harmonic distortion due to data center load:** The integration of data center loads leads to a significant increase in harmonic distortion levels. The benchmark system exhibits a low THD of approximately 0.7%, which increases to about 2.2% under constant data center loading and remains elevated at around 2% for the time-varying AI training load. This increase is primarily attributed to the presence of power-electronic interfaces within the PSU and UPS, which introduce nonlinearities and switching harmonics into the system.

Overall, the comparison demonstrates that while the constant data center load behaves as a steady aggregated demand with relatively predictable dynamic characteristics, the time-varying AI training load introduces continuous power fluctuations that act as persistent disturbances to the system. These variations lead to increased modal complexity, frequency shifts, reduced damping, and higher harmonic distortion, reflecting stronger interactions between generators and converter-based load dynamics. Consequently, although the dominant inter-area mode is preserved, the system exhibits a more dynamic and potentially less stable response under time-varying loading conditions. These findings highlight the importance of incorporating realistic, time-varying load profiles in stability assessments and emphasize the need for enhanced dynamic analysis and control strategies in future power systems with high penetration of data center loads.

5.2.2 Modal Characteristics and Harmonic Analysis under Voltage Disturbance, and Grid–UPS Mode Transition

5.2.2.1 Test System & Event Details

The study is conducted using the standard Kundur two-Area power system in Fig. 5.2, connected to the data center in Fig. 5.3. To evaluate the dynamic interactions between the power system and the data center load, three disturbance scenarios are considered for this study as follows:

- **Temporary Voltage Sag:** A short-duration voltage sag of 0.75 pu is applied for three cycles at the middle of the tie line at the grid side at $t = 15$ s for a constant data center load study and at $t = 16$ s for an AI training data center load profile study. During this event, the data center remains connected to the grid, and no grid-to-UPS switching occurs.
- **Grid-to-UPS Transition with Load Loss:** A three-phase fault is applied for three cycles at the Area-1 bus on the grid side at $t = 30$ s, causing the grid voltage to temporarily collapse to zero. The data center immediately transitions to UPS operation. As a result, the grid experiences a temporary loss of the 100 MW data center load until reconnection.
- **UPS-to-Grid Reconnection:** At $t = 50$ s, the data center reconnects from UPS operation back to the utility grid.

These disturbance scenarios allow assessment of the dynamic response of the grid–data center interaction, modal characteristics, and harmonic distortion under voltage disturbances, load disconnection, and reconnection events, with a constant and an AI-trained data center load profile. The test system was conducted in MATLAB/Simulink (50 μs discrete time step).

5.2.2.2 Simulation Results

For the simulation demonstration, two data center load types are used. The first is a constant data center load of 100 MW, and the second is an AI training profile mentioned in [9] shown in Fig. 4.8b. For this study, the frequency, voltage, and active power response at various points are evaluated.

Constant Data Center Load: The dynamic responses of the system under constant data center load and disturbance scenarios are illustrated in Figs. 5.4–5.6, which show the variations in grid voltage and system frequency at Areas 1 and 2 during the considered disturbance events, with the data center load held constant (100 MW). The corresponding frequency deviation characteristics and settling times are summarized in Table 5.4, while the modal characteristics and harmonic distortion associated with these events are summarized in Table 5.5. The percentage change in system frequency due to a disturbance is calculated by summing the maximum rise and drop of the frequency deviation and normalizing it to the nominal system frequency f_{nom} . This is expressed as:

$$\Delta f_{\%} = \frac{f_{\text{rise}} + f_{\text{drop}}}{f_{\text{nom}}} \times 100\%, \quad (5.0.1)$$

where f_{rise} and f_{drop} are the peak positive and negative deviations of frequency, respectively, and f_{nom} is the nominal system frequency (60Hz).

A constant data center load is used as a full-scale dynamic model of the data center, with a constant AI training profile, for this simulation.

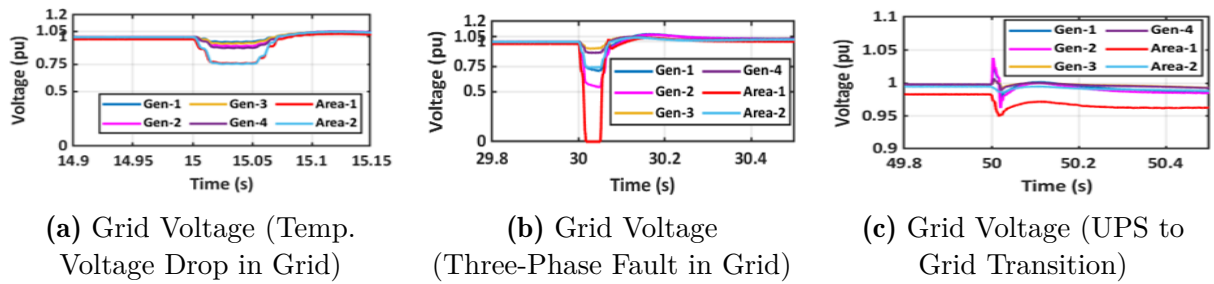


Figure 5.4: Grid Voltage with constant data center load.

AI Training Data Center Load Profile: The dynamic responses of the system under the AI training data center load profile are illustrated in Figs. 5.7–5.9. These figures present the variations of grid voltage and system frequency at Areas 1 and 2 under the considered disturbance events, with the AI training data center load profile. The corresponding frequency deviation characteristics and settling times are summarized in Table 5.4, while

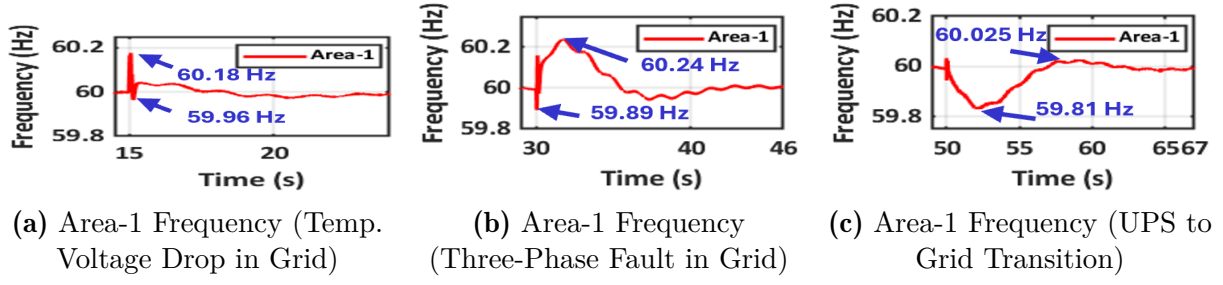


Figure 5.5: Area-1 (Bus 7) Frequency with constant data center load.

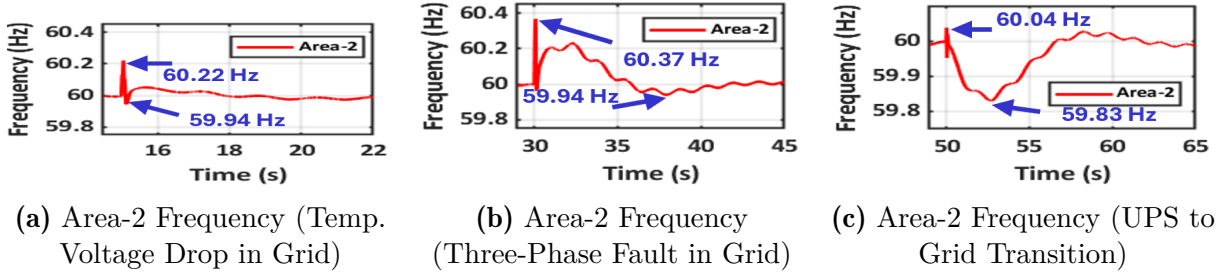


Figure 5.6: Area-2 (Bus 9) Frequency with constant data center load.

the modal characteristics and harmonic distortion under these scenarios are summarized in Table 5.5.

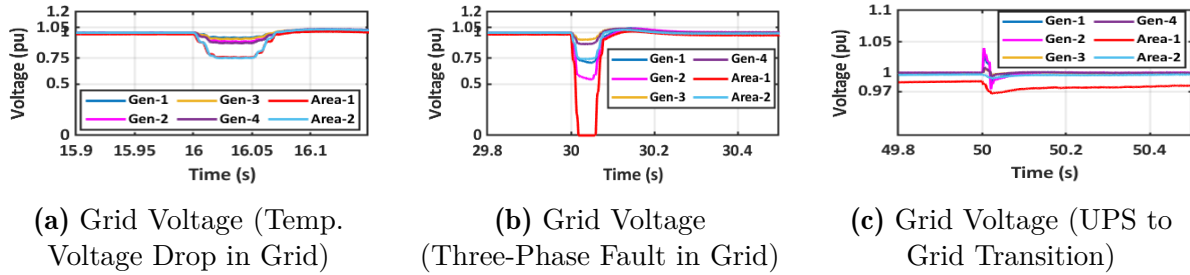


Figure 5.7: Grid Voltage with AI Training Data Center Load.

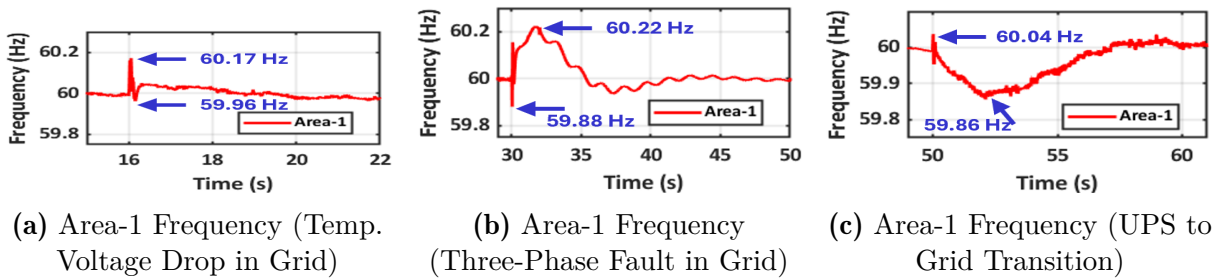


Figure 5.8: Area-1 (Bus 7) Frequency with AI Training Data Center Load Profile.

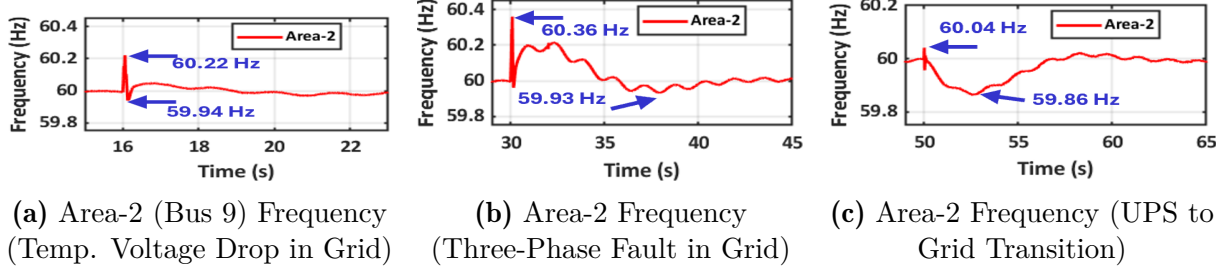


Figure 5.9: Area-2 (Bus 9) Frequency with AI Training Data Center Load Profile.

Table 5.4: Frequency dynamic characteristics under different data center load profiles and event scenarios

Data Center Load	Parameters	Event Scenario		
		Temporary Voltage Drop at Grid	Grid to UPS + Load Loss	UPS to Grid Reconnection
Constant Data Center Load	Area-1: Frequency Deviation (R-D)	0.18–0.04 Hz (0.37%)	0.24–0.11 Hz (0.58%)	0.025–0.19 Hz (0.36%)
	Area-1: Settling Time	5 s	11 s	12 s
	Area-2: Frequency Deviation (R-D)	0.22–0.06 Hz (0.47%)	0.36–0.07 Hz (0.72%)	0.04–0.14 Hz (0.30%)
	Area-2: Settling Time	4.5 s	10 s	12 s
AI Training Data Center Profile	Area-1: Frequency Deviation (R-D)	0.17–0.04 Hz (0.35%)	0.22–0.12 Hz (0.57%)	0.04–0.14 Hz (0.30%)
	Area-1: Settling Time	4 s	10 s	7 s
	Area-2: Frequency Deviation (R-D)	0.22–0.06 Hz (0.47%)	0.36–0.07 Hz (0.72%)	0.04–0.14 Hz (0.30%)
	Area-2: Settling Time	4 s	10 s	7 s

Note: R $\Rightarrow$ Rise and D $\Rightarrow$ Drop.

Table 5.5: Modal characteristics and harmonic distortion under different data center load profiles and event scenarios

Parameter	Load Condition	Event Scenario			
		Standard Kundur System with Ring Down	Temporary Voltage Drop at Grid	Grid to UPS + Load Loss	UPS to Grid Reconnection
Mode (Hz)	Constant Data Center	0.65, 1.13, 1.18	0.64, 1.11, 1.15	0.63, 1.10, 1.27	0.64, 1.11, 1.29
	AI Training Profile	(Spot Load)	0.64, 0.97, 1.13, 2.30	0.63, 1.10, 1.27	0.65, 1.01, 1.83, 3.24
Damping Ratio (% ζ)	Constant Data Center	2.3, 10.85, 8.32	2.2, 9.88, 8.34	2.4, 10.7, 0.6	1.9, 9.94, 0.55
	AI Training Profile	(Spot Load)	1.41, 2.1, 8.46, 2.65	1.95, 10.81, 1.22	1.1, 3.01, 3.25
THD (%)	Constant Data Center	0.7	2.2	3.58	3.38
	AI Training Profile	(Spot Load)	2	3.07	2.53

Note: indicates mode shifting, and indicates newly introduced modes.

5.2.2.3 Result Discussions

Frequency Dynamic Response Under Disturbance Events: The frequency deviation and settling characteristics summarized in Table 5.4 highlight several important observations regarding the dynamic impact of data center load disturbances on the power system.

- Maximum frequency deviation: Area-2 consistently demonstrates the largest frequency deviation across all scenarios. During the Grid-to-UPS transition, a load-loss event results in a frequency deviation of approximately 0.36 Hz. These findings suggest that Area-2 undergoes the most significant transient response during disturbances.
- Most severe disturbance event: Among the three disturbance scenarios, the Grid-to-UPS transition with load loss produces the largest frequency excursions across the

system, indicating that the sudden disconnection of the data center load, along with the Grid-to-UPS transition, introduces the most significant transient imbalance.

- Settling time characteristics: The longest system recovery times are also observed during the Grid-to-UPS transition following a load-loss event, with system stabilization occurring in approximately 10–11 s.

Overall, the Grid-to-UPS transition with load loss represents the disturbance condition that produces the most pronounced system frequency dynamics, with Area-2 experiencing the largest deviations among the studied locations.

Oscillation Modes and Harmonic Characteristics: The modal characteristics and harmonic distortion summarized in Table 5.5 provide insight into the oscillatory behavior and power-quality impacts associated with data center load integration.

- Baseline modes of oscillation: The third column in Table 5.5 provides oscillation modes of the Standard Kundur system with a ring-down event. During this event, a fault (ring-down event) is initiated, and the oscillation modes are evaluated. For this, a spot $P - Q$ load is used instead of the dynamic model of the data center and data center load. The modes of oscillation are being seen (including 0.64 Hz, 1.11 Hz, and 1.15 Hz).
- Shift in dominant oscillation modes: Compared to the original two-area model without data center load (named as the baseline Kundur system), shifts in oscillation frequencies are observed after integrating the data center load. The dominant inter-area mode near 0.65 Hz varies between approximately 0.63 and 0.65 Hz across disturbance scenarios. In addition, several existing modes above 1 Hz shifts (such as 1.11 Hz shift to approximately 0.97 Hz and 1.01 Hz, 1.15 Hz shifts to 1.13 Hz, 1.29 Hz shifts to 1.83 Hz) as shown in Table 5.5, due to the effect of converter-based data center load dynamic characteristics with varying grid conditions.
- Emergence of additional oscillation modes: Additional oscillation modes appear when the system is exposed to the AI training data center load profile under disturbances. Higher-frequency oscillations at 2.3 Hz and 3.24 Hz appear during disturbances, which is associated with interactions among converter-based data center equipment, disturbance transitions, and the rapidly varying AI training load pattern.
- Reduction in damping of the system: Under the data center load condition, particularly under the AI training load profile, a noticeable reduction in the damping of the dominant inter-area oscillation mode has been observed. For example, the damping ratio of the 0.65 Hz inter-area mode reduces from 2.3% to about 1.1% (approximately 52%) under the considered disturbance scenarios.
- Increase of Harmonic distortion: The presence of converter-based data center loads introduces harmonics as summarized in Table 5.5 compared with the baseline Kundur system due to the power-electronic interfaces associated with data center loads.

These observations indicate that while the dominant inter-area oscillation mode of the system remains largely preserved, the integration of data center loads introduces notable impacts on both frequency dynamics and oscillatory behavior. From the frequency response perspective, the Grid-to-UPS transition with load loss produces the most severe disturbance, with Area-2 experiencing the largest frequency deviation (up to ~ 0.36 Hz) and the longest settling times (approximately 10–11 s), indicating a significant transient imbalance.

From the modal perspective, shifts in higher-frequency modes and the emergence of additional oscillation components under time-varying AI load highlight the influence of converter-based load behavior and continuous load variability. A reduction in damping is observed across disturbance scenarios, particularly for the dominant ~ 0.64 Hz mode, indicating reduced oscillation attenuation capability and tighter stability margins. In addition, harmonic distortion increases from approximately 0.7% in the benchmark case to about 2–3.5% under data center loading, reflecting the impact of power-electronic interfaces within the PSU and UPS.

Overall, these results demonstrate that time-varying data center loads increase modal complexity, degrade damping, elevate harmonic content, and intensify transient frequency responses, leading to a more dynamic and potentially less stable system operation.

6.0 Engineering Insights and Practical Implications

6.1 Introduction

Unlike conventional loads, large-scale data centers exhibit highly dynamic behavior due to rapid computational workload variations, extensive power-electronic interfaces, UPS operating-mode transitions, and interactions with co-located generation and energy storage. Their large power demand and these distinct dynamic characteristics can significantly influence grid response, making their behavior relevant to both system planning and real-time operation. This chapter extends the technical findings of the developed modeling framework to broader engineering considerations for utilities, system planners, data center operators, and other entities involved in large-load integration.

6.2 Implications for Grid Stability and Planning

The increasing size and dynamic complexity of data centers require explicit consideration of their characteristics in grid planning and stability studies. Unlike conventional loads, data centers incorporate extensive power-electronic interfaces, rapidly varying computational workloads, UPS-based transfer and ride-through functions, and potentially co-located generation and energy storage. As demonstrated in this work, these characteristics can influence frequency response, oscillatory modes, damping, and power quality. NERC has similarly identified limitations in existing planning and modeling practices for emerging large loads and emphasized the need for improved representation of their dynamic behavior [21, 22]. The developed framework can therefore support the following planning applications.

6.2.1 Interconnection and Transmission Planning

The developed model can support interconnection and transmission planning studies by evaluating different data center capacities, points of interconnection, and operating conditions. Its component-based structure allows assessment of the combined effects of the UPS, PSU, IT workload, cooling system, and co-located resources at the point of common coupling, rather than representing the entire facility as a single static load. This is consistent with NERC’s identification of gaps in current large-load interconnection practices related to dynamic modeling, disturbance ride-through, ramping, power quality, protection settings, and oscillatory behavior [21]. Such studies can help determine required grid reinforcements, interconnection requirements, and appropriate operating limits before a data center is interconnected or expanded.

6.2.2 Dynamic Stability Assessment

The framework can also support dynamic stability assessment under both grid disturbances and time-varying data center operation. The results show that data center integration can alter modal characteristics and damping, while rapidly varying AI work-

loads can introduce additional oscillatory components. Therefore, stability assessment should consider not only conventional voltage and frequency performance but also oscillation damping, converter-driven interactions, and power-quality effects. NERC has similarly identified rotor-angle, frequency, voltage, converter-driven, resonance, and forced-oscillation phenomena as relevant stability considerations for emerging large loads [21]. The developed model provides a platform to identify these interactions and determine whether additional detailed studies or mitigation measures are required.

6.2.3 Scenario-Based Planning and What-If Analysis

The developed framework also supports “what-if” studies because a data center’s grid impact can vary with workload, system condition, UPS operating mode, protection settings, and the availability of co-located resources. Relevant scenarios include different data center penetration levels, constant and time-varying AI workloads, peak and off-peak grid conditions, voltage disturbances, partial or complete load loss, Grid-to-UPS transfer, UPS-to-Grid reconnection, and different operating conditions of local generation and BESS. NERC specifically emphasizes the importance of scenario selection, noting that large-load impacts may become more significant during off-peak conditions when system loading and generation dispatch differ from conventional peak-load planning cases [21]. Evaluating these scenarios can help planners identify adverse conditions in advance and support decisions regarding transmission upgrades, local resources, protection and control settings, and acceptable dynamic performance.

Overall, the developed framework provides a practical platform for evaluating large data center interconnections from planning, stability, and scenario-based perspectives. Such assessments can help identify potential grid impacts in advance and support reliable large data center integration.

6.3 Challenges in Large Data Center Integration

Large-scale data center integration introduces challenges beyond supplying the required steady-state power. Rapidly varying IT workloads, power-electronic interfaces, UPS and protection controls, and the large magnitude of aggregated demand can create dynamic interactions with the surrounding grid. NERC has similarly identified ramping, disturbance ride-through, oscillations, power quality, and protection coordination as important reliability concerns for emerging large loads [21]. Based on the findings of this work, several challenges are particularly relevant.

6.3.1 Oscillatory Stability

The results of this study show that data center integration can alter system damping and introduce additional oscillatory components, particularly under time-varying AI workloads. At higher penetration levels or under weaker grid conditions, interactions among UPS and PSU controls, BESS converters, synchronous generators, and network dynamics may become increasingly important. NERC has also identified forced oscillations, converter-driven stability, and resonance stability as emerging concerns associated with large power-electronic loads [21]. Therefore, oscillation monitoring and damping

assessment should be considered in the dynamic evaluation of large data center interconnections.

6.3.2 Load Ramping and Energy Management

Rapid changes in computational workload can produce significant active-power ramps over short time intervals. These variations must be balanced by the grid or by local resources and may increase regulation requirements and generator response, particularly when multiple large loads vary simultaneously. NERC similarly identifies rapid and unpredictable large-load ramping as a potential source of frequency excursions, voltage violations, and stability concerns [21]. Coordinated operation of co-located GT, BESS, UPS batteries, and other flexible resources can help manage these variations; however, their power references and recovery actions should be coordinated to avoid introducing additional power ramps at the grid interface.

6.3.3 Disconnection, Source Transfer, and Reconnection

The response of a data center during and after a disturbance can be as important as its nominal demand. Transfer from the grid to UPS or local generation can appear to the power system as a sudden load reduction, while simultaneous reconnection and load restoration can create a large recovery demand. NERC has identified large-scale disconnection and reconnection as scenarios that are not adequately represented in existing planning and operational assessments and notes that these events can affect frequency, voltage, and system stability [21]. Consistent with this concern, the Grid-to-UPS transition with temporary load loss examined in this work produced the largest frequency deviation and longest settling time among the evaluated disturbance scenarios. Controlled transfer and gradual restoration are therefore important considerations for large data center operation.

6.3.4 Ride-Through and Protection Coordination

Data center protection and UPS controls are primarily designed to protect sensitive equipment and maintain service continuity, whereas transmission protection is designed to preserve grid reliability. Poor coordination between these objectives can cause a normally cleared grid disturbance to trigger unnecessary load reduction or transfer to backup sources. NERC specifically identifies a coordination gap between transmission-side protection and customer-side UPS and ride-through settings and emphasizes the need for improved information exchange between utilities and large-load facilities [21]. Appropriate coordination of ride-through, transfer, and protection settings is therefore necessary to prevent unintended large changes in grid demand during disturbances.

Overall, these challenges demonstrate that the grid impact of a large data center depends not only on its rated power but also on how rapidly its demand changes and how the facility transfers, protects, and restores that demand during disturbances. Addressing these characteristics is important for reliable large-scale data center integration.

6.4 Need for Advanced Control and Coordination

The dynamic challenges associated with large data centers indicate a need for coordinated control as their size and grid penetration increase. The internal objectives of the facility, including computational availability, equipment protection, and resilience, must be coordinated with grid requirements related to power balance, stability, and disturbance response. NERC similarly emphasizes improved coordination between large-load facilities and grid entities, particularly for ramping, protection, disturbance ride-through, and post-disturbance recovery [21].

6.4.1 Coordinated Facility-Level Control

Coordinated operation of the UPS, BESS, co-located generation, cooling system, and flexible IT load can reduce rapid variations in power exchanged with the grid. For example, BESS can respond to fast changes in IT demand while slower generation resources provide longer-term power support. A supervisory energy-management layer can coordinate these resources while considering battery state of charge, backup-power requirements, and facility operating constraints. Such coordination is particularly important because NERC identifies rapid large-load ramping and limited visibility of facility operating characteristics as potential reliability concerns [21].

6.4.2 Oscillation Detection and Mitigation

Advanced control can also provide a means of mitigating adverse oscillatory behavior. The results of this study demonstrate that time-varying data center demand can affect system damping and introduce additional oscillatory components. Once an adverse mode is detected, controllable resources such as BESS, co-located generation, or flexible IT demand can potentially provide supplementary active-power modulation to improve damping. This is relevant to NERC's identification of oscillation monitoring and forced oscillations as emerging considerations for large-load integration [21].

6.4.3 Disturbance and Transition Coordination

Grid-to-UPS transfer, operation on local resources, and subsequent reconnection should be coordinated to avoid unnecessary transfers and large instantaneous changes in grid demand. Post-disturbance restoration can similarly coordinate IT load recovery, cooling demand, battery charging, and local generation rather than restoring the complete facility demand simultaneously. NERC identifies coordination gaps associated with large-load ride-through, transfer behavior, and the timing and rate of post-disturbance load recovery [21]. Appropriate voltage, frequency, phase-angle, and recovery criteria can therefore support controlled transfer and reconnection.

Based on these considerations, the framework developed in this work can be extended toward a hierarchical detection–decision–control structure, in which abnormal conditions are first identified, an appropriate response is selected, and available data center resources are subsequently coordinated for mitigation. Effective implementation would also require appropriate telemetry, validated dynamic models, and information exchange between the data center and relevant grid entities.

6.5 Role of Data Centers in Future Power Systems

Although large data centers introduce new reliability challenges, their controllable power-electronic interfaces, energy storage, co-located generation, and potentially flexible computational demand also create opportunities for grid-supportive operation. With appropriate controls and coordination, future data centers may therefore evolve from largely passive electricity consumers toward more flexible and grid-responsive facilities.

6.5.1 Flexible Computational Demand

Some computational workloads can be scheduled or adjusted to reduce rapid power variations and peak grid demand. Where operationally feasible, this flexibility can help reduce the grid impact of large data centers.

6.5.2 Grid-Supportive Local Resources

Co-located BESS and generation can further increase data center flexibility. BESS can buffer short-term variations between computational demand and grid power, while local generation can provide longer-duration support. When sufficient converter capacity and stored energy are available, BESS may also support functions such as frequency response and voltage regulation. These resources should be coordinated with UPS operation and facility backup requirements so that grid-support functions do not compromise data center resilience.

6.5.3 Grid-Responsive Data Center Operation

The combination of flexible workload, BESS, UPS, and co-located generation creates the possibility of a grid-responsive data center in which net power at the PCC is actively managed. Under normal conditions, the facility can prioritize computational and energy-management objectives, while during stressed conditions it can potentially limit rapid load changes or coordinate local resources to reduce its impact on the grid. During disturbances, coordinated transfer, ride-through, and restoration can similarly avoid large instantaneous changes in grid demand.

Realizing this role requires accurate dynamic models, validated controls, and coordination between data center and grid operators. NERC similarly emphasizes improved data sharing and coordination for reliable large-load integration [21]. With these capabilities, future data centers can use their flexible loads and local resources to reduce grid impacts while maintaining reliable computational service.

6.6 Summary

Large-scale data centers exhibit highly dynamic behaviors that are driven by several factors (e.g., rapid computational workload variations, power-electronic interfaces, and UPS operations—that significantly impact power grid planning, transmission, and real-time operation). The aggregate demand of these large data centers introduces critical integration challenges (e.g., oscillatory instability, rapid power ramping, complex disconnection and source-transfer events, and protection coordination gaps between facility equipment and transmission systems).

Addressing these reliability concerns requires a multifaceted approach (e.g., advanced, coordinated facility-level control using co-located BESS, local generation, cooling systems, and flexible IT workloads to smooth power exchanges). Implementing hierarchical management frameworks, robust telemetry, and enhanced information exchange between utilities and operators is vital for mitigating adverse grid interactions. Ultimately, with proper dynamic modeling and coordinated controls, future data centers can successfully evolve from passive electricity consumers into flexible, grid-responsive assets that actively support overall power system reliability through local resource management, rigorous scenario planning, and proactive load shaping. The next chapter concludes this report with discussion regarding key findings, contributions offered in this study, remaining limitations, and directions for future work.

7.0 Conclusions and Future Work

7.1 Summary of Key Findings

This report presented a dynamic modeling and simulation framework for evaluating the impact of grid-integrated data centers on power system dynamics. The developed framework incorporated key data center components, including the uninterruptible power supply (UPS), power distribution unit (PDU), power supply unit (PSU), IT load, cooling load representation, co-located gas turbine generation, battery energy storage system (BESS), and fault ride-through protection logic.

The simulation studies demonstrated that data center loads can significantly influence system dynamic behavior. Under constant data center loading, the dominant inter-area oscillation mode of the benchmark system remained close to its original frequency; however, changes in damping and harmonic distortion were observed. Under the time-varying AI training load profile, additional oscillatory components appeared, and the damping of the dominant inter-area mode decreased more noticeably. This indicates that rapidly varying data center workloads can introduce persistent disturbances that affect system stability margins.

The disturbance studies further showed that grid-to-UPS transition with temporary data center load loss produced the most severe frequency response among the studied events. This event resulted in larger frequency deviations and longer settling times compared with temporary voltage sag and UPS-to-grid reconnection cases. In addition, harmonic distortion increased under data center loading conditions due to the presence of converter-based interfaces such as PSUs and UPS systems.

Overall, the results confirm that accurate dynamic modeling of data center loads is essential for assessing their impact on voltage response, frequency dynamics, oscillatory modes, damping characteristics, and power quality.

7.2 Contributions of the Study

The main contributions of this study are summarized as follows:

- A utilization-based IT load model was implemented to represent workload-driven data center power demand, including time-varying AI training load behavior.
- A PUE-based cooling load representation was incorporated to capture the aggregate effect of cooling demand on total data center power consumption.
- A modular data center dynamic modeling framework was developed, including data center components, PDU, PSU, IT load with considering cooling load, integrated with grid as well as co-located plant for grid-integrated data center power systems.
- Modal analysis and harmonic analysis were performed under benchmark, constant data center load, and AI training load conditions to evaluate the impact of data center operation on system dynamics.

- Disturbance case studies were conducted to assess the effects of voltage sag, grid-to-UPS transition, load loss, and UPS-to-grid reconnection on frequency response, oscillatory behavior, and harmonic distortion.

7.3 Limitations

Although the proposed modeling framework captures the major electrical and control dynamics of a grid-integrated data center, several limitations remain.

1. First, the cooling system is represented using an aggregated PUE-based formulation rather than detailed motor-driven and thermodynamic subsystem models. Therefore, individual cooling components such as chillers, pumps, cooling tower fans, and CRAH/CRAC units are not dynamically modeled.
2. Second, the IT load model represents aggregated workload-driven power demand and does not explicitly model individual server scheduling, workload migration, or detailed processor-level dynamics. The model captures the aggregate electrical impact of workload variation, but more detailed computational workload behavior may be required for higher-fidelity studies.
3. Finally, the case studies are performed using a benchmark power system. While the Kundur two-area system is suitable for studying electromechanical oscillations and inter-area dynamics, additional validation on larger and more realistic transmission networks is needed to generalize the findings.

7.4 Future Research Directions

A detailed dynamic cooling system model can be developed by explicitly representing motor-driven loads, variable-speed drives, chillers, pumps, and thermal control loops. This would allow more accurate evaluation of the interaction between IT workload variation, cooling demand, and grid dynamics.

The IT load model can also be enhanced by incorporating workload scheduling, server-level utilization diversity, and demand response strategies. Such extensions would enable the analysis of how data center operational controls can be coordinated with grid stability requirements.

Further work may also investigate advanced control strategies for UPS and BESS systems, including grid-forming operation, coordinated voltage and frequency support, and oscillation damping control. These controls may help mitigate the adverse impacts of rapidly varying data center loads on system stability.

Another promising research direction is the development of integrated energy and power management strategies for grid-connected data centers. Future studies may investigate coordinated scheduling and optimization of utility power, co-located generation, battery energy storage systems (BESS), and flexible IT workloads to improve operational efficiency while supporting grid reliability. Such frameworks can incorporate economic dispatch, peak shaving, renewable energy integration, demand response participation, and resilience enhancement under both normal and disturbed operating conditions.

In addition, the proposed framework can be tested on larger transmission systems with multiple data centers connected at different buses. This would provide broader insight into the cumulative impact of high data center penetration on inter-area oscillations, voltage stability, frequency response, and harmonic distortion.

Finally, future research can explore mitigation strategies such as active power modulation, adaptive load shaping, coordinated BESS control, and grid-supportive data center operation to improve the stability and reliability of future power systems with large-scale data center integration.

Appendix A Detailed Modeling of Co-located Generation Components

A.1 Gas Turbine Generator Detailed Model

In this work, each unit is implemented in MATLAB/Simulink using a synchronous machine model in the dq rotor reference frame, equipped with a steam turbine governor (STG), an IEEE Type-1 excitation system, and a Kundur-type power system stabilizer (PSS). The notation and control block structure follow the standard presentation in [1, 2].

A.1 Synchronous Machine Model (dq-axis)

Each generator is represented by a detailed synchronous machine model suitable for electromechanical transient and small-signal stability studies. Rotor dynamics are governed by the swing equations (Kundur notation):

$$\dot{\delta} = \omega_b(\omega - 1), \quad (\text{A.1})$$

$$2H \dot{\omega} = P_m - P_e - D(\omega - 1), \quad (\text{A.2})$$

where δ is rotor angle (electrical rad), ω is rotor speed in per-unit, ω_b is synchronous base speed, H is inertia constant (s), D is damping coefficient (p.u.), P_m is mechanical power (p.u.), and P_e is electrical air-gap power (p.u.).

The electromagnetic dynamics are modeled in the dq frame using transient internal voltage states consistent with the standard two-axis representation [1]:

$$\dot{E}'_q = \frac{1}{T'_{do}} \left(E_{fd} - E'_q - (X_d - X'_d)I_d \right), \quad (\text{A.3})$$

$$\dot{E}'_d = \frac{1}{T'_{qo}} \left(-E'_d + (X_q - X'_q)I_q \right), \quad (\text{A.4})$$

where E'_q and E'_d are transient internal voltages, E_{fd} is field voltage, I_d, I_q are stator currents, and X_d, X_q, X'_d, X'_q are d/q -axis synchronous and transient reactances. Subtransient and leakage effects are included in the Simulink machine block through the provided reactance set and internal initialization (see Table A.1).

A.2 Steam Turbine and Governor (STG)

Primary frequency control is provided by a droop-based speed governor and a tandem-compound steam turbine model. The governor forms a commanded gate signal based on speed deviation:

$$\Delta\omega = \omega - 1, \quad (\text{A.5})$$

$$u_g = P_{ref} - \frac{1}{R} \Delta\omega, \quad (\text{A.6})$$

where P_{ref} is the reference mechanical power (p.u.) and R is the permanent droop (p.u.). The actuator/servo is modeled as a first-order lag:

$$\dot{x}_g = \frac{1}{T_{sm}}(\text{sat}(u_g) - x_g), \quad (\text{A.7})$$

where x_g is the valve/gate position, T_{sm} is the servo-motor time constant, and $\text{sat}(\cdot)$ applies the physical rate and magnitude limits indicated in Table A.2.

A multi-stage turbine representation is used to capture steam-chest and reheater dynamics. A convenient Kundur-consistent form is:

$$\dot{y}_2 = \frac{1}{T_2}(x_g - y_2), \quad (\text{A.8})$$

$$\dot{y}_3 = \frac{1}{T_3}(y_2 - y_3), \quad (\text{A.9})$$

$$\dot{y}_4 = \frac{1}{T_4}(y_3 - y_4), \quad (\text{A.10})$$

$$\dot{y}_5 = \frac{1}{T_5}(y_4 - y_5), \quad (\text{A.11})$$

$$P_m = F_2 y_2 + F_3 y_3 + F_4 y_4 + F_5 y_5, \quad (\text{A.12})$$

where $T_2 \sim T_5$ are turbine time constants and $F_2 \sim F_5$ are torque fractions (Table A.2). The initial mechanical power P_{m0} is set from the steady-state load flow to match scheduled generation.

A.3 IEEE Type-1 Excitation System

Voltage regulation is implemented using an IEEE Type-1 excitation system. The measured terminal voltage is first filtered:

$$\dot{V}_{t,f} = \frac{1}{T_r}(V_t - V_{t,f}), \quad (\text{A.13})$$

where T_r is the transducer time constant. The regulator error signal is

$$e_v = V_{ref} - V_{t,f} + V_{stab}, \quad (\text{A.14})$$

where V_{ref} is the voltage reference and V_{stab} is the stabilizing signal from the PSS.

The AVR amplifier is modeled as

$$\dot{V}_R = \frac{1}{T_a}(K_a e_v - V_R), \quad V_R \in [E_{fmin}, E_{fmax}], \quad (\text{A.15})$$

where K_a and T_a are regulator gain and time constant, and E_{fmin}, E_{fmax} are limiter bounds.

The exciter dynamics are represented by

$$\dot{E}_{fd} = \frac{1}{T_e} (K_e V_R - E_{fd}), \quad (\text{A.16})$$

with exciter gain K_e and time constant T_e . In your implementation $T_e = 0$ (Table A.3), yielding an effectively algebraic exciter path. Auxiliary transient gain reduction and damping feedback are disabled ($T_b = T_c = 0$, $K_f = 0$), consistent with the mask configuration.

A.4 Power System Stabilizer (Kundur PSS)

A Kundur-type PSS is included to improve damping of electromechanical oscillations by modulating excitation through V_{stab} . The implemented structure follows the conventional washout plus lead-lag compensator form [1]:

$$V_{stab}(s) = K_{pss} \left(\frac{sT_w}{1 + sT_w} \right) \left(\frac{1 + sT_1}{1 + sT_2} \right) \left(\frac{1 + sT_3}{1 + sT_4} \right) U_{in}(s), \quad (\text{A.17})$$

where U_{in} is the stabilizer input (e.g., $\Delta\omega$ and/or accelerating power ΔP_a), T_w is the washout time constant, K_{pss} is stabilizer gain, and (T_1, T_2, T_3, T_4) provide phase compensation. In your Simulink implementation, both $\Delta\omega$ -based and ΔP_a -based stabilizing paths are present and summed into V_{stab} as shown in the generator subsystem.

A.5 Model Parameters (from Simulink Masks)

Tables A.1–A.3 summarize the key parameters used in the Simulink generator, governor, and excitation blocks (as configured in your masks). These parameters are applied consistently across all four machines.

Table A.1: Synchronous machine parameters (per generator) used in the Kundur model implementation.

Parameter	Value
Rated apparent power, S_n	900 MVA
Rated line-to-line voltage, V_n	20 kV
Frequency, f_n	60 Hz
$[X_d, X_q, X'_d, X'_q, X''_d, X_l]$ (p.u.)	[1.8, 3.25, 1.7, 0.55, 0.25, 0.2]
$[T'_{do}, T''_{do}, T'_{qo}, T''_{qo}]$ (s)	[8, 0.03, 0.4, 0.05]
Stator resistance, R_s (p.u.)	0.0025
Inertia and damping set, $[H, F, p]$	[6.5, 0, 4]

Table A.2: Steam turbine and governor (STG) parameters (per generator) used in the Kundur model implementation.

Parameter	Value
Governor gain, droop, dead zone $[K_p, R, D_z]$	$[1, 0.05, 0]$
Speed relay and servo time constants $[T_{sr}, T_{sm}]$ (s)	$[0.001, 0.15]$
Gate opening limits $[v_{gmin}, v_{gmax}, g_{min}, g_{max}]$	$[-0.1, 0.1, 0, 4.496]$
Nominal speed (rpm)	3600
Turbine time constants $[T_2, T_3, T_4, T_5]$ (s)	$[0, 10, 3.3, 0.5]/10$
Torque fractions $[F_2, F_3, F_4, F_5]$	$[0, 0.36, 0.36, 0.28]$
Initial mechanical power, P_{m0} (p.u.)	0.7778

Table A.3: IEEE Type-1 excitation system parameters (per generator) used in the Kundur model implementation.

Parameter	Value
Transducer time constant, T_r (s)	20×10^{-3}
Regulator gain and time constant $[K_a, T_a]$	$[200, 0.001]$
Exciter gain and time constant $[K_e, T_e]$	$[1, 0]$
Transient gain reduction $[T_b, T_c]$	$[0, 0]$
Damping feedback $[K_f, T_f]$	$[0, 0]$
Regulator output limits $[E_{fmin}, E_{fmax}]$ (p.u.)	$[-12.3, 12.3]$
Initial values $[V_{t0}, V_{f0}]$ (p.u.)	$[1, 1.83395]$

A.2 Battery Energy Storage System Detailed Model

The BESS model used for this study is based on a Lithium-Ion battery. The inverter model used in this study for the BESS to convert DC to AC is a reference-voltage-controlled average model of a three-level neutral-point-clamped (NPC) power converter. The general control diagram for the inverter of BESS is shown in Fig. 4.14. GFL mode is considered for BESS for this study. BESS mode is the current control mode for BESS's GFL operation. In this model, for transient events like fault analysis, inverter control involves dq0 transformation in reference voltage generation for the BESS inverter shown in Fig. A.1. The transformation matrix that is required for the ABC to dq0 transformation is as follows:

$$\mathbf{T}_\theta = \frac{2}{3} \begin{bmatrix} \cos(\theta) & \cos(\theta - \frac{2\pi}{3}) & \cos(\theta + \frac{2\pi}{3}) \\ -\sin(\theta) & -\sin(\theta - \frac{2\pi}{3}) & -\sin(\theta + \frac{2\pi}{3}) \\ \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \end{bmatrix}. \quad (\text{A.18})$$

Where θ is the reference angle achieved with PLL block. The formula for the abc to dq0 component is

$$A_{dq0} = \mathbf{T}_\theta A_{abc} \quad (\text{A.19})$$

Where A is the signal to be transformed. The grid-side voltage and current, denoted as V_{abc_prim} and I_{abc_prim} are converted to their dq component using the formula in (A.19) in an abc to dq0 block, and then the dq component of the reference voltage for the inverter are created in the current control block of the BESS model like mentioned in Fig. A.1. By using inverse transformation matrix $\mathbf{T}_\theta^{-1}$, reference sinusoidal voltage is generated for the inverter control shown in Fig. A.1. The formula for the dq0 component to the abc component is as follows:

$$A_{abc} = \mathbf{T}_\theta^{-1} A_{dq0} \quad (\text{A.20})$$

where,

$$\mathbf{T}_\theta^{-1} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 1 \\ \cos(\theta - \frac{2\pi}{3}) & -\sin(\theta - \frac{2\pi}{3}) & 1 \\ \cos(\theta + \frac{2\pi}{3}) & -\sin(\theta + \frac{2\pi}{3}) & 1 \end{bmatrix}. \quad (\text{A.21})$$

The outer loop in Fig. A.2 contributes to generating the reference dq component of current for the generation of the inverter reference voltage dq component in Fig. A.1.

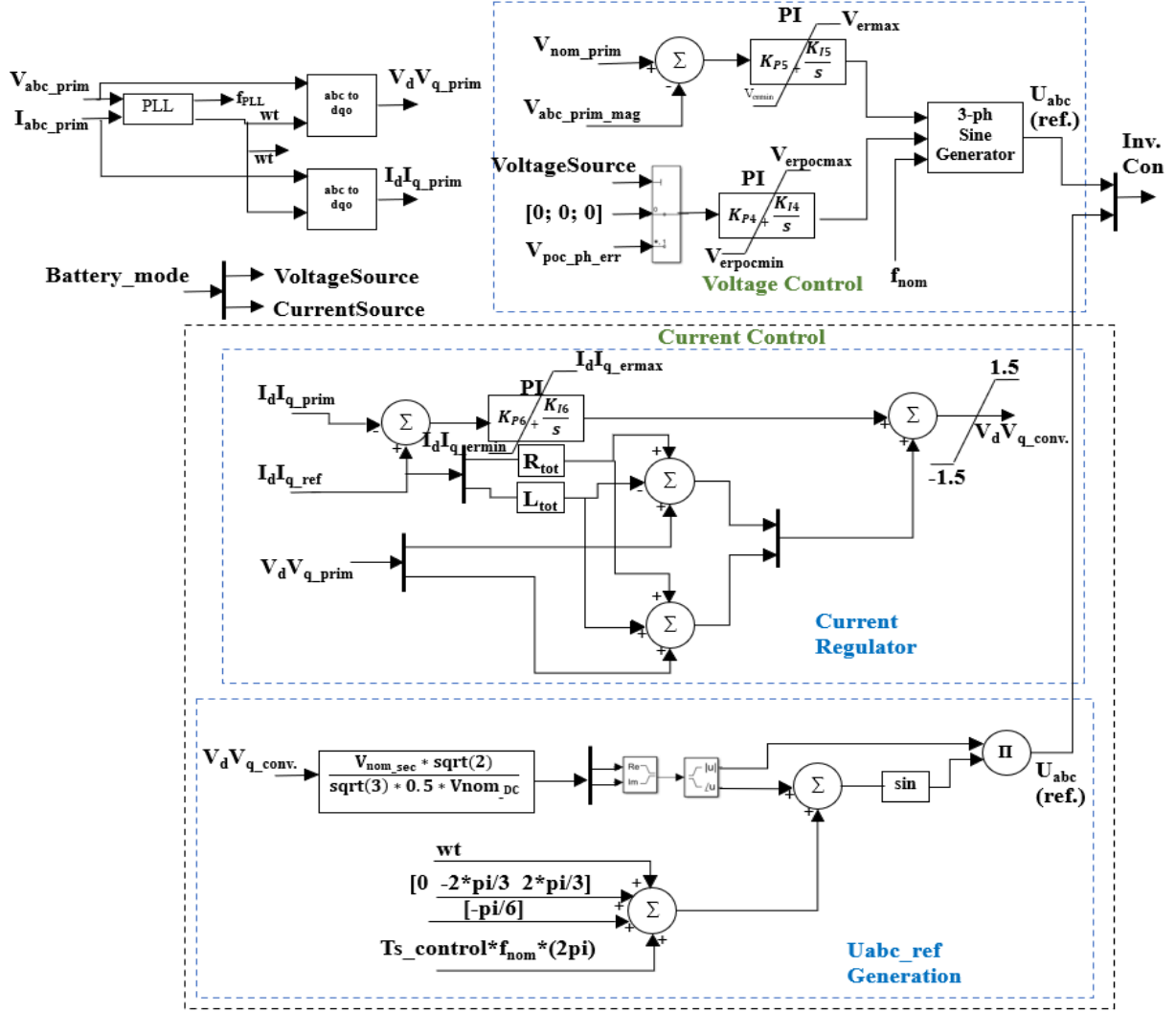


Figure A.1: Detailed control diagram for BESS inverter.

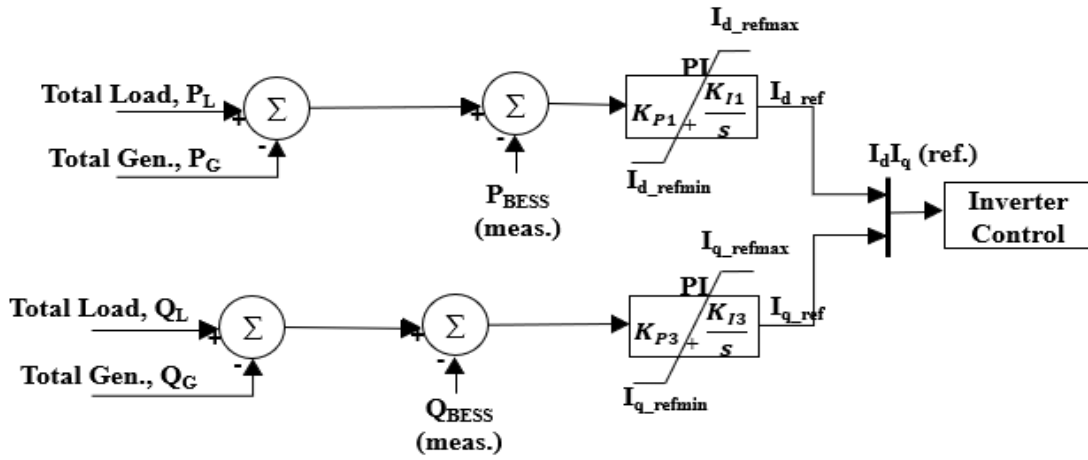


Figure A.2: BESS outer loop.

References

- [1] P. Kundur, *Power System Stability and Control*. New York, NY, USA: McGraw-Hill, 1994.
- [2] MathWorks, “Two-area power system model,” <https://www.mathworks.com/help/sps/ug/two-area-power-system.html>, 2023, accessed: 2025-01-15.
- [3] Alibaba Group, “Alibaba cluster trace program,” <https://github.com/alibaba/clusterdata>, accessed: 2026.
- [4] Siemens, “100 MW Hyperscale AI Blueprint,” <https://resources.sw.siemens.com/sr-RS/brochure-data-center-100-mw-hyperscale-ai-blueprint/>, 2025, accessed: 2026.
- [5] NVIDIA, “Vera Rubin NVL72,” <https://www.nvidia.com/en-us/data-center/vera-rubin-nvl72/>, accessed: 2026.
- [6] S. Li, L. Yang, H. Yu, S. Yao, T. Wu, X. Jiang, H. Lu, K. Wang, C. Wang, S. Xu *et al.*, “Heterogeneity at hyperscale: Characterization and scheduling of large production AI clusters at alibaba (operational systems),” in *20th USENIX Symposium on Operating Systems Design and Implementation (OSDI 26)*, 2026, pp. 2187–2203.
- [7] M. Dayarathna, Y. Wen, and R. Fan, “Data center energy consumption modeling: A survey,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 732–794, 2016.
- [8] K. M. U. Ahmed, M. H. Bollen, and M. Alvarez, “A review of data centers energy consumption and reliability modeling,” *IEEE access*, vol. 9, pp. 152 536–152 563, 2021.
- [9] North American Electric Reliability Corporation, “Characteristics and risks of emerging large loads, Alibaba Cluster Trace Program, Alibaba Group,” <https://github.com/alibaba/clusterdata>, North American Electric Reliability Corporation, Tech. Rep., 2025, accessed: 2026.
- [10] J. Sun, S. Wang, J. Wang, and L. M. Tolbert, “Dynamic model and converter-based emulator of a data center power distribution system,” *IEEE Transactions on Power Electronics*, vol. 37, no. 7, pp. 8420–8432, 2022.
- [11] J. Sun, M. Mihret, M. Cespedes, D. Wong, and M. Kauffman, “Data center power system stability—part ii: System modeling and analysis,” *CSEE Journal of Power and Energy Systems*, vol. 8, no. 2, pp. 420–438, 2022.
- [12] P. Ren, W. Sun, Y. Wang, and G. Harrison, “Grid frequency stability support potential of data center: A quantitative assessment of flexibility,” *IEEE Transactions on Industry Applications*, 2026.

- [13] W. E. Gnibga, A. Blavette, and A.-C. Orgerie, “Renewable energy in data centers: The dilemma of electrical grid dependency and autonomy costs,” *IEEE transactions on sustainable computing*, vol. 9, no. 3, pp. 315–328, 2023.
- [14] N. Rasmussen, “Eco-mode: Benefits and risks of energy-saving modes of ups operation,” *White Paper 157*, 2012.
- [15] I. IEC *et al.*, “Uninterruptible power systems (UPS)-part 3: method of specifying the performance and test requirements,” *Uninterruptible power systems (UPS)-Part*, vol. 3, 2011.
- [16] X. Chen, X. Wang, A. Colacelli, M. Lee, and L. Xie, “Electricity demand and grid impacts of ai data centers: Challenges and prospects,” *arXiv preprint arXiv:2509.07218*, 2025.
- [17] R. E. McFarlane, “Ashrae standards and practices for data centers,” *Data Center Handbook: Plan, Design, Build, and Operations of a Smart Data Center*, pp. 175–191, 2021.
- [18] P. P. Gyang, P. Chakraborty, L. Meegahapola, and X. Yu, “Dynamic modeling of a data center for power system stability studies,” *IEEE Transactions on Power Systems*, 2025.
- [19] R. Rahmani, I. Moser, and M. Seyedmahmoudian, “A complete model for modular simulation of data centre power load,” *IEEE Transactions on Automation Science and Engineering*, 2017.
- [20] P. Kundur, N. J. Balu, and M. G. Lauby, *Power System Stability and Control*. New York: McGraw-Hill, 1994, vol. 7.
- [21] North American Electric Reliability Corporation, “Assessment of gaps in existing practices, requirements, and reliability standards for emerging large loads,” North American Electric Reliability Corporation, Atlanta, GA, USA, NERC Large Loads Working Group White Paper, Mar. 2026.
- [22] —, “Data center load modeling technical reference,” North American Electric Reliability Corporation, Atlanta, GA, USA, Technical Reference, 2026.