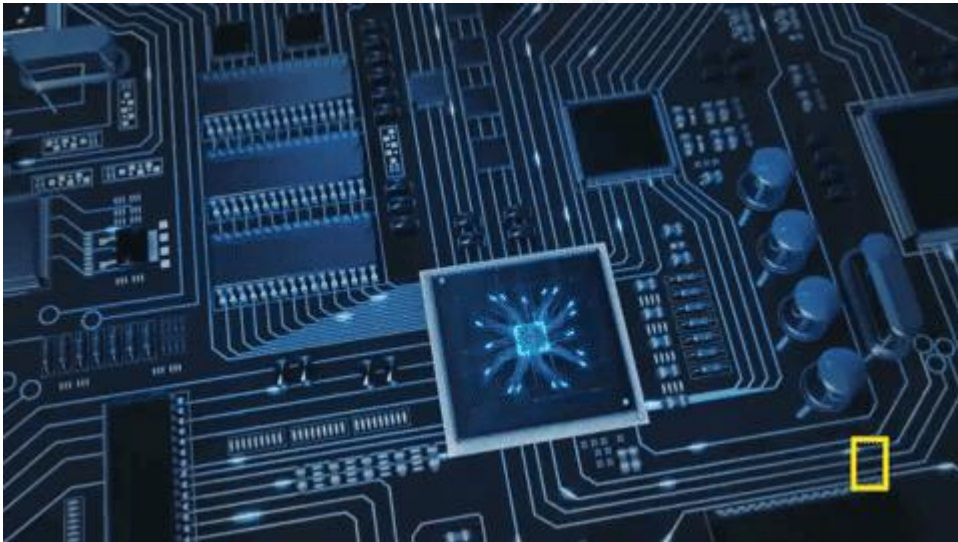


BEE714D

Big Data Analytics in Power Systems

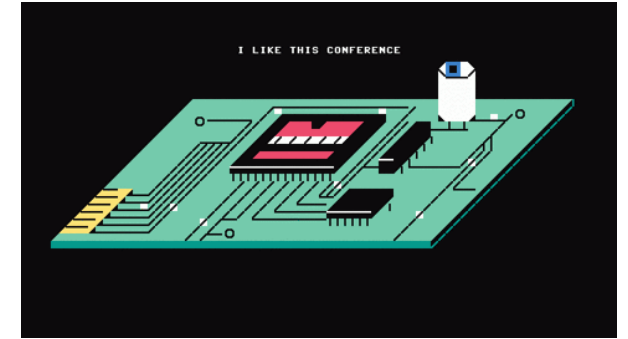
Module-2: Role of Big Data in Smart Grid Communications



Presented by,
Mr.Shreeshayana R
Assistant Professor
Electrical and Electronics Engineering
ATME College of Engineering, Mysuru

Course Overview

- **Course Code:** BEE714D
- **Course Title:** Big Data Analytics in Power Systems
- **Type:** Professional Elective
- **Prerequisite:** Power System Analysis-I
- **Contact Hours:** 40 Hours



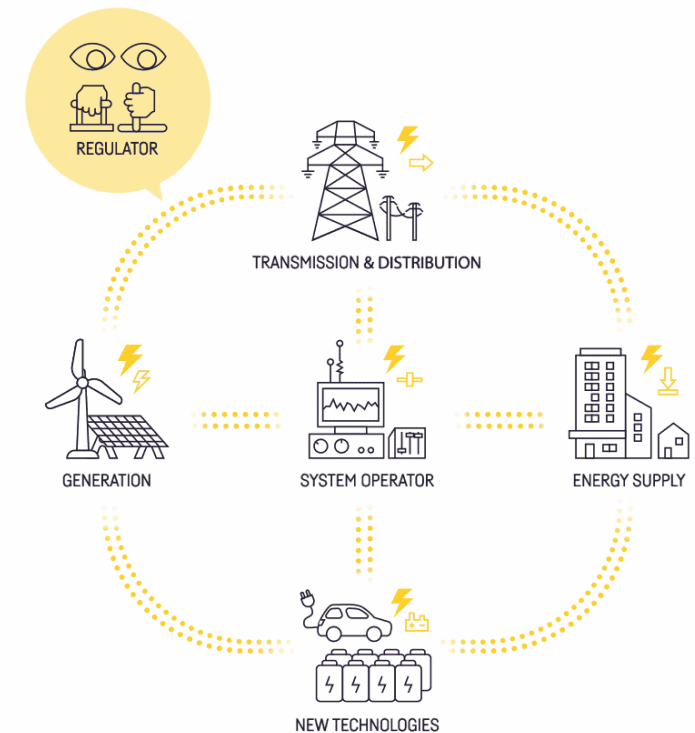
Course Objectives

1. To define big data and to explain **big data application** and analytics to **power systems**.
 2. To explain the role of big data in **smart grid communications and optimization** of big data in electric power systems.
 3. To explain security methods for **the infrastructure communication** and data mining methods for theft detection in power systems.
 4. To explain the application of **unit commitment method** in the control of smart grid.
- To explain **protection algorithm** for transformer based on data pattern recognition



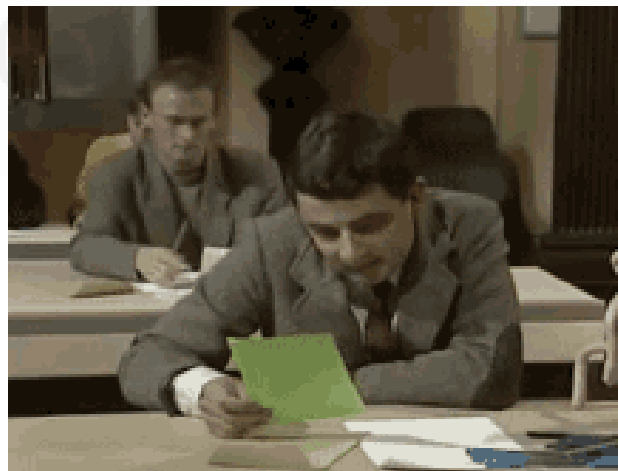
Module 2: Role of Big Data in Smart Grid Communications

- **Role of Big Data in Smart Grid Communications:** Introduction, The Grid Modernization, The Grid Interconnection with the Internet of Things, Data Traffic Pattern in a Smart Grid Environment, The Massive Flow of Information in a Smart Scenario, The Volume of Generated Data in a Smart Distribution System: A Case of Study.
- **Big Data Optimization in Electric Power Systems:** Introduction, Background, Scientometric Analysis of Big Data, Big Data and Power Systems, Optimization Techniques Used in the Big Data Analysis.



Delivery Plan: Week-4 to Week-5

IA	Module	COs
IA-1	Module-1	CO-1
	Module-2 a	CO-1
IA-2	Module-3	CO-3
	Module-2 b	CO-2
IA-3	Module-4	CO-4
	Module-5	CO-4



Text Books

Big Data Analytics in Future Power Systems, Ahmed F. Zobaa and Trevor J. Bihl, CRC Press 2019. 2019.

List of Additional Reference Books/URLs, Text Books, Notes, Multimedia Content, etc

- 1. Big Data Analytics for Power Systems – Big Data Analytics in Power Systems**
- 2. Application of Big-Data Analytics in Power System Protection-Lec-37: Application of Big-Data Analytics in Power System Protection**

Course Outcomes & Bloom's Taxonomy

- CO-1:** Interpret the role of big data and machine-learning methods applicable to power systems and in particular to Smart Grid communications. [L2]
- CO-2:** Apply optimization methods which are suitable for big data models in power systems. [L3]
- CO-3:** Identify various cyber security issues, electricity theft detection and mitigation that exist in IoT-enable future power systems. [L3]
- CO-4:** Identify renewable energy planning concerns associated with planned future power systems that have high renewable penetration. [L3]

Course Code:	BEE714D	TITLE: Big Data Analytics in Power Systems							Faculty Member: SHREESHAYANA R			
List of Course Outcomes	Program Outcomes											
	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12
CO-1	2	-	-	-	2	-	-	-	2	-	-	2
CO-2	2	2	-	-	2	-	-	-	2	-	-	2
CO-3	2	2	-	-	2	-	-	-	2	-	-	2
CO-4	2	2	-	-	2	-	-	-	2	-	-	2

Course Code:	BEE714D	TITLE: Big Data Analytics in Power Systems	Faculty Member: SHREESHAYANA R
List of Course Outcomes	Program Specific Outcomes		
	PSO1	PSO2	
CO-1	2		
CO-2	2	-	
CO-3	2	-	
CO-4	2	-	

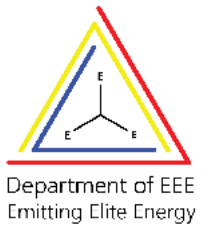
SYLLABUS

Role of Big Data in Smart Grid Communications:

- 2.1 Introduction,
- 2.2 The Grid Modernization,
- 2.3 The Grid Interconnection with the Internet of Things,
- 2.4 Data Traffic Pattern in a Smart Grid Environment,
- 2.5 The Massive Flow of Information in a Smart Scenario ,
- 2.6 The Volume of Generated Data in a Smart Distribution System: A Case of Study.

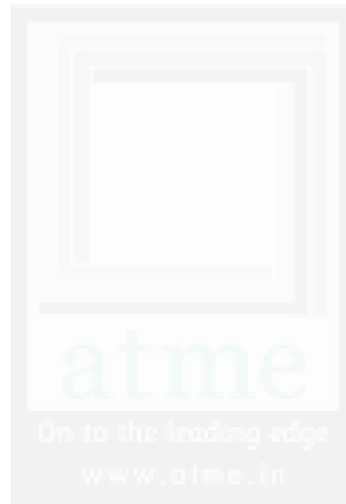


A T M E
College of Engineering



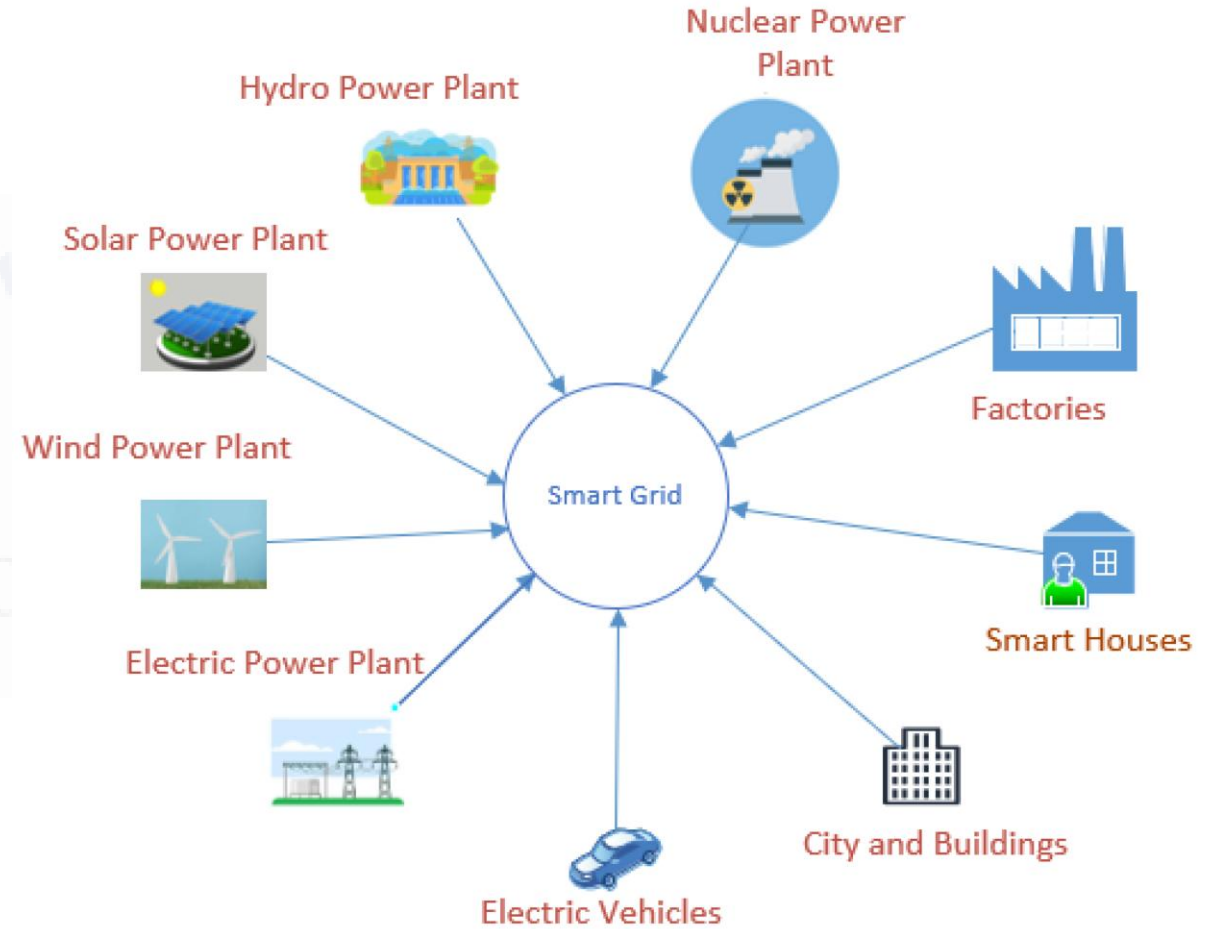
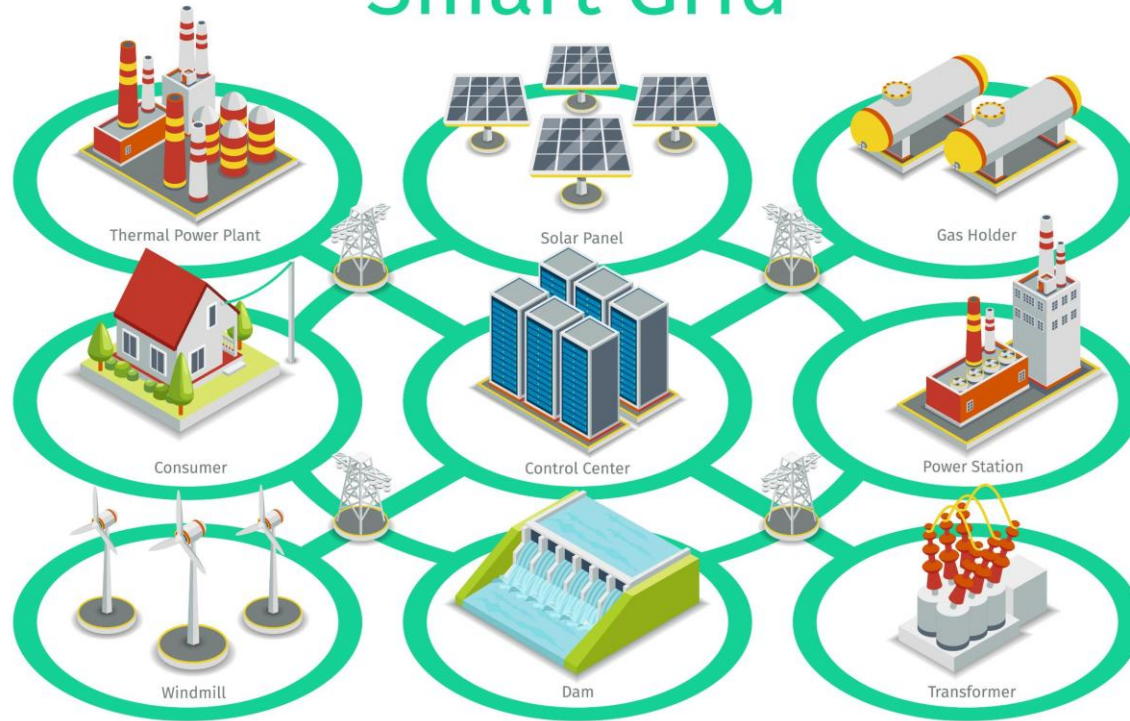
Module-2

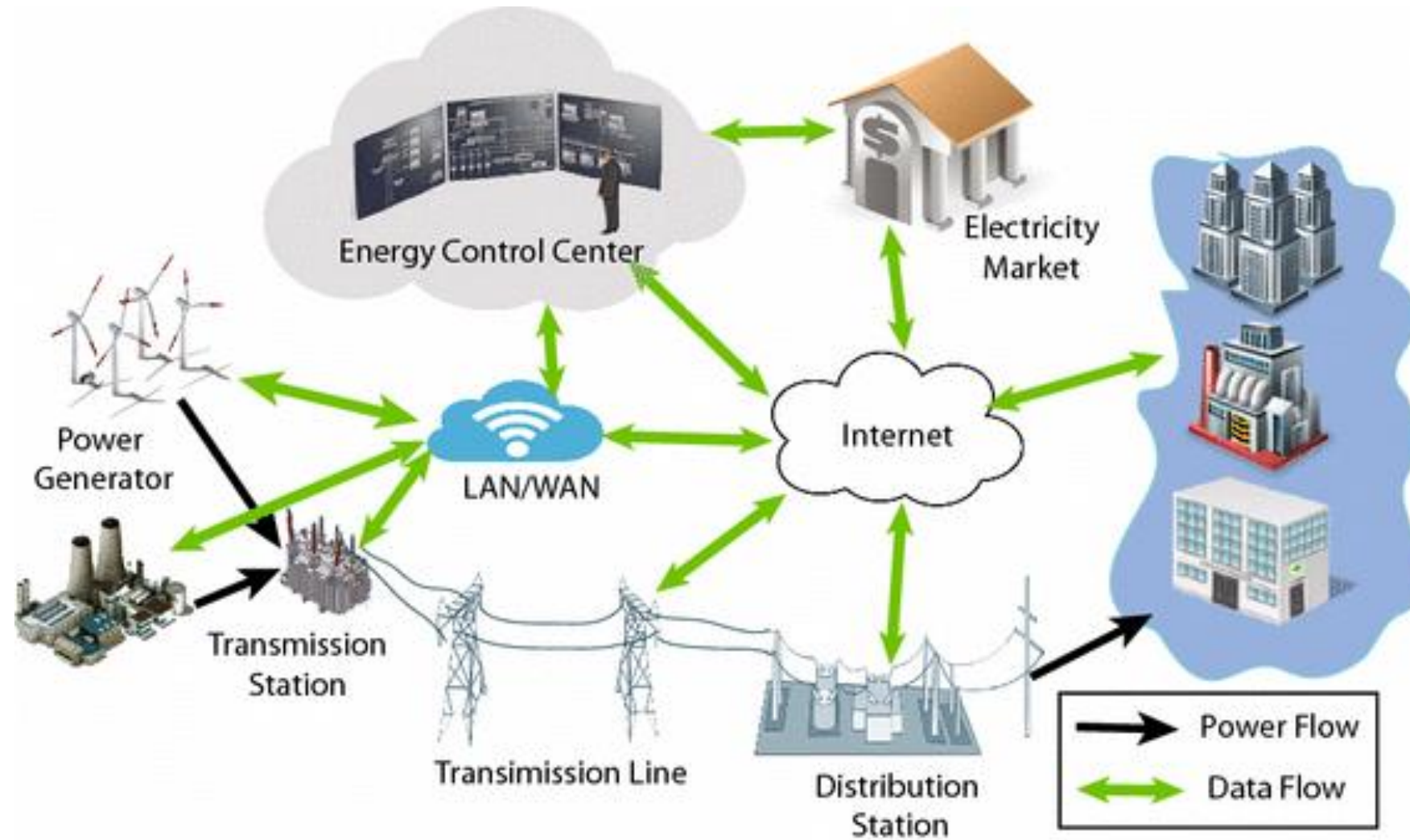
Introduction



A T M E
College of Engineering

Smart Grid



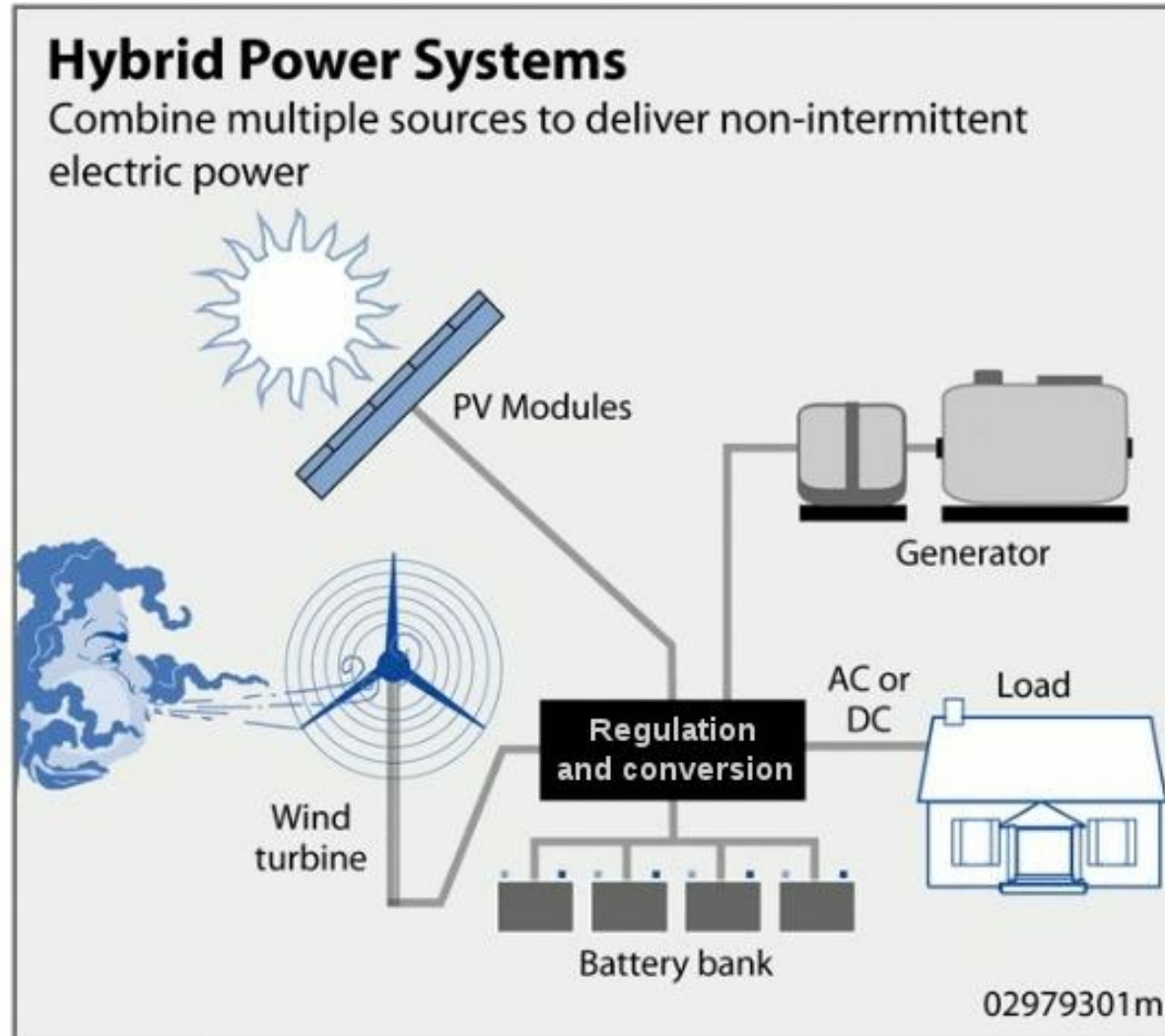


2.1 Introduction

- **Smart Grid as a Sensor Network** – Consists of many connected smart devices.
- **Rising Data Flow** – Growth of devices and operational needs increases information exchange.
- **Big Data Role** – Provides efficient management of large data volumes.
- **Benefits** – Helps utilities understand customer behavior, demand, weather, outages, and failures.
- **Robust Methodologies** – Needed to make the grid smarter and more reliable.
- **Key Concern** – Quantifying and utilizing data from devices effectively.

2.1 Introduction

- **Chapter Aim** – To characterize and evaluate data growth in smart grid communication networks.
- **Example** – Active distribution system shows massive data generation for monitoring and control.



2.2 The Grid Modernization

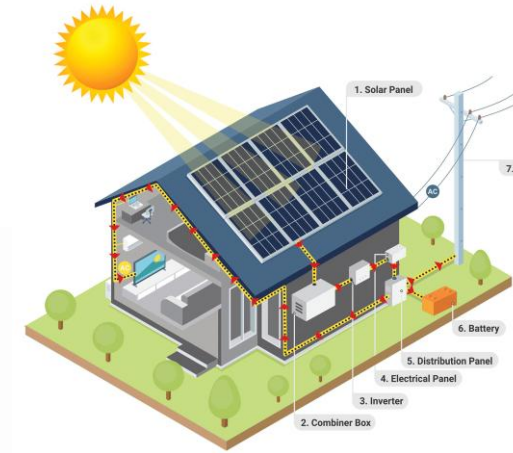
• **Dynamic Nature** – Power generation variations directly affect customer supply → requires real-time monitoring and operation.

• **Smart Grid** – Integrates ICT with electrical system to improve reliability, robustness, safety, reduce losses and failures (especially at distribution level where 90% failures occur).

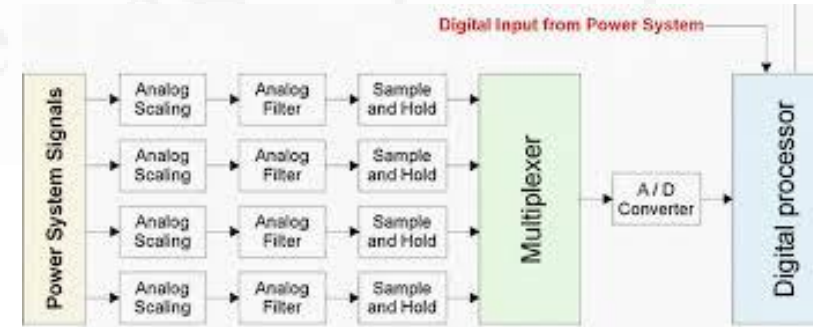


2.2 The Grid Modernization

• **Distributed Generation (Solar/Wind)** – Higher penetration increases need for monitoring and efficient info exchange among all agents.

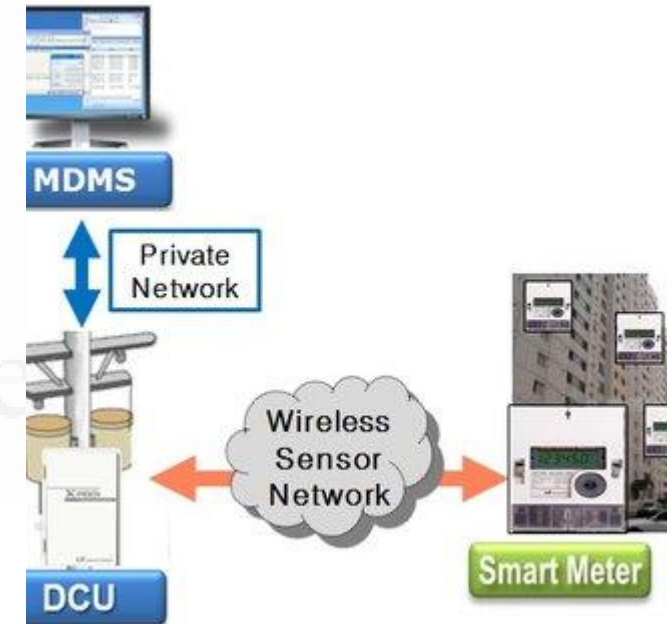


• **Role of IEDs – Real-time automation** through data from user profiles, PMUs, distributed resources, AMI, sensors, actuators, breakers, capacitor banks, and utilities.

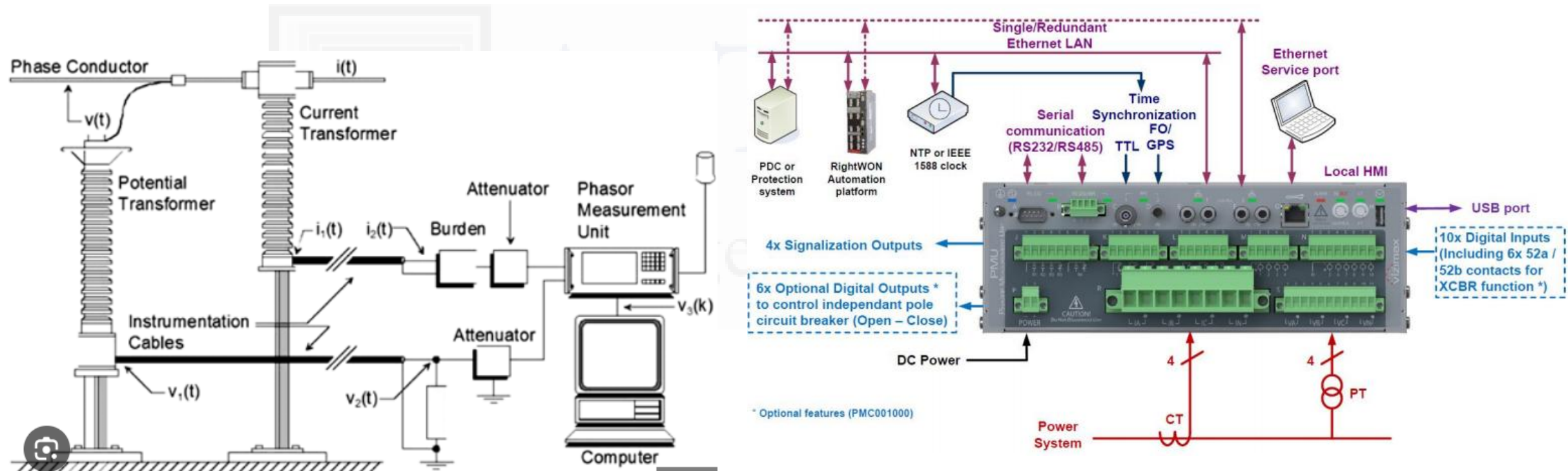


• **Advanced Metering Infrastructure (AMI)** – First step in modernization; enables two-way communication between users and utilities, supports demand response, pricing, state estimation, and real-time supply-demand balance.

• <https://www.youtube.com/watch?v=Lby1sDHpXKo>



• **Phasor Measurement Units (PMUs)** – Applied in protection, state estimation, voltage/frequency control, renewable integration, and islanding monitoring.



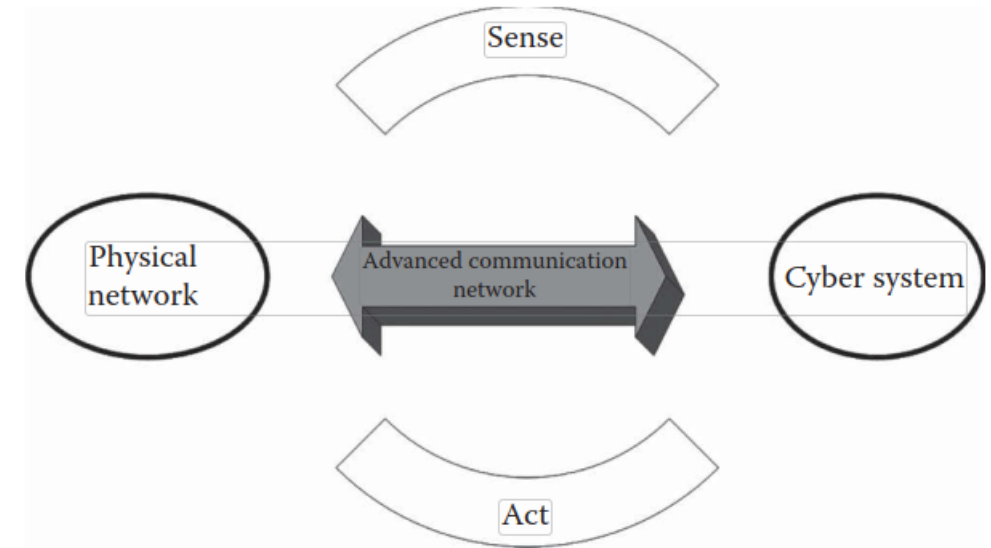
- **Massive Data Generation** – Intelligent devices and sensors create huge volumes of information requiring proper management.
- **Big Data as Solution** – Provides tools to handle and analyze large datasets, improving decision-making, grid robustness, and profitability.
- **Key Requirement** – Quantify which devices generate data and how much, to effectively apply Big Data technologies.

2.3 The Grid Interconnection with the Internet of Things

- **Intelligent Communication Network** – Must be high-performance, reliable, robust, flexible → handles data collection, routing, monitoring, and management.

- **Scale of Communication** – Thousands of devices send data to substations and control centers with minimal human interaction; each distribution system has hundreds–thousands of smart meters, PMUs, and IEDs.

- **IoT in Smart Grids** – Enhances flexibility, scalability, and automation → forms a cyber-physical system integrating power and communication networks.



FIGURE

Smart grid as a cyber–physical system. Adapted from Yu and Xue (2016).

•Key Communication Requirements:

- Latency** – network delay must be minimal.
- Bandwidth** – critical for choosing wired/wireless tech.
- Interoperability/Flexibility** – systems must work together.
- Data Throughput** – max info transfer capacity.
- Cybersecurity** – authentication, authorization, privacy.
- Data Characteristics** – IoT devices generate big data: heterogeneous, varied, unstructured, noisy, redundant.

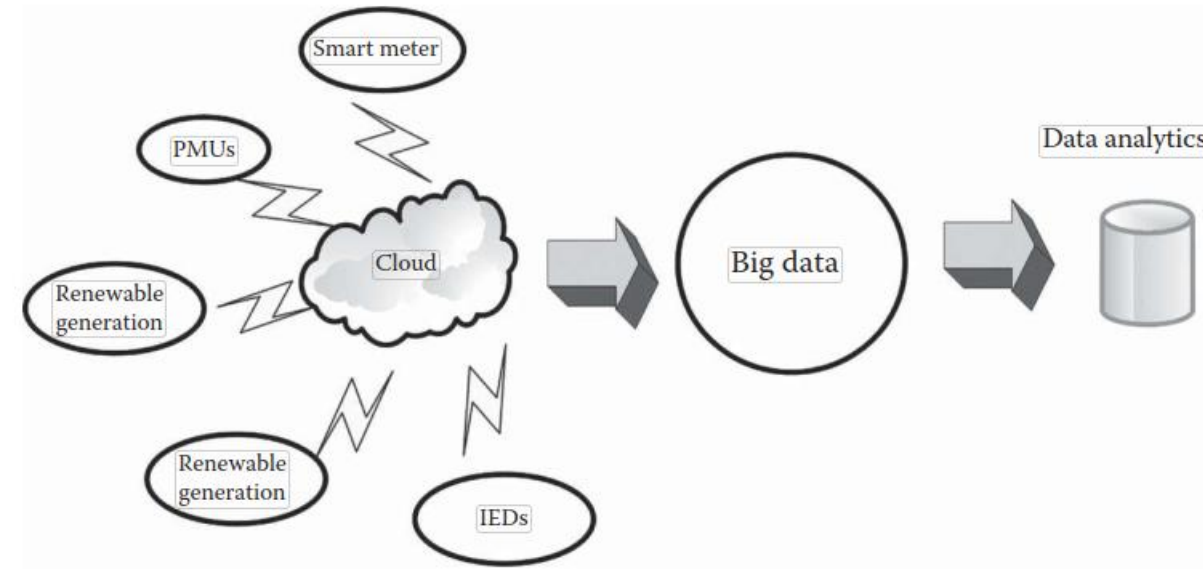


FIGURE
The integration of IoT and Big Data technology. Adapted from Marjani et al. (2017).

- **Technology Choices** – Must be reliable, secure, cost-effective, and adaptable to future needs.
- **Wired** – high data rate, costly, low mobility.
- **Wireless/Mobile** – scalable, flexible, lower cost, supports **M2M communication**.
- **Mobile Standards** – LTE-based enhancements for IoT.
- **NB-IoT (3GPP Release 13 & 14)** – efficient spectrum use, wide coverage, low-cost devices, high capacity, supports massive machine-to-machine (M2M) communication.

2.4 Data Traffic Pattern in a Smart Grid Environment

- Traditional power systems are **unidirectional**, with limited intelligence and no real-time communication (Bouhafs et al., 2012).
- Smart grids require **advanced communication systems** for protection, monitoring, and control.
- Existing systems like **SCADA** and **AMR** are inadequate for future smart grid demands (Lai and Lai, 2015).
- Future grids need **bidirectional communication**, ensuring robustness, reliability, and security.

Main functions (Marjani et al., 2017; Fan et al., 2012):

- 1. Distribution Control & Protection – IED devices detect/locate faults** and exchange reporting messages.
- 2. Wide Area Monitoring System** – Collects info from substations and large areas for critical decisions.
- 3. Demand Response (DR)** – Manages multiple distributed energy resources with intermittent variables.
- 4. Advanced Metering Infrastructure (AMI)** – Smart meters for billing, consumer interaction, load control, DR, and islanding detection.

- Each application has **different latency and data requirements** (Kuzlu et al., 2014):
 - **Protection:** 1–10 ms, few bytes
 - **Control:** ~100 ms, few bytes
 - **Monitoring:** ~1 s, few–medium bytes
 - **Metering:** minutes–hours, medium size
- **AMI** enables functions like **choosing energy sources, demand monitoring, and dynamic billing** (Sánchez-Ayala et al., 2013).
- Smart grids will produce **huge amounts of data**, e.g., PMUs (sampled at 10–60 per cycle) and AMI (every 1–15 min).
- **Big data analytics** is crucial for maintaining reliability, robustness, and grid modernization (Daki et al., 2017).

Phasor Measurement Units (PMUs)

- PMUs placed strategically perform **accurate voltage and current phasor measurements** using GPS (Liu et al., 2012).
- Data is sent to **Phasor Data Units (PDUs)** for processing.
- PMUs generate measurements per cycle (10, 20, 30, 60 most common).
- Required bandwidth depends on:
 - **$BW = N_{frame} \times f_s \times N_{PMU}$**
 - N_{frame} = frame size (bytes)
 - f_s = sampling frequency
 - N_{PMU} = number of PMUs connected
- **Higher sampling frequency = higher transmission capacity needs.**
- **Increasing PMU deployment = more synchrophasor data to analyze.**

Advanced Metering Infrastructure (AMI)

- Smart meters enable **numerous smart applications** when paired with reliable communication infrastructure.
- Each meter generates data with **different message size and sample rates** (Luan et al., 2010).
- Smart meters will be deployed in **residential, commercial, and industrial** facilities, creating massive data flow.
- **Estimating number of smart meters** depends on communication network design (coverage, antenna gain, propagation losses, geography).
- Formula for number of smart meters:

$$NSM = \rho \pi d^2$$

- ρ = smart meter density (per m²)
- d = coverage radius of base station (urban, suburban, rural variations).
- **Minimum data rate per smart meter = 64 kbps** for critical operations (Persia et al., 2015).

2.5 The Massive Flow of Information in a Smart Scenario

•Massive data generation:

- Growth of **IEDs** (Intelligent Electronic Devices) → exponential data flow.
- Utilities already generate **hundreds of millions of gigabytes**, rising to **terabytes annually**.
- **Synchrophasors alone produce hundreds of terabytes per year** (Asad et al., 2017).

•Opportunities from big data:

- Improves **grid reliability** and enables applications such as:
 - **Predictive analytics**
 - **Demand-side management**
 - **Real-time grid awareness**
 - **Outage detection**
 - **Asset management**
 - **Theft detection**

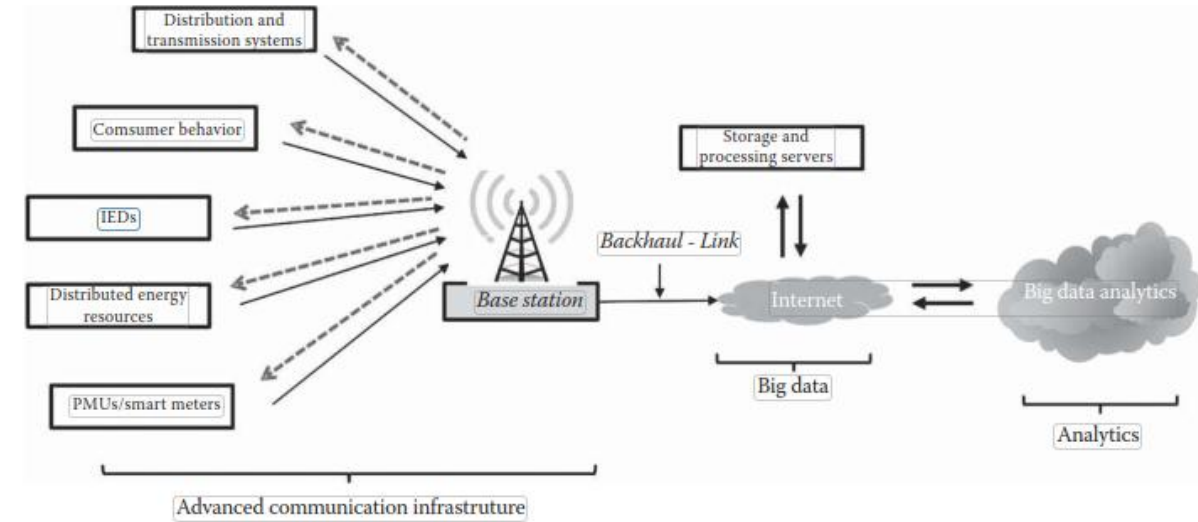


FIGURE 3.3
The integration of information and communication technologies in a smart grid context.

Role of big data (Jiang et al., 2016):

- Enhances **prediction, management, and processing** of grid operations.
- Recognizes **patterns in data** for better operational decisions.

Key application areas:

- **Demand Response (DR):** Predicting/analyzing user patterns → accurate demand forecasting.
- **Distributed Energy Resources (DER):** Forecasting & scheduling intermittent energy sources for planning.
- **AMI:** Smart meter data → customer behavior, load forecast, demand management.
- **Distribution Automation:** Monitoring and sensing → predict outages, improve robustness.

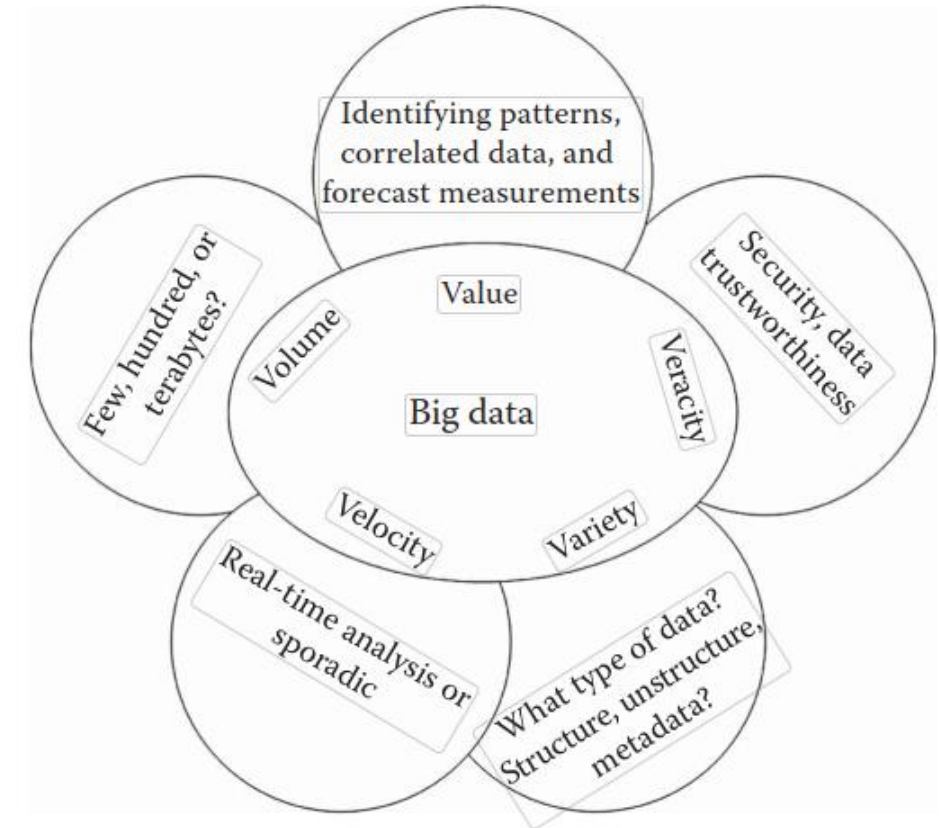


FIGURE 3.4

The 5 V's of big data in a smart grid scenario. Adapted from Subhani et al. (2015).

Big data foundation – 5 V's (Subhani et al., 2015):

- **Volume** – massive data scale
- **Velocity** – high-speed generation & transfer
- **Variety** – diverse data formats (structured/unstructured)
- **Value** – usefulness of data for decisions
- **Veracity** – data quality and accuracy

Advanced data analytics methods:

- Predictive analytics, data mining, AI, fuzzy theory

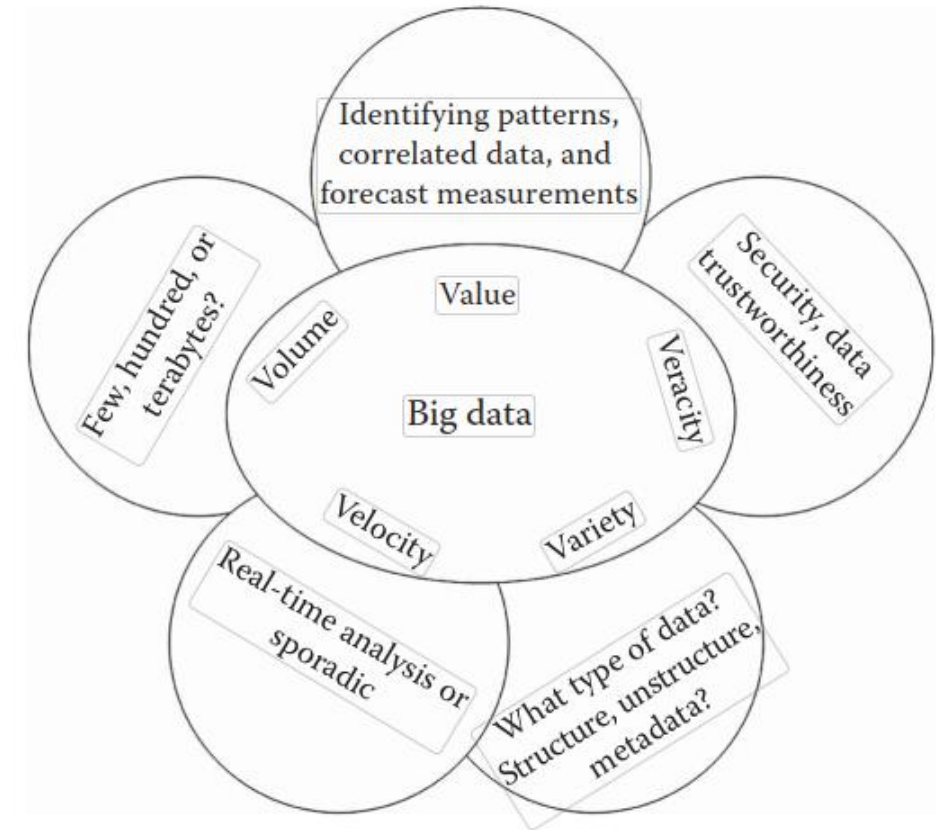


FIGURE 3.4

The 5 V's of big data in a smart grid scenario. Adapted from Subhani et al. (2015).

2.6 Volume of Generated Data in Smart Distribution – Case Study'

General

- **Future smart grids = huge data** from IEDs, PMUs, and smart meters.
- In normal mode: periodic traffic (e.g., metering every 15 min).
- In critical mode: continuous messages, DR, load flow, outage info.

Case Study I – Data Generated by PMUs

Step-1: Setup

- IEEE 123 bus system is selected (high-load topology, widely used in research).
- PMU placement follows Jamil et al. (2014).
- **49 PMUs** deployed → optimal number for this system.

Step-2: Data Frame Structure

- Each PMU sends **frame packets** as per IEEE Std. C37.118 (2011).
- Packet = **80 fixed bytes** + fields (phasors, transducers, digital signals).
- Each PMU has **8 phasor channels** + **2 digital channels**.

Step-3: Sampling Rates

- Brazilian system frequency = **60 Hz**.
- Sampling rates considered: **10, 20, 30, 60 synchrophasors/sec**.
- Higher sampling rate = higher data generation.

Step-4: Bandwidth Requirement Formula

$$BW = N_{frame} \times f_s \times N_{PMU}$$

- N_{frame} = frame size (bytes)
- f_s = sampling frequency
- N_{PMU} = number of PMUs (49 in this case)

Step-5: Results

- At 10 samples/sec → data volume already in **hundreds of Mbps**.
- At higher rates → data grows drastically.

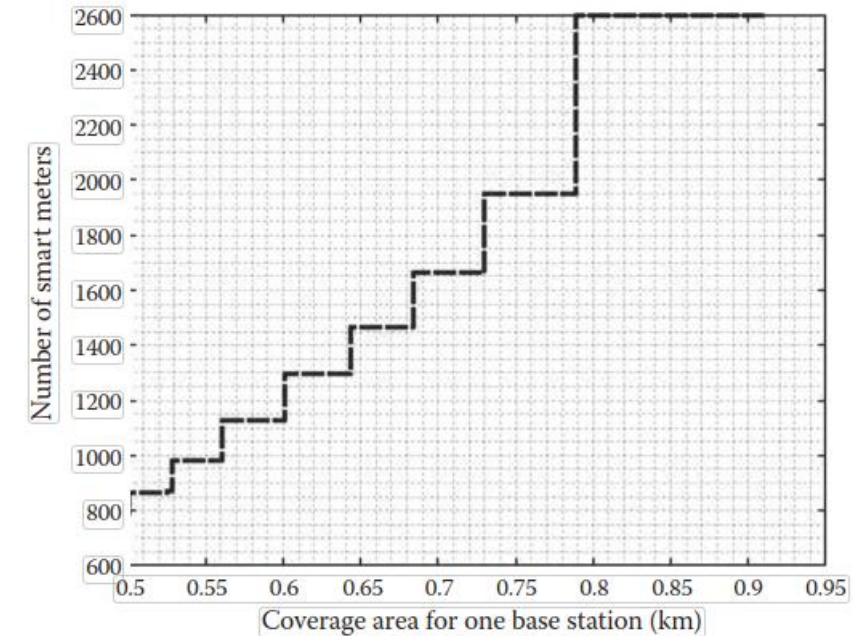
Step-6: Inference

- PMUs generate **massive continuous data streams**.
- Without **big data analytics**, it is difficult to handle, store, and process.
- Organized analysis = better grid performance and reliability.

•Simulation: **IEEE 123 bus system** with NB-IoT communication.

Volume of Generated Data by PMUs

#PMU	#Bytes	Sampling Rate (Mbps)					
		10	12	15	20	30	60
1	112	8.96	10.752	13.44	17.92	26.88	53.76
2	224	16.06	19.27	24.08	32.11	48.17	96.34
3	336	43.16	51.79	64.74	86.32	129.48	258.96
6	672	232.02	278.43	348.04	464.05	696.07	1392.15
9	1008	1871.05	2245.26	2806.57	3742.10	5613.15	11226.29
10	1120	16764.60	20117.51	25146.89	33529.19	50293.79	100587.57
15	1680	225316.16	270379.39	337974.24	450632.32	675948.47	1351896.95
20	2240	4037665.548	4845198.66	6056498.323	8075331.097	12112996.6	24225993.29
30	3360	108532449.9	130238940	162798674.9	217064899.9	325597350	651194699.7
49	5488	4765008682	5718010419	7147513023	9530017364	1,4295E+10	28590052093
50	5600	2,13472E+11	2,5617E+11	3,20209E+11	4,26945E+11	6,4042E+11	1,28083E+12
60	6720	1,14763E+13	1,3772E+13	1,72144E+13	2,29526E+13	3,4429E+13	6,88577E+13



Case Study II – Data Generated by AMI (Smart Meters)

Step-1: Setup

- Smart meters deployed across the same **IEEE 123 bus system area** ($\approx 2 \text{ km}^2$, urban scenario).
- Meters randomly placed, exact optimal allocation not studied here.

Step-2: Data Rate Assumption

- Each smart meter requires **64 kbps** in critical mission scenarios (Persia et al., 2015).

Step-3: Coverage Analysis

- Communication via **NB-IoT wireless**.
- Factors: distance, propagation loss, antenna gain, coverage area.
- One base station coverage = **0.5 km^2** \rightarrow supports **~ 800 meters**.

Step-4: Scaling to Whole System

- Total area = $\sim 2 \text{ km}^2$.
- Need **5 base stations** to cover entire system.
- Supports **~ 4000 smart meters**.

Step-5: Inference

- More smart meters = **exponential data growth**.
- Data must be processed by **advanced analytics** to avoid overload.
- Correct treatment ensures **efficient grid + communication flow**.

Inference

Case I (PMUs): Generate huge continuous data; needs big data analytics for reliability.

Case II (AMI): Thousands of smart meters → **exponential data; requires efficient communication & analytics.**

•**Overall:** Smart grids produce massive data; without proper treatment it's a burden, with analytics it ensures reliability and efficiency.

Big Data Optimization in Electric Power Systems:

- 2.6 Introduction, Background,**
- 2.7 Scientometric Analysis of Big Data,**
- 2.8 Big Data and Power Systems,**
- 2.9 Optimization Techniques Used in the Big Data Analysis.**

2.6 Introduction

• **Big Data Characteristics:** Commonly defined by five Vs –

- **Volume:** Large-scale data (terabytes to zettabytes)
- **Velocity:** Speed of data processing (batch or streaming)
- **Variety:** Structured, semi-structured, and unstructured data
- **Veracity:** Accuracy and reliability of data
- **Value:** Usefulness of insights gained from data

• **Power Systems Context:**

- Comprise networks to generate, transmit, and use electricity
- Require predictive models and optimization for efficient operation

•**Optimization Needs:**

- Growing data size demands advanced optimization algorithms
- Problems often involve nonconvex, nonsmooth, and large-scale computations
- Aim is to maximize or minimize objective functions within constraints

•**Opportunities in Big Data:**

- Rich datasets can enhance model accuracy
- Effective learning depends on uncovering low-dimensional data structures

Scope of the Chapter:

- Reviews recent research on **big data optimization in power systems**
- Includes **scientometric analysis** using keywords like “big data” and “power system”
- Covers keyword **trends, network visualizations, journal mapping, and bibliographic coupling**
- Highlights common optimization techniques and metaheuristic methods used in the field
- Provides insights into current efforts and future directions

Big Data Projects:

- Unlock valuable insights but require complex setup and management.*

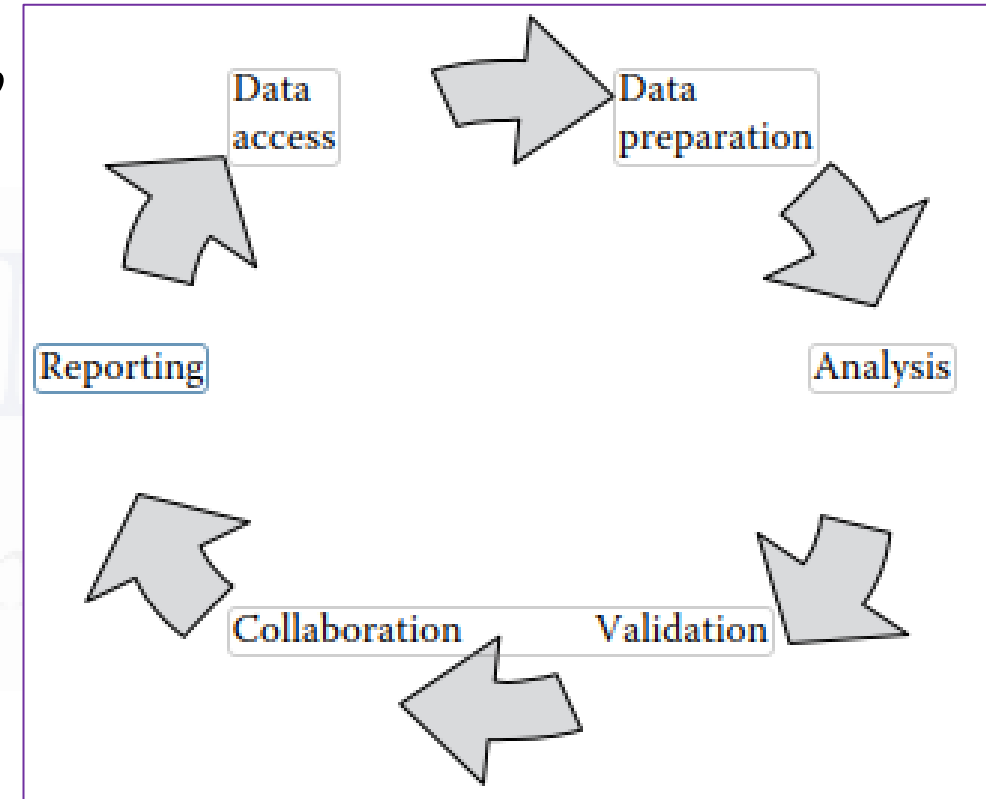
•Key Steps in Big Data Workflow:

- 1.Data Preparation** – Collecting, cleaning, and organizing data
- 2.Data Analysis** – Inspecting and modeling data to gain insights
- 3.Data Validation** – Ensuring accuracy and quality of data
- 4.Data Collaboration** – Sharing and visualizing data effectively
- 5.Data Reporting** – Submitting and presenting data with context
- 6.Data Access** – Retrieving and managing data from repositories

•Applications:

- Supports decision-making in fields like **education**, **healthcare**

- (e.g., personalized medicine), and **manufacturing** (e.g., transparency and efficiency).



2.7 Scientometric Analysis of Big Data,

4.3 Scientometric Analysis of Big Data

•Big Data Relevance:

- **Essential across sectors:** In the 21st century, data is integral to nearly all activities—finance, research, transport, security, business, and more.
- **Challenges:** data storage, processing, and traditional tech limitations

•Scientometric & Social Network Analysis (SNA):

- Used to study trends, collaboration, and growth in scientific research
- Helpful for tracking the evolution of big data in power systems

Definition

“Scientometric is a key enabler that observes scientific publications to explore the structure and growth of a specific science using some quantitative measures of scientific information, as the number of scientific articles published in a given period, their citation impact, etc.”

VOSviewer is a popular software tool used in scientometric analysis and bibliometric mapping. It's especially useful for visualizing relationships in large sets of scientific literature.

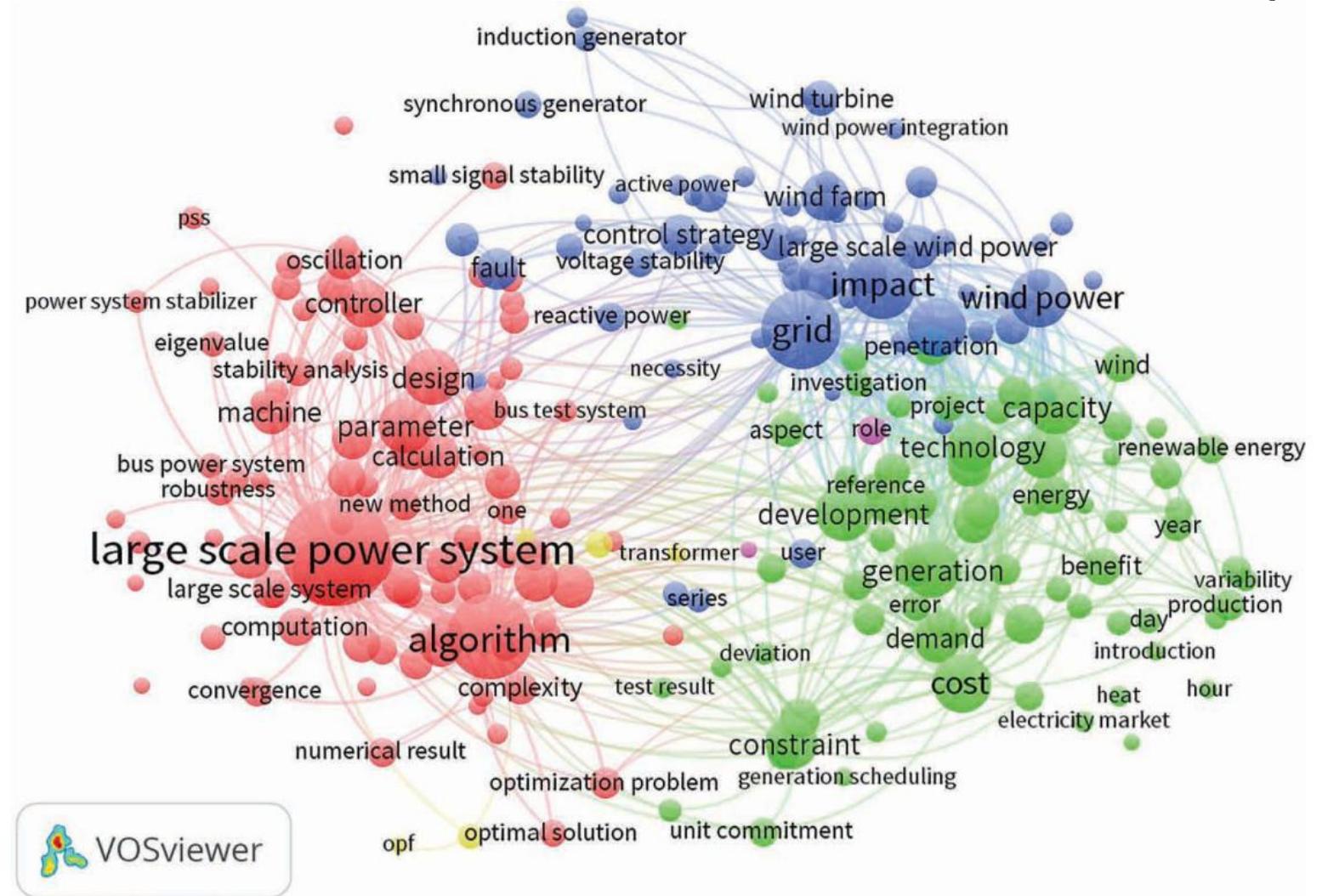
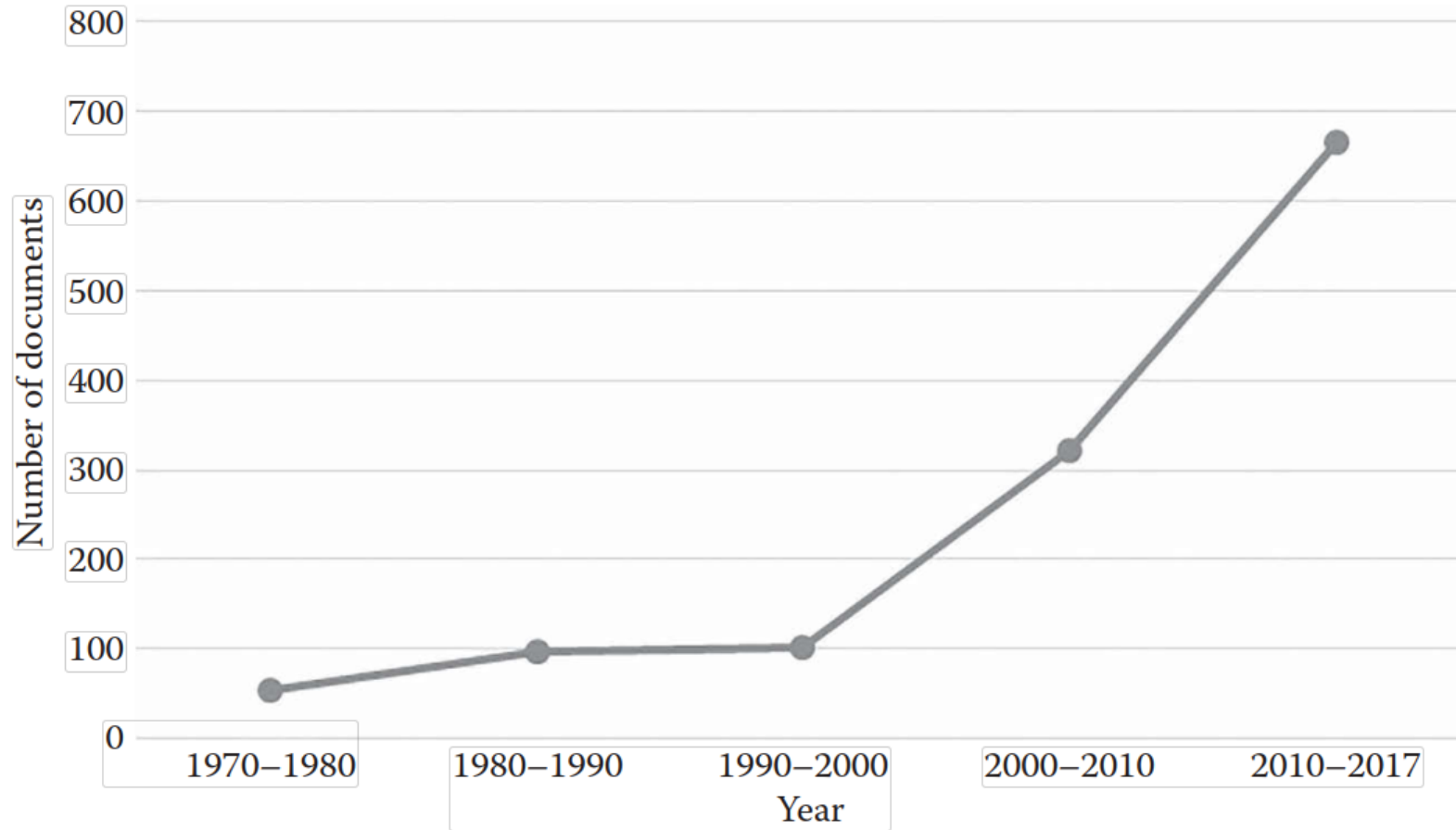


FIGURE
Cognitive map (keyword search based on co-occurrences).



FIGURE

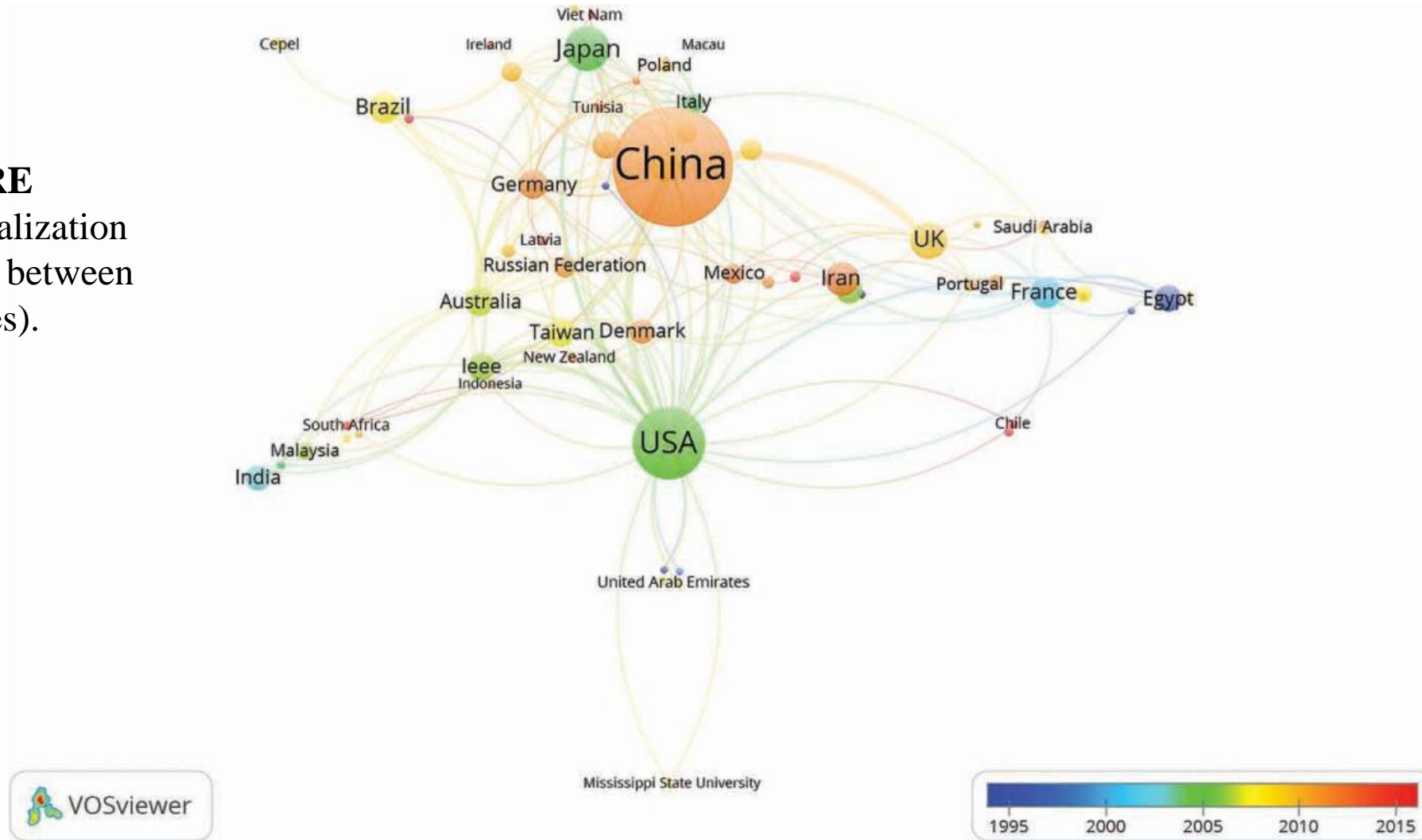
A number of publications on “large-scale” power system.

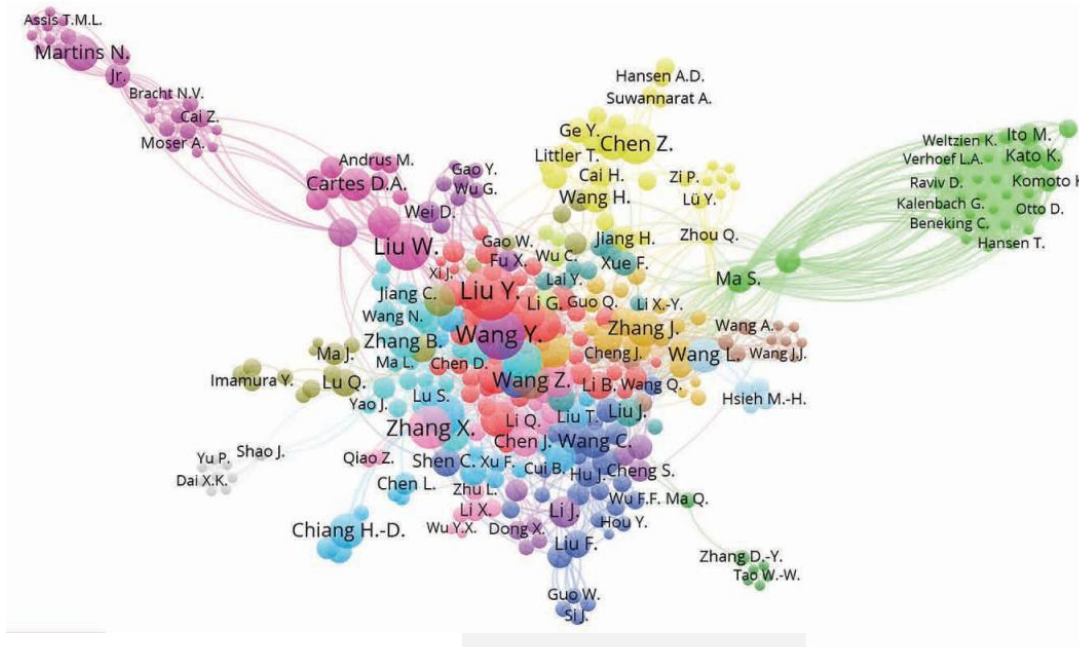
TABLE 4.2

The Most Commonly Used Keywords in Big Data Optimization Literature

No.	Keyword	Occurrences
1	Large-scale power system	399
2	Algorithm	248
3	Grid	211
4	Technique	166
5	Impact	152
6	Wind power	119
7	Cost	116
8	Integration	114
9	Capacity	110
10	Development	107

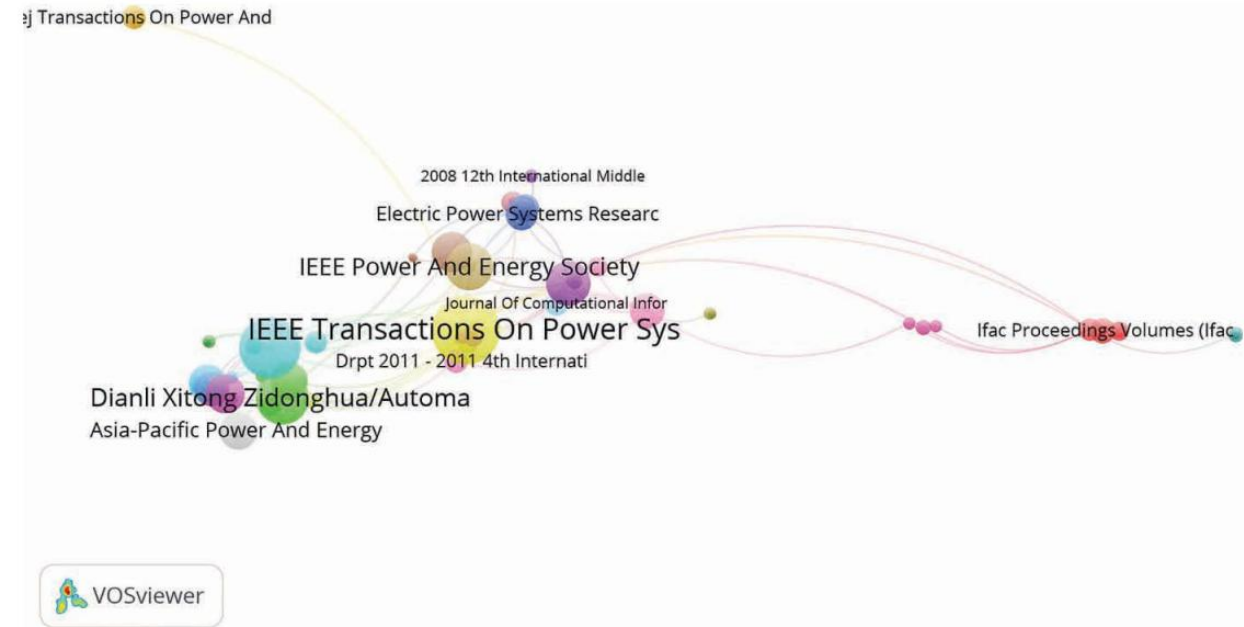
FIGURE
Network visualization
(collaboration between
countries).





FIGURE

Scientific community (coauthor) working on the large-scale power system.



FIGURE

Journal map (title) based on citation analysis.

- **Figures** show citation-based network and density visualizations of active journals.
- **Density Map**
- Red = high journal activity; blue = low activity.
- **Network Visualization**
- The right side shows the densest area, focused on power system research.

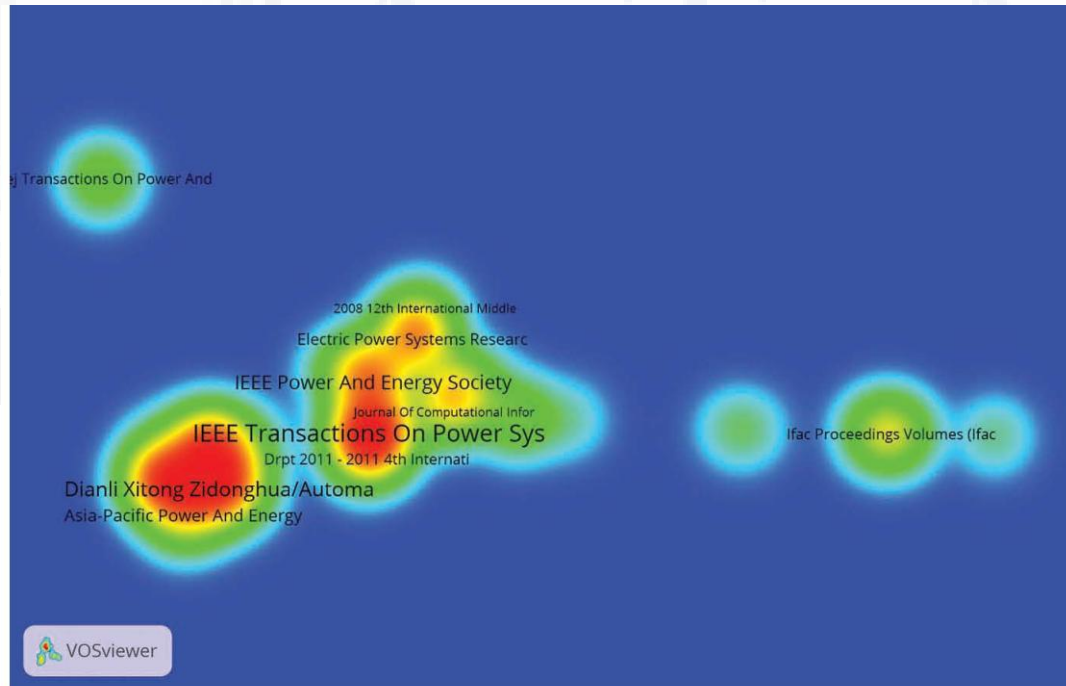


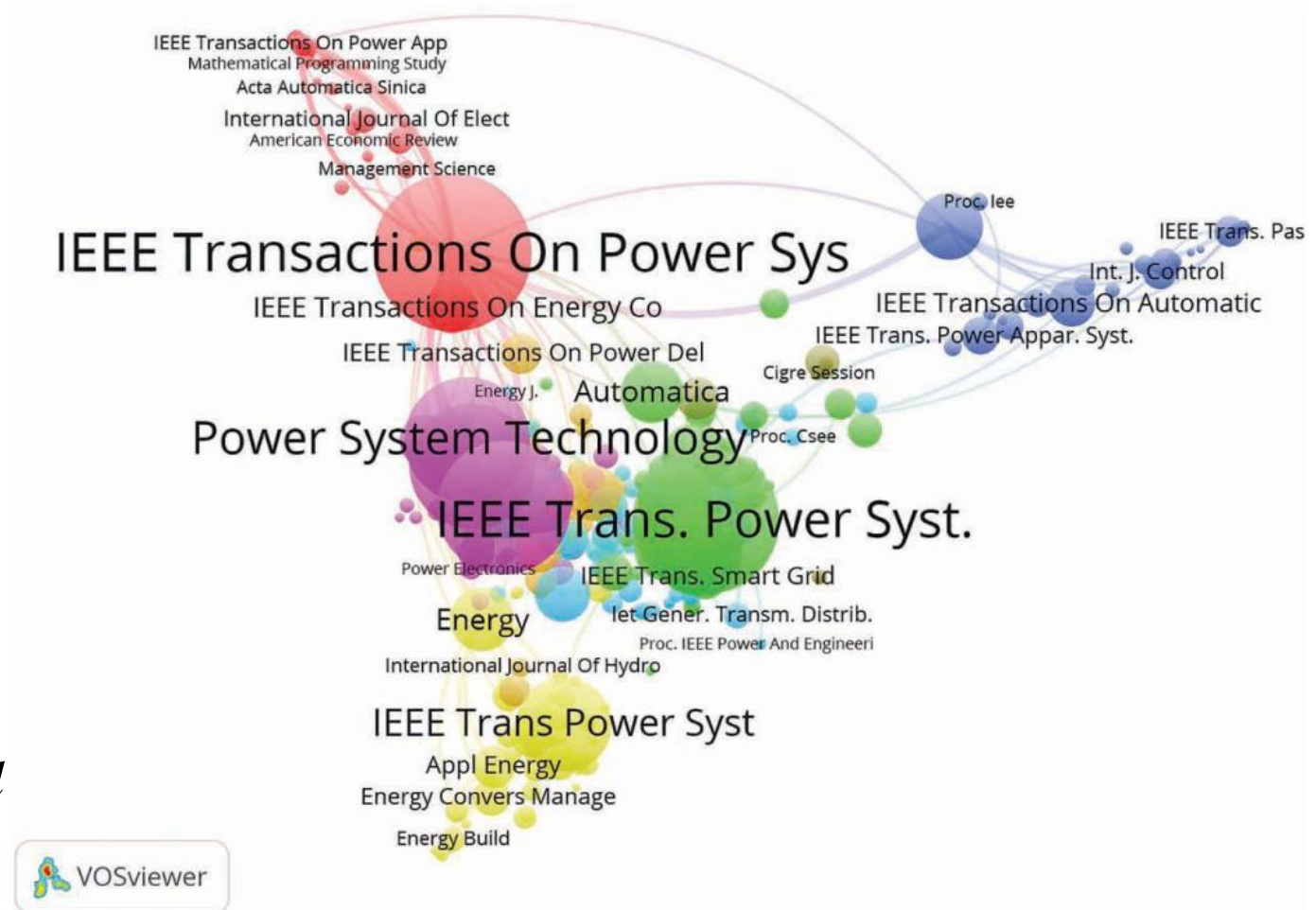
FIGURE
Density map (Journal title) based on
citation analysis.

FIGURE 4.

Network visualization (co-citation analysis).

Key Journals Identified:

- *IEEE Transactions on Power Systems*
- *Applied Mechanics and Materials*
- *Power System Protection and Control*
- *Automation of Electric Power System*
- *IEEE Power and Energy Society General Meeting*
- *International Journal of Electrical Power and Energy Systems*
- *Proceedings of the Chinese Society of Electrical Engineering*



Big Data in Power Systems

- Power systems face **large-scale, complex optimization problems**, often with uncertain inputs.
- **Key areas of research:**
 - **Optimal operation** of energy retailers and hydro units.
 - **Stochastic programming and uncertainty modeling** (e.g., Monte Carlo methods).
 - **Emission and security planning.**
- Some models involve **millions of variables and equations**, highlighting the **big data scale**.



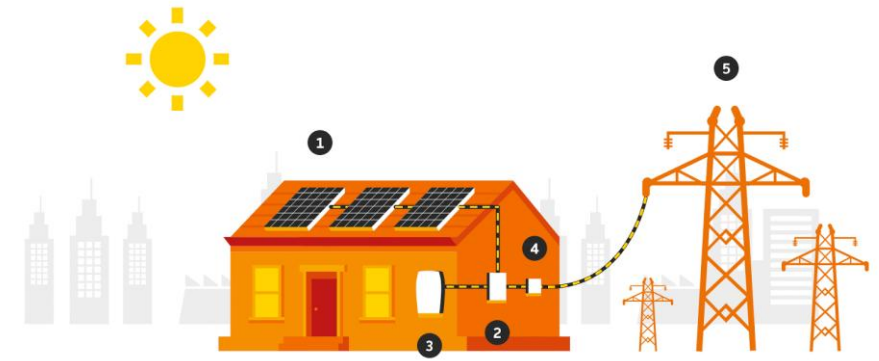
Big Data Optimization:

Key Challenges: Includes data privacy, massive data volume, and complex data management (Zicari et al., 2016).

•Application Areas:

- Social network science
- Machine learning
- Biology
- Power systems (e.g., energy demand, scheduling, dispatch)

•Need for New Algorithms: Traditional optimization methods are inadequate for large-scale problems; more powerful, scalable techniques are required.



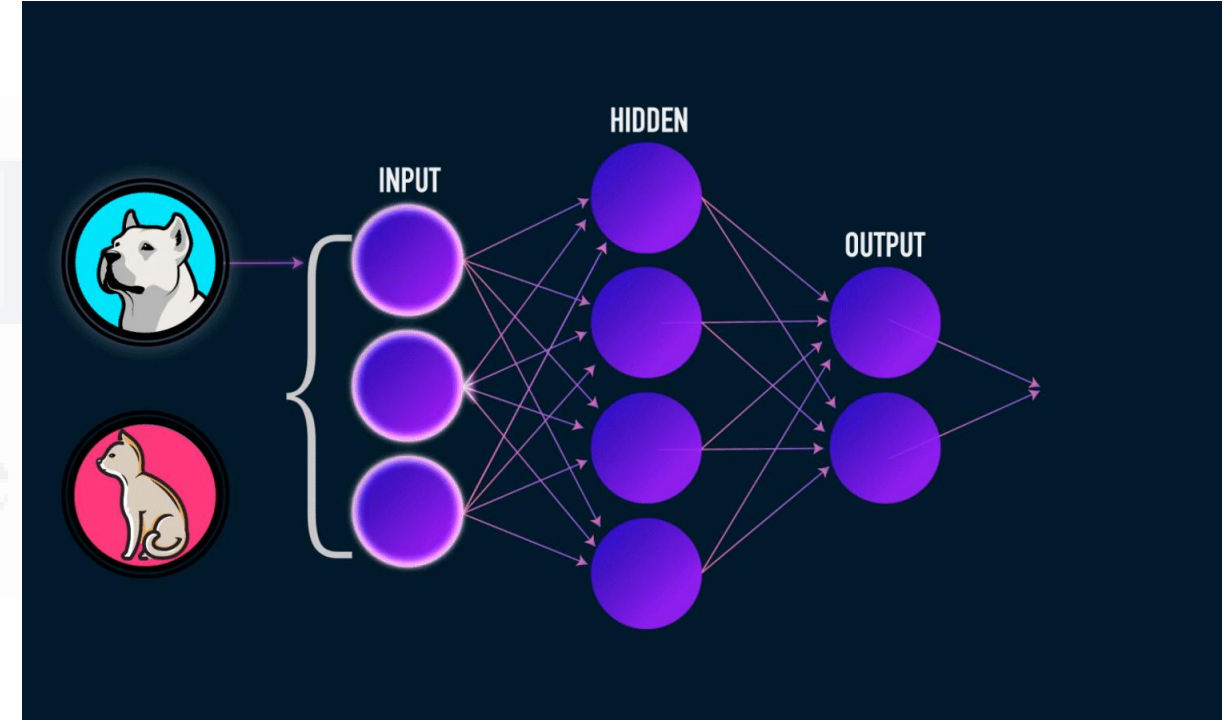
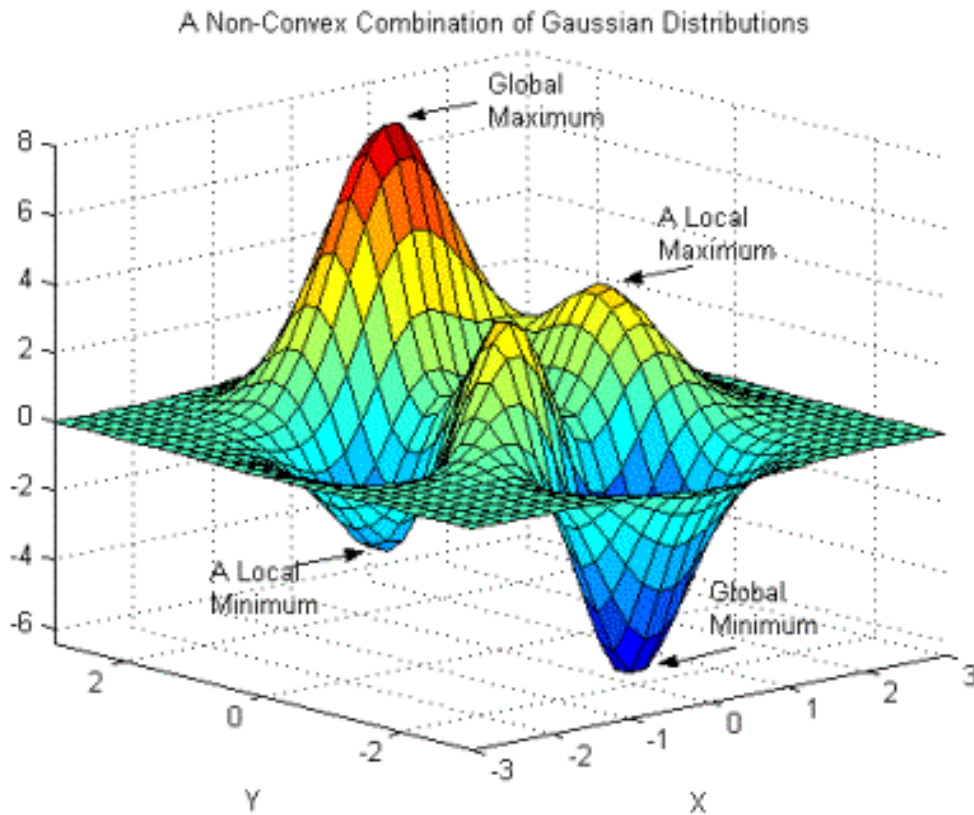
Examples of Optimization Types in Big Data:

- **Logistics & Supply Chain Optimization** (Brouer et al., 2016; Gunasekaran et al., 2017)
- **Nonconvex Optimization** (Gong et al., 2016)
- **Unconstrained Optimization** (Babaie-Kafaki, 2016)
- **Non-smooth Optimization** (Karmitsa, 2016)



Logistics & Supply Chain Optimization

Nonconvex Optimization Example: Training deep neural networks for image recognition involves minimizing a complex, nonconvex loss function over massive datasets.



Training a deep neural network (DNN) for image classification.

Unconstrained Optimization

Example: Linear regression on large-scale customer data minimizes prediction error without any constraints on model parameters.



Use Case: Predicting Customer Churn in a Telecom Company

•**Objective:** Use customer data (e.g., call duration, data usage, complaints) to **predict the likelihood of churn** (i.e., leaving the service)

- **Non-smooth Optimization Example:** Minimizing functions like the absolute value or hinge loss, which have sharp corners or discontinuities, in large-scale machine learning tasks.

Application of Big Data in Power System Studies:

- **Challenges:** Data storage, analysis, visualization, and sharing.
- **Uses:** Trend detection, problem spotting, and predictive analysis.
- **Key Need:** Data must be easy to understand and use.
- **Industry Focus:** Cost-effective storage, processing, and data tools.
- **Barriers:** Lack of skilled staff and need for training.

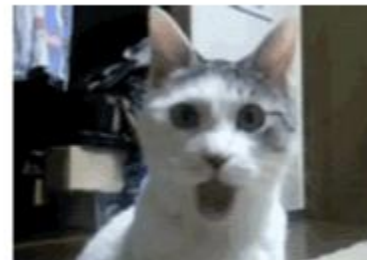
Optimization Techniques in Big Data Analysis:

Traditional methods can't handle large-scale data.

- New techniques are needed for scalability and efficiency.

- **Key approaches include:**

- Big image optimization
 - Intelligent data reduction
 - Hadoop-based optimization
 - Mathematical and metaheuristic methods (Emrouznejad & Marra, 2016)
- Widely applied in power system operations.



Before compression - 477kb

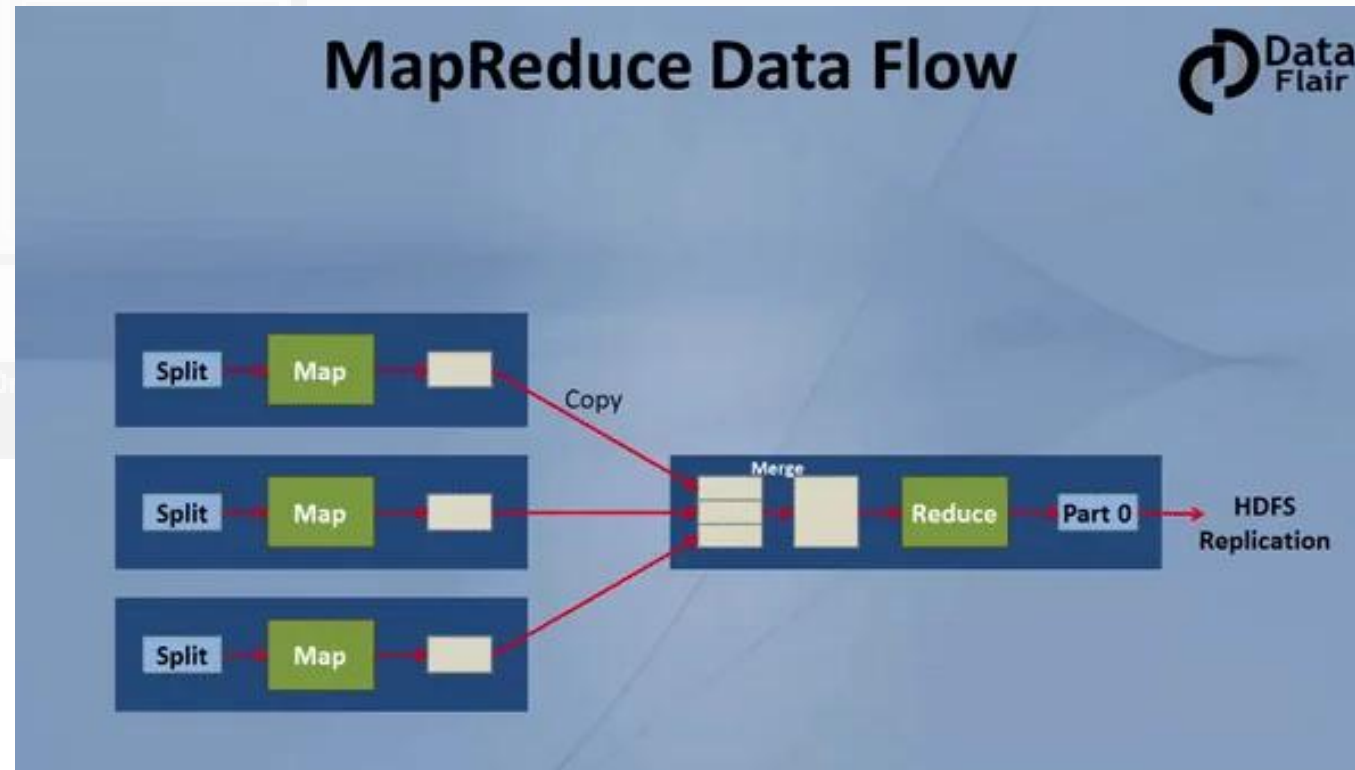


After light compression - 393kb



After heavy compression - 263kb

MapReduce is the heart of **Hadoop**. It is a programming model designed for processing huge volumes of data (both structured as well as unstructured) in parallel by dividing the work into a set of independent sub-work (tasks).



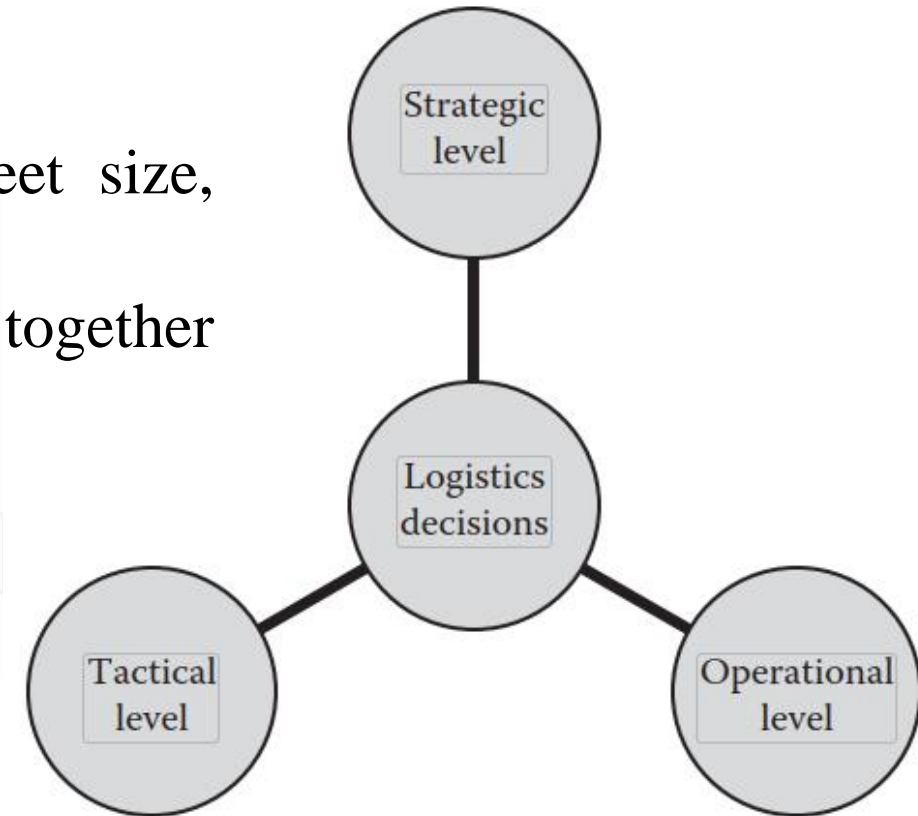
Big Data in Logistics Optimization

Step 1: Understand Logistics vs. Supply Chain

- **Logistics:** Internal company operations (e.g., fleet size, routing).
- **Supply Chain:** Network of organizations working together to deliver products.

Step 2: Know the 3 Decision Levels

- **Strategic:** Long-term decisions (e.g., fleet size, facility location).
- **Tactical:** Medium-term (e.g., production schedules, service planning).
- **Operational:** Daily actions (e.g., routing, loading, vessel landing).



Step 3: Focus on Marine Logistics

- ~80% of global logistics is by sea (seaborne transport).
- **Network design** (e.g., port selection, routing) is key for efficiency.

Step 4: Apply Optimization Models

- Big data helps manage **large customer networks and complex routes**.
- Example: **Multi-objective optimization** used in Taiwan's **nuclear power logistics** (Sheu, 2008).
- The author has considered risk reduction in the formulation. The result depicts the improvement of performance from 7.41% to 18.37%, and risks were also reduced by 37.75%.

Real-World Example: Optimizing Marine Logistics for a Global Shipping Company

The Company:

A global shipping company, **BlueOcean Logistics**, operates cargo ships across Asia, Europe, and the Americas. They need to **optimize their logistics network** to reduce fuel costs, improve delivery time, and meet rising customer demands.

Step 1: Strategic Level – Big Picture Decisions

- Problem:** Where should BlueOcean build its next port or distribution hub?
- Data Used:** Years of shipping data, demand forecasts, port congestion reports.
- Action:** Use big data analytics to identify **high-demand regions** (e.g., Southeast Asia) and plan for **fleet size** and **facility location**.

Step 2: Tactical Level – Planning Schedules and Services

- Problem:** How to schedule ships and assign deliveries efficiently?
- Data Used:** Cargo bookings, weather forecasts, port availability.
- Action:** Use optimization algorithms to create **production and delivery schedules**, ensuring ships arrive and depart with minimal delays.

Step 3: Operational Level – Day-to-Day Execution

- Problem:** How to choose the best route for each ship?
- Data Used:** Real-time GPS, fuel prices, sea currents, and weather.
- Action:** Use real-time routing systems to decide the **most fuel-efficient and fastest route** for each ship while avoiding storms or delays at ports.

Step 4: Big Data for Smart Decisions

- Example:** During peak holiday season, demand spikes in Europe.
- BlueOcean uses **predictive analytics** to reroute more ships to Europe, increase port staffing, and adjust fuel purchasing based on trends — all powered by big data.

Computational Method for Large-scale Unconstrained Optimization

Unconstrained optimization is widely used in engineering, industry, and economics, especially in large-scale problems where **memory usage is critical**.

It also arises from transforming constrained problems by embedding penalty terms into the objective function.

A common and effective method for large-scale unconstrained optimization is the

- **Conjugate Gradient (CG) method**, particularly the **Hestenes–Stiefel (HS)** variant.

Steps for Using the Hestenes–Stiefel Conjugate Gradient Method:

Step 1: Initialization

- Choose initial guess x_0
- Compute initial gradient $g_0 = \nabla f(x_0)$
- Set initial direction $d_0 = -g_0$

Step 2: Iteration

For each iteration k :

- Compute step size α_k (using line search)
- Update position: $x_{k+1} = x_k + \alpha_k d_k$
- Compute new gradient $g_{k+1} = \nabla f(x_{k+1})$
- Compute $\beta_k^{HS} = \frac{g_{k+1}^T (g_{k+1} - g_k)}{d_k^T (g_{k+1} - g_k)}$
- Update direction: $d_{k+1} = -g_{k+1} + \beta_k^{HS} d_k$

Step 3: Compute β (Hestenes–Stiefel formula)

Use:

$$\beta_k^{HS} = \frac{g_{k+1}^T y_k}{d_k^T y_k}, \quad \text{where } y_k = g_{k+1} - g_k$$

Step 4: Check Conjugacy Condition

Ensure:

$$d_{k+1}^T y_k = 0 \quad (\text{conjugacy condition})$$

This ensures the new direction maintains conjugacy with the previous.

Step 5: Line Search to Find Step Size α_k

Determine α_k by minimizing along direction d_k .

Step 6: Update Variables

- $x_{k+1} = x_k + \alpha_k d_k$
- $g_{k+1} = \nabla f(x_{k+1})$
- $d_{k+1} = -g_{k+1} + \beta_k d_k$

Step 7: Convergence Check

Stop if $\|g_{k+1}\|$ is below a threshold.

Example-1: Real-World Scenario: Minimizing Fuel Cost in Power Generation

Suppose you're managing a **power plant**, and your job is to **minimize fuel cost** for electricity generation.

Your fuel cost (in \$/hour) as a function of power output x (in MW) is:

$$f(x) = 5x^2 - 30x + 200$$

You want to find the value of x that minimizes fuel cost.

This is a **unconstrained optimization** problem.

Example-2: Simple Analogy: Finding the Bottom of a Valley

Imagine you're standing somewhere on a **bumpy hill**, blindfolded. Your goal is to **find the lowest point in the valley**.

- You can feel the slope under your feet (**this is the gradient**).
- You want to **walk downhill**.
- But you don't want to go back and forth in zigzags forever (which is what **steepest descent does**).
- So you use a smarter way: **Conjugate Gradient**.

Step 1: Start Somewhere (Initial Guess)

You start at a **random spot on the hill**. You don't know if it's high or low — just a guess.

Step 2: Feel the Slope (Gradient)

You feel the slope under your feet. It tells you which direction is downhill.

Step 3: Start Walking Downhill (Direction)

You start walking in the direction that takes you down fastest.

Step 4: Decide How Far to Go (Step Size)

As you walk, you keep checking if the ground is getting lower. When it stops getting lower, you stop. That's your new position.

Step 5: Feel the New Slope

You again feel the slope from the new position — this tells you the new downhill direction.

Step 6: Combine Old and New Directions

You don't just go straight down again. You **combine** the new slope with the old direction to make a smarter path — like adjusting your steering instead of turning all the way.

Step 7: Repeat Until Flat

Keep repeating these steps until you reach a spot where you **can't feel any more slope** — meaning you're at the bottom (minimum cost).

Numerical Approach for Non-smooth Large-scale Optimization

Step 1: Understand Nonsmooth Functions

- Smooth functions have a clear slope (derivative) everywhere, like a smooth curve.
- Nonsmooth functions have **sharp corners, kinks, or discontinuities**, making them harder to optimize.
- **Examples** include real-world problems in **economics, engineering, control, and data analysis**.

Step 2: Recognize the Challenges

- These problems are often **large-scale**, but even small non-smooth problems can be hard to solve.
- Standard optimization methods (like gradient descent) don't work well because the gradient may not exist everywhere.

Step 3: Use the Bundle Method

- The **Bundle Method** is a powerful technique for solving nonsmooth problems.
- It builds up a collection (bundle) of past information to guide each optimization step.
- Two main types:
 - **LMBM (Limited Memory Bundle Method)** – uses a smaller memory footprint
 - **D-Bundle (Diagonal Bundle Method)** – efficient for large problems

Step 4: Apply in Real-world Problems

- Bundle methods have been successfully used in:
 - **Power systems** (e.g., uncertainty handling, unit commitment, scheduling)
 - **Decomposition algorithms** for large systems
 - Studies in **engineering, economics, and control systems**

Big Data Analytics Based on Convex and Nonconvex Optimization

Step 1: Optimization Formulation

A single-objective optimization problem can be expressed as:

$$\min(\max) OF(x) \quad \text{s.t. } g_i(x) \leq 0, \quad i = 1, \dots, m, \quad x \in D$$

Here, x is the decision vector, D is the feasible region, OF is the objective function, and g represents constraints.

Step 2: Convex vs. Nonconvex

- If both OF and g are convex, the problem is a **convex optimization** and a global solution can be found.
- If nonconvex, only **local or approximate global solutions** are generally possible, making it more complex.

Step-3: Economic Dispatch (ED) in Power Systems

- ED aims to allocate power demand among plants economically while meeting constraints.
- The cost function (fuel cost + valve-point effects) makes ED **nonconvex**.
- Advanced **Particle Swarm Optimization (PSO)** methods (e.g., with worst-position learning and local random search) improve convergence and global solution finding.

Step-4: Multi-Objective Optimization

- Real-world problems often involve multiple conflicting objectives (e.g., cost vs. emissions).
- Since one objective's improvement may worsen another, researchers seek **Pareto optimal solutions** — compromise solutions balancing all objectives.

Multi-objective optimization seeks **Pareto solutions**, where no single solution can improve one objective without worsening another; such solutions are called **non-dominated**.

Inference:

Pareto optimization helps balance trade-offs and avoids bias toward a single objective.

Example-1: Real-World Example:

In **power systems**, minimizing **fuel cost** and **emissions** simultaneously — reducing cost may increase emissions, so Pareto solutions offer the best trade-off.

Example-2: Choosing the Fastest & Cheapest Travel Option

Imagine you want to travel from **Mysuru** to **Bengaluru**.

Your **objective**:

- Spend the **least money**
- Reach in the **least time**

Case 1: Convex Optimization (Easy Problem)

- Suppose bus fare increases steadily with speed (e.g., faster bus → slightly more fare).
- The relationship is **smooth** and **predictable**.

👉 You can easily find the **best bus** that balances cost and time → **Global Best Solution**.

Case 2: Nonconvex Optimization (Hard Problem)

• Now imagine:

- Some buses have **special discounts** 🎫 (sudden drop in fare).
 - Others have **traffic jams** 🚦 (sudden increase in travel time).
 - The cost-time curve now has **ups and downs** (not smooth).
- 👉 It's harder to find the absolute best option — you may end up with only a **good-enough local solution**

Takeaway:

- **Convex** → smooth, one clear best answer.
- **Nonconvex** → bumpy, multiple local answers, hard to find the true best.

- **TLBO** – Teaching–Learning-Based Optimization
- **PSO**-Particle Swarm Optimisation
- **GSA**-Gravitational Search Optimisation
- **OTLBO** – Orthogonal Teaching–Learning-Based Optimization
- **CPSO** – Culture Particle Swarm Optimization

Quick Comparison

- **TLBO:** Teacher–student model, simple, efficient.
- **OTLBO:** TLBO + Orthogonal design → better diversity & global search.
- **CPSO:** PSO + cultural knowledge → avoids local optima, more robust.

Metaheuristic Algorithms for Big Data Optimization

- **Big Data Challenge** – Optimization in big data involves large-scale problems (hundreds to millions of variables), where metaheuristic algorithms are effective but face performance issues with increasing dimensions.
- **Metaheuristic Strengths & Limits** – Evolutionary algorithms are powerful for global optimization but struggle with scalability, leading to the need for hybrid or advanced methods.
- **Hybrid PSO–GSA (Beigvand et al., 2017)** – Proposed for large-scale, complex power system dispatch; it outperformed PSO, TLBO, OTLBO, CPSO, and GSA in robustness, solution quality, convergence speed, and computational efficiency.
- **Renewable Energy Application** – Differential evolution (Rajesh et al., 2016) applied to solar plant optimization minimized both cost and emissions, showing promise in sustainable power systems.

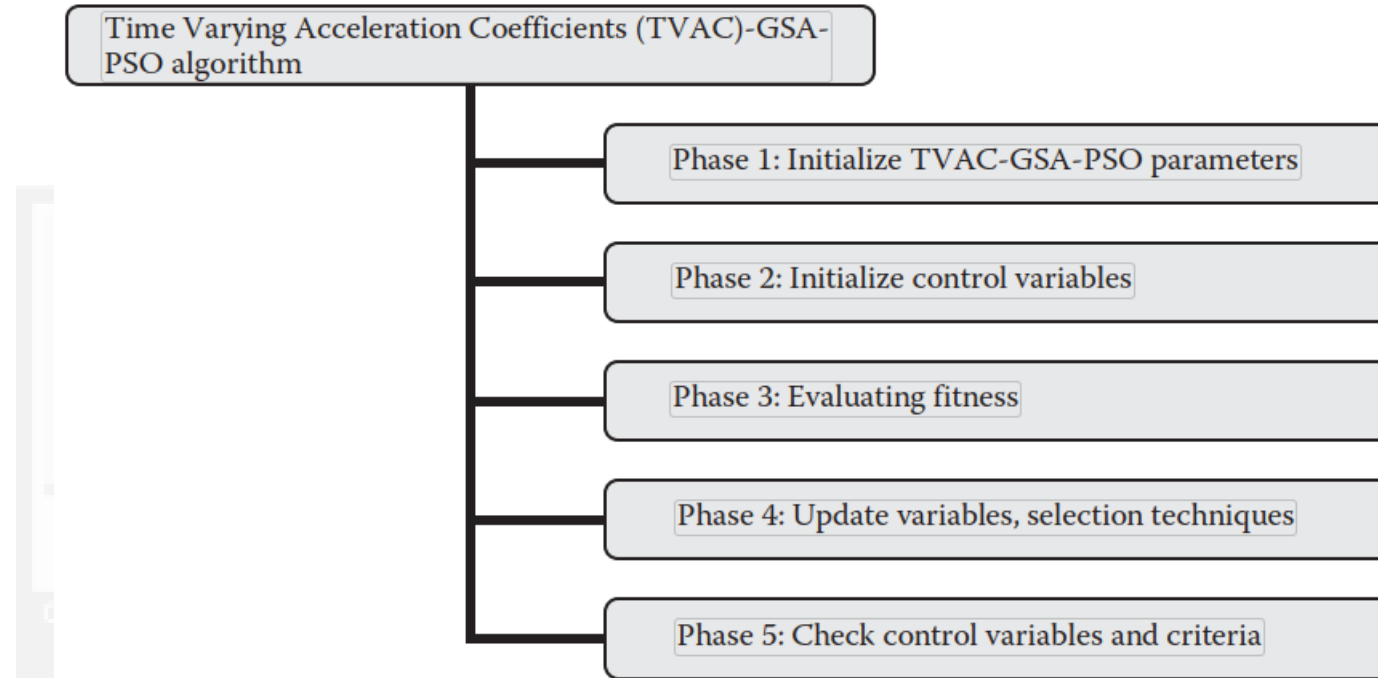
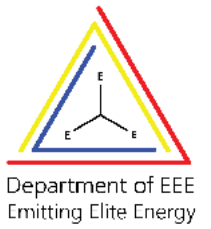


FIGURE: Phase classification for hybrid algorithm.



A T M E
College of Engineering



Thank You