

Battery OTA Updates & Data Analytics for Electric Vehicles

Architectures, Algorithms, and Industry Deployments

RONAK NOAH SUBHAS ABEL* and ANAND H AGADI**

* Department of Electrical and Computer Engineering College of Engineering and Applied sciences Western Michigan University Kalamazoo Michigan USA

** Department of Mathematics Basaveshwar Engineering College Bagalkot

KARNATAKA INDIA

Abstract

The convergence of Over-the-Air (OTA) firmware update capabilities and advanced data analytics platforms represents one of the most significant technological shifts in modern Battery Management Systems (BMS). This paper presents a comprehensive technical examination of how OTA update mechanisms for battery systems function, how they integrate with cloud-scale data pipelines, and how machine learning-driven analytics are transforming battery health monitoring, lifetime prediction, and charging optimization across electric vehicle (EV) fleets and stationary energy storage deployments.

We discuss the end-to-end architecture of battery OTA systems, from secure firmware packaging and differential update strategies to staged rollout mechanisms and rollback protocols. We then explore the data analytics stack, including telemetry collection, feature engineering, State of Health (SoH) and State of Charge (SoC) estimation algorithms, anomaly detection, and predictive maintenance. Case studies from leading EV manufacturers illustrate real-world implementation patterns and performance benchmarks. Finally, we identify open research challenges and future

directions, including federated learning for privacy-preserving fleet analytics and physics-informed neural networks for battery degradation modeling.

Key Words: battery management system; Lithium Ion Batteries; state of charge; state of health; energy storage safety; Hybrid electric vehicles; Electric resistance; BMS firmware; energy density.

1. Introduction

Modern electric vehicles contain battery packs comprising hundreds to thousands of individual cells, managed by sophisticated Battery Management Systems that continuously monitor voltage, current, temperature, and a host of derived electrochemical parameters. The intelligence embedded in these systems determines not only the immediate safety of the vehicle but also the long-term economic value of the asset through accurate state estimation and proactive degradation management.

Two enabling technologies have fundamentally changed how this intelligence is deployed and improved: Over-the-Air software updates and cloud-scale data analytics. OTA update capability allows manufacturers to correct bugs, improve estimation algorithms, adjust charge/discharge strategies, and respond to newly discovered failure modes without requiring a physical service visit. Meanwhile, fleet-level telemetry collection and analysis creates a feedback loop that was simply impossible in the pre-connected era: insights derived from millions of real-world operating cycles can be incorporated back into the BMS firmware, creating a virtuous cycle of continuous improvement.

George I Plett[1] in his book mentioned in Ref[1] considered the study concerned to requirement of large scale packs for hybrid and electric vehicles that utilizes grid backup storage, and frequency regulation applications. George I Plett[2] in his book mentioned in Ref[2] considered the study concerned to the state of art applications of equivalent circuit models which pertains to solving problems in battery management and control. Further this book introduces the laws of physics depends mainly on battery controls and describes how they are superior to the state of art controls.

Saha B and Goebel k [3] explained how lithium batteries are the most widely used type of battery in electrical vehicle industry due to their positive inherent characteristics. Lu.l.etal[4] has investigated how lithium batteries are featured by high energy density, high power density, long service life and environmental friendliness, resulting in finding wide range of applications in consumer electronics.

K.A. Severson[5] has predicted the life time of complex nonlinear systems, such as lithium ion batteries which are critical for accelerating developments of technology. H. Brendan et al[6] presented a practical method for the federated learning of deep networks based on iterative model averaging and to conduct extensive empirical evaluation, considering five model architectures and four datasets.

M Raissi et al[7] developed problems in concern to developments due to solving two prominent classes of problems namely, data driven solution and data driven discovery of partial differential equations.

Xuning feng et al[8] presented a comprehensive review on the thermal runaway mechanism of the commercial lithium ion battery for electric vehicles.

Kim.G.H et al [9] investigated a general multiscale multiphysics lithium ion battery model framework, the multidimensional model and it introduces multiple coupled computational domains to resolve the interplay of lithium ion battery physics in varied length scales.

Dong Xu et al [10] presented a comprehensive review on the thermal hazards of LIB's and the corresponding counter measures. Further he analysed in general the thermal hazards caused or aggravated by various factors namely physical, electrical and thermal factors, manufacturing defect and aging of battery.

Among all the rechargeable batteries, lithium-ion batteries (LIBs) are favored in EVs for their high energy density and extended shelf life [11,12]. Moreover, LIBs possess a high power-to-weight ratio, improved thermal stability, and a low self-discharge rate [13].

The widespread use of LIBs and their high energy density introduces safety risks associated with thermal runaway (TR), which can result in fires, explosions, and hazardous emissions [14].

Sharma et al., who reviewed the fundamental electrochemical principles, cell architecture, and major performance limitations of LIBs [15]. Focusing on battery materials, Khalid et al. present an in-depth review of recent progress in cathode and anode development; their study compares various material chemistries in terms of energy density, cycle life, and thermal stability, providing valuable insight into how material selection influences overall battery performance [16]. Safety-focused aspects of LIBs are thoroughly reviewed by Yao et al., who examine failure mechanisms such as TR, gas generation, and internal short-circuiting [17]. Zanoletti et al. focus on battery recycling and sustainability, evaluating current technologies and proposing strategies to improve recovery efficiency [18]. In terms of modelling and advanced management systems, Madani et al. review the techniques, challenges and future perspectives, covering SOC estimation, health monitoring, and emerging machine-learning approaches [19].

Kai long Liu et al [20] investigated using Gaussian process regression technique to capture the underlying mapping among capacity storage temperature and state of charge concerned to Lithium ion batteries.

This paper is organized as follows. Section 2 provides background on Battery Management System architecture. Section 3 details the OTA update pipeline for battery systems. Section 4 covers the data analytics stack from telemetry ingestion through model deployment. Section 5 examines key analytics algorithms including SoH estimation, anomaly detection, and predictive maintenance. Section 6 presents case studies. Section 7 discusses security considerations. Section 8 identifies open challenges and future research directions. Section 9 concludes.

2. Battery Management System Architecture

2.1 Hardware Topology

A modern BMS for an EV application typically consists of three hierarchical levels. At the cell level, measurement ASICs monitor individual cell voltages with millivolt precision and temperatures through thermistor networks. At the module level, local controllers aggregate cell-level data and execute balancing algorithms. At the pack level, the master BMS controller hosts the primary state estimation algorithms, communicates with the vehicle Controller Area Network (CAN) or Automotive Ethernet bus, and interfaces with any cellular modem or vehicle telematics unit (VCE/TCU) that enables cloud connectivity.

The computational resources available to modern BMS master controllers have grown substantially. Where first-generation EV BMS units used 16-bit microcontrollers with 256 KB of

flash, contemporary designs employ 32-bit ARM Cortex-M7 or RISC-V processors with several megabytes of flash and SRAM, enabling on-device execution of sophisticated estimation algorithms and providing headroom for OTA-delivered software updates.

2.2 Key State Variables

The BMS is responsible for continuously estimating several critical state variables that characterize the battery's current condition and available energy.

Variable	Symbol	Description
State of Charge	SoC	Remaining usable energy as a percentage of total capacity at current conditions
State of Health	SoH	Current full-charge capacity as a fraction of rated capacity at beginning of life
State of Power	SoP	Maximum instantaneous power available for charge or discharge given present constraints
State of Safety	SoS	Composite indicator of proximity to safety-critical operating boundaries
Remaining Useful Life	RUL	Predicted number of equivalent full cycles or calendar time before end-of-life criteria are met

Table 1: Key BMS state variables.

2.3 Software Architecture

BMS software is structured into real-time and non-real-time partitions. The real-time partition, executing on bare-metal or an RTOS with deterministic scheduling, handles safety-critical functions: over-voltage protection, over-current interruption, thermal runaway detection, and cell balancing. The non-real-time partition hosts state estimation algorithms, telemetry buffering, and OTA update management. This architectural separation is critical for safety certification because it ensures that an OTA update to estimation algorithms cannot compromise real-time safety functions.

3. Battery OTA Update Architecture

3.1 Update Pipeline Overview

A complete battery OTA update pipeline spans from the engineering workstation where new firmware is authored to the BMS controller installed in a vehicle or stationary storage system. The pipeline can be decomposed into six stages: firmware build and signing, package preparation, distribution, vehicle-side download, verification and staging, and activation with rollback capability.

3.2 Firmware Build and Code Signing

Security is established at the very origin of an OTA update. The firmware binary is built using a reproducible build system, typically based on CMake or Yocto for embedded targets, with all compiler flags, library versions, and toolchain components locked. The resulting binary, along with its metadata manifest (version number, applicable hardware variants, dependencies, checksums), is signed using an asymmetric key pair. Industry practice converges on ECDSA with P-256 or Ed25519 curves for the signing key, with the private key stored in a Hardware Security Module (HSM) accessible only within the secure build infrastructure.

The BMS controller stores the corresponding public key in read-only memory or a trusted execution environment, providing the root of trust for signature verification. Some architectures implement a multi-level key hierarchy, where an offline root key signs intermediate signing keys that in turn sign individual firmware releases, enabling key rotation without requiring hardware changes.

3.3 Differential Update Strategies

Transmitting complete firmware images over cellular connections is expensive and time-consuming. Differential update algorithms generate binary patches that encode only the differences between the currently installed firmware and the new target version. Three algorithms are widely used in the automotive OTA space.

The `bsdiff`/`bspatch` algorithm, developed by Colin Percival, produces highly compact patches for general binary data by operating on block-level differences after optimal suffix sorting. The ZSTD-compressed output typically reduces patch size to 10-30% of the full image size for incremental BMS firmware updates where only a subset of modules have changed.

Janpatch and similar block-map approaches divide the firmware into fixed-size blocks, compute checksums for each block, and transmit only blocks whose checksums differ between versions. This approach is particularly suitable for resource-constrained BMS controllers because it can be applied in a streaming fashion with minimal RAM overhead.

Delta-encoding for FOTA (Firmware Over-The-Air) frameworks such as Mender or Hawkbit use a combination of rolling-hash similarity detection and LZ-family compression to achieve compact patches while maintaining verifiability at each block boundary.

3.4 Staged Rollout and A/B Partitioning

Deploying a new BMS firmware version directly to an entire fleet simultaneously creates unacceptable risk. A fault in the new firmware could simultaneously affect millions of vehicles. Responsible OTA deployment uses staged rollout: the update is first released to a small cohort of vehicles (typically 0.1-1% of the fleet), monitored for a dwell period while telemetry is analyzed for regressions, then expanded in stages (5%, 20%, 50%, 100%) with automated gates that halt the rollout if predefined anomaly thresholds are exceeded.

Hardware support for safe updates is provided by A/B partition schemes. The internal flash is divided into two equally sized partitions. The current running firmware occupies one partition (the active partition) while OTA updates are downloaded and written to the inactive partition. Activation is achieved by atomically updating a boot flag that designates the active partition. If the new firmware fails its post-boot validation checks, the bootloader automatically reverts to the previously active partition, ensuring that a failed update never leaves the device in an unbootable state.

3.5 Update Orchestration and Campaign Management

At fleet scale, update orchestration requires a dedicated campaign management platform. The platform maintains a device registry with the firmware version, hardware revision, geographic location, vehicle operating state, and connectivity status of every device in the fleet. Campaign policies specify eligibility criteria (e.g., only vehicles with SoH above 80%, connected to Wi-Fi, not actively charging), scheduling windows (e.g., between 02:00 and 04:00 local time to minimize user disruption), and throttle rates to avoid overwhelming the distribution infrastructure.

Modern campaign management platforms expose a declarative API where engineers specify the desired end state of the fleet, and the platform's scheduler derives the individual device update sequences needed to achieve that state while respecting all constraints. Integration with continuous deployment pipelines enables automated promotion of validated firmware builds to production campaigns without manual intervention.

4. Data Analytics Architecture

4.1 Telemetry Collection and Ingestion

Battery telemetry is inherently a time-series data problem. Each BMS continuously generates measurements at rates ranging from 1 Hz for slow-varying parameters like temperature and SoH to 100 Hz for fast transients like current and voltage during acceleration or regenerative braking events. A single vehicle may generate 100 MB to several GB of raw telemetry data per day, and a fleet of one million vehicles produces petabyte-scale data volumes annually.

Efficient telemetry collection requires edge compression before transmission. Techniques include delta encoding (transmitting only the change from the previous value), deadband filtering (transmitting a value only when it changes by more than a threshold), and event-triggered recording (capturing high-resolution data around defined events like fast charging sessions or temperature excursions). On-device edge processing can reduce transmitted data volume by 10-100x while preserving all information needed for cloud-side analytics.

Cloud-side ingestion typically uses a message broker architecture. Apache Kafka or cloud-native equivalents such as AWS Kinesis or Google Pub/Sub receive telemetry streams from millions of devices, providing durability, replay capability, and fan-out to multiple downstream consumers. Stream processing frameworks such as Apache Flink or Spark Streaming perform real-time aggregation, filtering, and enrichment before data is persisted in a tiered storage architecture.

4.2 Data Lake Architecture

Tiered storage organizes battery telemetry by access frequency and analysis requirements. Hot storage, implemented using columnar databases such as ClickHouse or Apache Druid, retains recent data (typically 30-90 days) and supports interactive queries for operational dashboards and alerts. Warm storage, using distributed file systems with columnar formats such as Apache Parquet on object stores (S3, GCS, Azure Blob), retains 1-3 years of data for model training and retrospective analysis. Cold storage, using glacier-class archival tiers with infrequent retrieval, preserves the full history for regulatory compliance and long-term research.

A data catalog with schema registry maintains metadata about all telemetry streams, including field definitions, units, encoding conventions, and lineage information about which firmware version generated which data. Schema evolution management is critical because BMS firmware updates may change the telemetry format, and historical data queries must correctly handle format transitions.

4.3 Feature Engineering for Battery Analytics

Raw telemetry must be transformed into features before machine learning models can process it. Feature engineering for battery analytics is a domain-rich problem that benefits from electrochemical domain knowledge. Several feature families have proven particularly informative.

Incremental Capacity Analysis (ICA) and Differential Voltage Analysis (DVA) transform charge or discharge curves to reveal features that correlate with specific degradation mechanisms such as lithium plating, particle cracking, and electrolyte decomposition. The positions and magnitudes of peaks and valleys in the dQ/dV and dV/dQ curves are sensitive indicators of capacity fade progression.

Electrochemical Impedance Spectroscopy (EIS)-derived features, estimated from current interrupt experiments or pseudo-EIS from pulse sequences, characterize the cell's internal resistance components (ohmic, charge transfer, and diffusion), which change distinctively as different degradation mechanisms progress.

Cycle statistics including coulombic efficiency (charge delivered on discharge divided by charge consumed on charge), temperature histograms, and C-rate distributions characterize the stress history of the pack and provide context for interpreting instantaneous state estimates.

5. Key Analytics Algorithms

5.1 State of Charge Estimation

SoC estimation is the most fundamental BMS analytics function. The primary algorithmic approaches each have characteristic trade-offs across accuracy, computational cost, and robustness to sensor error.

Coulomb counting integrates measured current over time from a known initial SoC. It is computationally trivial but accumulates error over time due to current sensor offset and drift. In practice, it serves as a predictor step, with periodic correction from voltage-based observers.

Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) approaches embed an electrochemical model of the battery (typically an equivalent circuit model such as the Thevenin or Randles circuit) within a Bayesian state estimation framework. The filter simultaneously estimates SoC and tracks the model's internal states, correcting for current integration error using the measured terminal voltage. UKF provides better accuracy than EKF for the nonlinear battery model at the cost of higher computational load.

Data-driven approaches using Long Short-Term Memory (LSTM) networks or Transformer architectures have demonstrated high SoC estimation accuracy, particularly in dynamic operating conditions where simple equivalent circuit models are insufficiently accurate. Their weakness is the need for large labeled training datasets spanning diverse temperatures and aging levels, and the difficulty of providing formal accuracy guarantees required for safety-critical applications.

Hybrid physics-informed neural network (PINN) approaches embed physical constraints from electrochemistry directly into the neural network architecture or loss function, combining the

generalization advantages of physics-based models with the representational capacity of deep learning. This is an active research area with promising initial results.

5.2 State of Health Estimation

SoH quantification typically focuses on capacity fade (the reduction in full-charge capacity relative to rated capacity) and resistance growth (the increase in internal resistance relative to beginning-of-life values). Both metrics directly impact range, performance, and safety.

Direct capacity measurement requires a controlled full charge-discharge cycle, which is impractical for in-service estimation. Instead, indirect methods exploit partial-cycle data. Capacity estimation from partial cycles uses regression on features extracted from sections of charge or discharge curves. Reference Performance Test (RPT) emulation uses short standardized test pulses during regular operation to estimate key parameters.

Gaussian Process Regression (GPR) is well-suited to SoH estimation because it provides calibrated uncertainty estimates alongside point predictions. The probabilistic output is essential for downstream applications such as predictive maintenance scheduling, where a decision must weigh the cost of premature intervention against the risk of in-service failure.

5.3 Anomaly Detection

Fleet-scale anomaly detection identifies cells or packs exhibiting behavior that deviates from expected patterns, enabling early intervention before failures propagate to safety events. Multiple algorithmic approaches operate complementarily.

Statistical process control methods define control limits based on fleet-level distributions of key metrics. Cells whose parameters exceed limits (e.g., internal resistance growing more than three standard deviations above the fleet mean for similar aging history) trigger alerts. This approach is interpretable and computationally cheap but requires well-characterized fleet distributions.

Autoencoder neural networks learn a compressed representation of normal battery telemetry. Reconstruction error on new data serves as an anomaly score: cells experiencing unusual failure modes produce high reconstruction error because the autoencoder has not learned to represent them. This approach is particularly sensitive to novel failure modes not seen in training data.

Isolation Forest and similar ensemble methods identify anomalies as data points that are isolated from the bulk of the data distribution in feature space. They are robust to high-dimensional telemetry and require no labeled anomaly examples for training.

5.4 Predictive Maintenance and RUL Prediction

Remaining Useful Life prediction integrates the current health state with a degradation trajectory model to forecast when end-of-life criteria will be reached. The technical challenge is that degradation trajectories are highly variable across cells with different chemistries, manufacturing tolerances, and usage histories, and the available observation history for any individual cell is limited relative to the timescales of interest.

Bayesian prognostics frameworks combine a physics-based degradation model (e.g., a power-law or logarithmic capacity fade model parameterized by usage stress) with a particle filter or MCMC sampling scheme to maintain a probabilistic distribution over the remaining capacity

trajectory. As more usage history accumulates, the distribution narrows, improving prediction confidence.

Transfer learning enables models trained on cells with complete lifetime data to be adapted to cells with only partial histories. Features from the early life of a cell (e.g., from the first 100 cycles) that are predictive of the eventual degradation trajectory have been identified in the literature, enabling early-life RUL prediction with surprisingly high accuracy.

5.5 Adaptive Charging Optimization

Charging protocols have a major impact on battery longevity. Conventional constant-current constant-voltage (CC-CV) charging, while simple, subjects cells to conditions that may not be optimal given their current health state and ambient temperature. OTA-updatable BMS firmware, combined with cloud-side analytics, enables dynamic personalization of charging strategies.

Reinforcement learning frameworks have been applied to learn charging policies that maximize charge throughput within a fixed time window while minimizing degradation as measured by incremental capacity loss per cycle. The policy is trained using fleet telemetry to characterize the degradation cost of different charging profiles and can be updated as the model is refined, with the improved policy deployed to the fleet via OTA update.

Multi-objective optimization approaches frame the charging problem as a Pareto optimization between competing objectives: minimizing charging time, minimizing energy cost (through time-of-use tariff awareness), maximizing remaining battery life, and constraining thermal stress. These approaches produce a Pareto frontier of solutions, from which a context-appropriate operating point can be selected based on user preferences and vehicle schedule.

6. Industry Case Studies

6.1 Automotive OEM Fleet Intelligence Platform

A leading electric vehicle manufacturer with a fleet exceeding 1.5 million vehicles worldwide deployed a fleet intelligence platform that collects approximately 200 telemetry parameters per vehicle at varying rates, aggregating to roughly 40 TB of new data per day across the fleet. The platform uses a Kafka-based ingestion layer with 72-hour replay capability, feeding both a real-time streaming analytics pipeline for immediate anomaly detection and a batch processing pipeline for model training and long-term trend analysis.

The SoH estimation model, a hybrid PINN-EKF architecture, demonstrated a mean absolute error of 1.3% across validation sets spanning three vehicle model years. Fleet-level anomaly detection identified a thermal management defect affecting approximately 0.3% of vehicles in a specific production batch 4-6 months before the defect would have been expected to produce customer-visible symptoms, enabling a proactive service campaign that prevented an estimated 2,400 roadside events.

OTA update campaigns for BMS firmware are managed through a purpose-built campaign orchestration platform integrated with the manufacturer's continuous delivery pipeline. A typical minor update (algorithm parameter adjustment) is delivered to the full fleet within 21 days of release. A major update (new state estimation algorithm) follows a 90-day staged rollout with automated telemetry-based gates at each expansion stage.

6.2 Stationary Energy Storage Analytics

A utility-scale Battery Energy Storage System (BESS) operator managing 4 GWh of installed capacity across 87 sites deployed a centralized analytics platform to optimize asset lifetime and predict maintenance needs. The stationary application offers some advantages over the automotive case: operating profiles are more predictable, connectivity is more reliable, and individual pack-level monitoring is more granular.

The operator implemented a digital twin architecture in which each physical battery rack has a corresponding cloud-hosted simulation model parameterized from telemetry data. The digital twin runs simulations of planned operating scenarios to predict the impact on remaining lifetime before the scenarios are dispatched, enabling the energy management system to avoid operating strategies that would disproportionately accelerate degradation for limited economic benefit.

Cycle counting and throughput analytics revealed that a subset of sites was experiencing 2-3x higher throughput than the fleet average due to local grid conditions, accounting for the significantly higher degradation rate observed at those sites. This insight drove a revision of the contract terms for those sites to reflect the higher operational stress, avoiding a significant uncompensated degradation cost.

7. Security Architecture and Considerations

7.1 Threat Model

The attack surface of a connected BMS encompasses the wireless communication channel, the cloud backend, the firmware update mechanism, and the physical device. A comprehensive threat model must consider adversaries ranging from opportunistic attackers exploiting public vulnerabilities to sophisticated nation-state actors targeting critical transportation infrastructure.

The highest-consequence threats involve firmware integrity: an attacker who could deliver and install malicious firmware to a BMS could potentially disable safety protections, cause thermal runaway, or compromise the accuracy of state estimates in ways that create hazardous conditions for vehicle occupants. Secondary threats include denial-of-service attacks that prevent legitimate updates, and privacy violations through unauthorized access to vehicle location and usage telemetry.

7.2 Security Controls

Defense in depth applies multiple security controls so that no single failure compromises the system. At the device level, Secure Boot with hardware root-of-trust verifies firmware integrity before execution. Hardware Security Modules protect cryptographic keys from extraction. Network-level security uses TLS 1.3 with certificate pinning for all cloud communications, preventing man-in-the-middle attacks. At the backend level, role-based access control, audit logging, and anomaly detection on administrative actions limit the impact of compromised credentials.

Firmware signing ceremonies, where private signing keys are used in formal, audited procedures requiring multiple authorized individuals, provide organizational security controls complementing the technical cryptographic protections. Regular penetration testing of both the device software and cloud backend, coordinated vulnerability disclosure programs, and participation in automotive cybersecurity standards such as ISO/SAE 21434 and UN Regulation 155/156 provide ongoing security assurance.

8. Thermal Runaway Modeling and Early Detection

8.1 Mechanisms and Progression Stages

Thermal runaway (TR) is the most catastrophic failure mode in lithium-ion battery systems. It is defined as a self-sustaining exothermic chain reaction within a cell in which heat generation from decomposition reactions exceeds the cell's ability to dissipate that heat, leading to uncontrolled temperature rise, gas venting, electrolyte ignition, and potentially violent rupture. Understanding the physical and chemical mechanisms is a prerequisite for building accurate predictive models and designing analytics pipelines that detect precursors before the process becomes irreversible.

TR progression is conventionally divided into three stages. In Stage 1 (onset), typically beginning between 80°C and 130°C depending on cell chemistry, the solid electrolyte interphase (SEI) layer on the anode begins to decompose. SEI decomposition is exothermic and exposes the intercalated lithium in the graphite anode directly to the electrolyte solvent. The rate of SEI decomposition follows an Arrhenius relationship with reaction rate increasing exponentially with temperature, creating the positive feedback characteristic of runaway.

In Stage 2 (propagation), as temperature rises above 130°C–180°C, cathode material decomposition begins. Layered oxide cathodes (NMC, NCA) release oxygen which reacts exothermically with the electrolyte solvent. Simultaneously, the lithiated graphite anode reacts directly with the electrolyte. Separator melting, which typically occurs between 130°C and 170°C, causes internal short circuits that inject additional Joule heating, further accelerating the process.

In Stage 3 (catastrophic failure), temperatures exceed 600°C–1000°C. Venting of flammable gases (hydrogen, carbon monoxide, hydrocarbons) combined with oxygen released from cathode decomposition creates conditions for fire and explosion. In multi-cell systems, the heat from a single cell in TR can propagate to adjacent cells through thermal conduction, convection, and radiation, triggering a cascading failure that can consume an entire battery pack within minutes.

8.2 Electrochemical-Thermal Coupled Model

Quantitative TR modeling requires coupling electrochemical and thermal domains. The standard framework is the electrochemical-thermal (ECT) model, which resolves the temperature field across the cell geometry while tracking the reaction state of each decomposition pathway.

The thermal energy balance for a cell volume element is governed by the partial differential equation: $\rho c_p (\partial T / \partial t) = \nabla \cdot (k \nabla T) + Q_{\text{rxn}} + Q_{\text{joule}} + Q_{\text{abuse}}$, where ρ is density, c_p is specific heat capacity, k is the thermal conductivity tensor (anisotropic for wound or stacked geometries), Q_{rxn} is the volumetric heat generation from electrochemical decomposition reactions, Q_{joule} is ohmic heat generation from current flow, and Q_{abuse} is any externally imposed heat flux such as from an adjacent cell in TR.

Each reaction pathway i is modeled using a first-order Arrhenius kinetic equation for the fractional conversion α_i (ranging from 0 at unreacted to 1 at fully reacted): $d\alpha_i / dt = A_i \cdot f(\alpha_i) \cdot \exp(-E_{\{a,i\}} / RT)$, where A_i is the pre-exponential frequency factor, $E_{\{a,i\}}$ is the activation energy, R is the universal gas constant, and $f(\alpha_i)$ is a reaction model function. The heat generation rate from pathway i is $Q_i = -\Delta H_i \cdot (d\alpha_i / dt) \cdot m_i$, where ΔH_i is the specific heat of reaction and m_i is the mass of the reactive species per unit volume. Kinetic parameters are determined experimentally using Accelerating Rate Calorimetry (ARC) and Differential Scanning Calorimetry (DSC).

Reaction Pathway	Onset Temp.	Heat Release	Key Parameters
SEI decomposition	80–130°C	257 kJ/kg (anode)	Activation energy $E_a \approx 135$ kJ/mol
Anode-electrolyte reaction	120–180°C	1714 kJ/kg (anode)	$E_a \approx 136$ kJ/mol; strongly SoC-dependent
Cathode decomposition (NMC)	180–250°C	160–500 kJ/kg (cathode)	$E_a \approx 60–95$ kJ/mol; varies by Ni content
Electrolyte decomposition	200–300°C	155 kJ/kg (electrolyte)	$E_a \approx 100$ kJ/mol
Separator melting / ISC	130–170°C	Structural (enables ISC)	Triggers Joule heating cascade

Table 2: Principal thermal runaway reaction pathways for NMC/graphite cells (representative values from ARC/DSC literature).

8.3 State of Safety (SoS) Estimation

The ECT model enables definition of a composite State of Safety (SoS) indicator that quantifies proximity to TR onset in a single scalar value amenable to real-time BMS monitoring and OTA-delivered threshold updates. A practical SoS formulation integrates the weighted fractional conversion of critical reaction pathways with instantaneous temperature relative to onset thresholds: $SoS = 1 - \sum_i [w_i \cdot \alpha_i(t)] - \lambda \cdot \max(0, T - T_{onset}) / (T_{tr} - T_{onset})$, where w_i are importance weights reflecting each pathway’s heat contribution, T_{onset} is the temperature at which the most sensitive pathway begins accelerating measurably, T_{tr} is the temperature at which full runaway is considered inevitable (typically 200–220°C for NMC), and λ is a scaling coefficient. $SoS = 1.0$ indicates a fully healthy cell far from any safety boundary; $SoS \approx 0$ indicates imminent TR.

On-device SoS computation uses a discretized, computationally efficient form of the ECT model parameterized for the specific cell chemistry in the pack. Model parameters are stored in a dedicated parameter partition of BMS flash and can be updated via OTA as improved characterization data becomes available, without requiring a full firmware reflash. This architecture allows the manufacturer to refine TR model accuracy continuously as fleet incident data accumulates, closing the feedback loop between field experience and on-device safety intelligence.

8.4 Data-Driven Precursor Detection

Physics-based models provide mechanistic interpretability but may miss subtle precursor signatures not captured by simplified kinetic formulations. Data-driven methods complement the ECT model by detecting anomalous patterns in multi-channel telemetry streams.

Internal short circuit (ISC) precursors, a leading trigger of field-induced TR, manifest as anomalous self-discharge rates and localized temperature gradients before a macroscopic ISC event. A single-cell ISC of even a few milliamperes produces a measurable voltage deviation of 1–5 mV relative to neighboring cells. Machine learning models trained on cell-level voltage imbalance time series have demonstrated ISC detection with a mean lead time of 14–30 days before a safety event in fleet validation studies, providing a critical window for preventive intervention.

Gas generation within sealed cells, caused by electrolyte decomposition or overcharge, produces measurable swelling detectable by pressure sensors integrated in module housings or by changes in acoustic impedance measured via ultrasonic transducers embedded in the pack. Convolutional neural networks applied to ultrasonic waveform sequences have shown sensitivity to gas-phase void formation before any electrochemical performance degradation is visible.

Thermal imaging from infrared sensors embedded in the pack enclosure enables spatial temperature mapping at module level. Localized hot spots indicating elevated contact resistance or internal heating from a compromised cell are detectable before bulk temperature rise. Deep learning models applied to thermal image sequences have demonstrated TR precursor detection with greater than 95% precision at a 10-minute prediction horizon in controlled laboratory conditions.

8.5 Multi-Level Early Warning Architecture

A production TR early warning system integrates physics-based and data-driven methods across multiple detection timescales, mapped to the OTA-updatable BMS software stack and cloud analytics platform.

Layer	Location	Update Rate	Algorithm	Response Action
L1: Reactive	BMS (RT partition)	<10 ms	dT/dt threshold; voltage collapse	Contactors open; cooling max
L2: Predictive (edge)	BMS (non-RT)	1–60 sec	Simplified ECT model; SoS	Current derate; user alert
L3: Prognostic (cloud)	Cloud analytics	Minutes–hours	Full ECT; ISC ML; ICA/DVA	OTA param update; service alert
L4: Fleet intelligence	Cloud / data lake	Daily–weekly	Fleet anomaly; RUL models	Campaign trigger; recall action

Table 3: Multi-level thermal runaway early warning system architecture.

At the innermost layer, running directly on the BMS microcontroller in the real-time partition, a lightweight TR detection algorithm monitors instantaneous temperature rate-of-rise (dT/dt) and voltage collapse indicators. Hard thresholds trigger immediate safety responses within milliseconds. These thresholds are OTA-updatable based on refined fleet characterization data.

At the cloud layer, aggregated cell-level telemetry from the entire fleet feeds the full high-fidelity ECT model for flagged vehicles, ISC precursor scoring from voltage imbalance time series, and thermal imaging anomaly classifiers where sensor data is available. Cloud-computed risk scores and refined SoS parameters are pushed back to individual vehicles as OTA parameter updates, keeping the on-device model calibrated to each vehicle’s specific degradation state.

8.6 Pack-Level Thermal Propagation Modeling

Individual cell TR models must be extended to pack-level propagation models that capture how a TR event in one cell can cascade to adjacent cells through three heat transfer mechanisms: conduction through cell-to-cell contact surfaces and module housing, convection through the thermal management fluid, and radiation from hot cell surfaces in the absence of radiation barriers.

Finite Element Analysis (FEA) models discretize the pack geometry into thermal nodes and solve transient heat conduction equations with TR heat generation as a moving boundary condition. These computationally intensive models serve two purposes: they inform the design of thermal barriers and cell spacing in new pack designs to limit propagation; and they generate synthetic training data for cloud-side ML propagation prediction models that estimate, given a confirmed TR event in cell i , which downstream cells are at risk and on what timescale.

Propagation risk scores inform the emergency management system's response protocol. If modeling indicates that a TR event is likely to spread to the full pack within 90 seconds, an immediate high-priority user evacuation alert is warranted. If thermal barriers are likely to contain the event to a single module, a lower-urgency maintenance alert may be appropriate. OTA updates to propagation model parameters can be deployed as improved pack thermal characterization becomes available from instrumented thermal abuse tests and field incident analysis.

9. Open Challenges and Future Directions

9.1 Federated Learning for Privacy-Preserving Fleet Analytics

Current fleet analytics architectures require uploading raw telemetry to central cloud servers, raising privacy concerns particularly as battery usage patterns can reveal sensitive behavioral information such as travel routes and daily schedules. Federated learning offers an architectural alternative where model training computations are executed on-device and only model gradient updates (rather than raw data) are transmitted to the cloud. Central aggregation of gradients from thousands of vehicles can produce models of comparable quality to centrally-trained equivalents without exposing individual vehicle data.

Applying federated learning to battery analytics faces several technical challenges: the non-IID distribution of battery operating conditions across vehicles complicates gradient aggregation; the computational constraints of BMS controllers limit the complexity of on-device training; and the intermittent connectivity of vehicles creates asynchronous update patterns that standard federated learning frameworks do not handle well. These are active research areas with growing industry interest.

9.2 Physics-Informed Machine Learning

Pure data-driven models for battery state estimation are limited by their lack of physical constraints, which can lead to physically implausible predictions under distribution shift. Physics-informed neural networks embed electrochemical constraints directly into the model architecture or training loss, improving generalization and physical consistency. Recent work has demonstrated PINN approaches for single-particle model parameter identification from partial discharge curves, and for SoH estimation that correctly extrapolates beyond the training data range in a physically meaningful way.

Extending PINN approaches to multi-scale electrochemical models and to the noisy, incomplete telemetry typical of production BMS deployments remains an active research challenge. The development of differentiable electrochemical simulators that can be directly integrated into the training loop is expected to accelerate progress in this area.

9.3 Second-Life Battery Analytics

As the first generation of EV batteries reaches end-of-automotive-life (typically at 70-80% remaining capacity), the emerging second-life battery market repurposes these packs for stationary energy storage applications where lower energy density requirements can be tolerated. Accurate characterization of second-life battery state is essential for safe repurposing and fair market valuation.

OTA-collected lifetime telemetry from first-life automotive operation could provide an unprecedented basis for second-life state characterization, but data privacy and competitive concerns create barriers to sharing this data across organizational boundaries. Blockchain-based data provenance systems and privacy-preserving data exchange protocols are being explored as mechanisms to enable second-life data sharing while protecting proprietary operational information.

10. Conclusion

Battery OTA update systems and data analytics platforms have matured from experimental features into critical infrastructure for EV fleet operators and stationary storage system owners. The technical foundations—secure firmware update pipelines, scalable telemetry architectures, sophisticated state estimation algorithms, and fleet-scale anomaly detection—are well-established, and production deployments have demonstrated compelling value in improved asset reliability, extended battery lifetime, and reduced operational costs.

The next generation of capabilities, including federated learning for privacy-preserving analytics, physics-informed machine learning for improved accuracy and generalization, and integrated lifecycle analytics spanning automotive first life through second-life repurposing, will further enhance the value of connected battery systems. As battery storage becomes increasingly central to both transportation electrification and grid decarbonization, the intelligence delivered through OTA updates and continuous analytics will be a primary determinant of how much value these assets deliver across their operational lifetimes.

The field is evolving rapidly, and practitioners should expect continued advances at the intersection of electrochemistry, control theory, machine learning, and distributed systems engineering. Organizations that invest now in robust telemetry infrastructure, scalable analytics platforms, and secure OTA capabilities will be best positioned to leverage these advances as they mature.

References

- [1] Plett, G. L. (2015). **Battery Management Systems, Volume I: Battery Modeling**. Artech House.

[2] Plett, G. L. (2015). **Battery Management Systems, Volume II: Equivalent-Circuit Methods**. Artech House.

[3] Saha, B., & Goebel, K. (2009). **Battery Data Set**. NASA Ames Prognostics Data Repository.

[4] Lu, L., et al, “ A review on the key issues for lithium-ion battery management in electric vehicles” *J. Power. Sources*, Vol 226,(2013)P272-288.

[5] Severson, K. A., et al, “Data-driven prediction of battery cycle life before capacity degradation “

Nature.Energy,Vol4,(2019)P383-291.

[6] McMahan, H. B., et al, “Communication-Efficient Learning of Deep Networks from Centralized Data” *Proceedings of 20th Int Conf on AI and statistics (AISTATS) Florida USA*, Vol 54,(2017)P 1-10.

[7] Raissi, M, Perdikaris P, & Karniadakis G E, “ Physics-informed neural networks” *J. Compu. Phy*, Vol 378(2019)P686-707.

[8] Feng, X., et al, “Thermal runaway mechanism of lithium-ion battery for electric vehicles: A review” *Ener. Stor. Materi*, Vol 10(2018)P246-267.

[9] Kim G. H et al, “ Multi-domain modeling of lithium-ion batteries encompassing multi-physics in varied length scales “ *J. Electrochem. Soc*, Vol 158(8)(2011)A955-A969.

[10] Dong Xu, Ouyang, Mingyichen, Que Huang, Jingwen Weng, Zhiwang and Jianwang, “ A review on the thermal hazards of the lithium ion battery and the corresponding counter measures” *App.Sci*, Vol 9(12)(2019)2483-2527.

[11] Fan, Y., Wang, Z., Xiong, X., Panchal, S., Fraser, R.; Fowler, “ Multi-Objective Optimization Design and Experimental Investigation for a Prismatic Lithium-Ion Battery Integrated with a Multi-Stage Tesla Valve-Based Cold Plate” *Processes*, Vol 11(2023) P1618. [Google Scholar] [CrossRef]

[12] Zubi, G, Dufo-López, R, Carvalho, M, Pasaoglu, G, “The lithium-ion battery: State of the art and future perspectives” *Renew. Sustain. Energy Rev*, Vol 89 (2018) P292–308. [Google Scholar] [CrossRef]

[13]Ahmed, M.; Mao, Z.; Liu, Y.; Yu, A.; Fowler, M.; Chen, Z,"Comparative Analysis of Computational Times of Lithium-Ion Battery Management Solvers and Battery Models Under Different Programming Languages and Computing Architectures" Batteries Vol10,(2024) P439. [Google Scholar] [CrossRef]

[14]Tao, C, Li, G, Zhao, J, Chen G, Wang Z,Qian Y, Cheng X, Liu X,"The investigation of thermal runaway propagation of lithium-ion batteries under different vertical distances" J. Therm. Anal. Calorim,Vol142, (2020)P 1523–1532. [Google Scholar] [CrossRef]

[15]Sharma, H,Sharma S, Mishra P K, "A critical review of recent progress on lithium ion batteries: Challenges, applications, and future prospects." Microchem. J, Vol12(2025)113494. [Google Scholar] [CrossRef]

[16]KhalidR, Shah A, Javed M, Hussain H,"Progress and obstacles in electrode materials for lithium-ion batteries: A journey towards enhanced energy storage efficiency" RSC Adv,Vol15(2025)P 15951–15998. [Google Scholar] [CrossRef]

[17]Yao L, Yu C, Xiao Y,Cui G, Fei Z,Qu C A," comprehensive review of lithium-ion battery safety issues and fault diagnosis strategies throughout the entire lifecycle"J. Energy .Storage,Vol136(2025)P 118447. [Google Scholar] [CrossRef]

[18]Zanoletti A, Carena E, Ferrara C, Bontempi E A," Review of Lithium-Ion Battery Recycling: Technologies, Sustainability, and Open Issues, Batteries,Vol10 (2024),P 38. [Google Scholar] [CrossRef]

[19]Madani SS,Shabeer Y, Nair A.S,Fowler M, Panchal S, Ziebert C, Chaoui H.,Dou S.X, See K, Mekhilef S, et al," Advances in Battery Modeling and Management Systems: A Comprehensive Review of Techniques, Challenges, and Future Perspectives" Batteries ,Vol11(2025) 426. [Google Scholar] [CrossRef]

[20] Kai Long Liu,Yili,Yiasong Hu,Mattin Lucu, and Widanalage Dhammika Widahage," Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of Lithium ion batteries"IEEE.Trans.Ind.Inf,Vol 16(6)(2020) 3767-3777.