Open Validation of Arrhenius Thermal Runaway Calibration for Lithium-Ion Cells — A Benchmark Against a Public ARC Dataset

Kazuo Abe
Shinonome Engineering LLC
ORCID: 0009-0005-0557-0779
Technical Report · 2026-06-19 (last revised 2026-06-24)
Preprint submitted to engrXiv

Abstract

Accurate prediction of battery thermal runaway is a prerequisite for Hazard Mitigation Analysis (HMA) mandated by NFPA 855 [8] for most battery energy storage system (BESS) installations; UL 9540A [7] provides the standardized test method for thermal runaway fire propagation characterization. Arrhenius kinetic parameters (activation energy Ea, pre-exponential factor A, reaction enthalpy ΔH, and heat capacity Cp) required for CFD source-term generation (FLACS, KFX) are typically obtained by fitting an adiabatic thermal runaway ODE to Accelerating Rate Calorimetry (ARC) data. Despite its central role in fire safety engineering, no openly reproducible benchmark of this calibration procedure — with released code and raw results — exists, to our knowledge, for modern 21700 lithium-ion cells.

We present **Calor**, an open inverse-analysis workflow that fits a two-state adiabatic ODE ($y = [T, \alpha]$, $n = 1$) to public ARC time-series data using a three-stage optimizer (L-BFGS-B → Nelder-Mead → default-value fallback). Applied to five cell chemistries from the Zenodo 21700 ARC dataset (DOI: 10.5281/zenodo.7707929, CC BY 4.0), the workflow achieves range-normalized RMSE (NRMSE) ranging from 0.076 (NMC) to 0.304 (LFP) for the four chemistries with meaningful fits. One chemistry, NCA-HEI, produced a complete non-fit: the reaction progress variable α remains essentially zero throughout the ARC window ($\alpha_f \approx 0$), with simulated peak temperature 139 °C versus measured 484 °C (absolute RMSE ≈ 97 K); its low range-normalized NRMSE (0.257) reflects the large temperature-range denominator rather than fit quality. We characterize two identifiability limits: (i) an Arrhenius compensation effect, along which Ea and $\log_{10}(A)$ co-vary at near-constant NRMSE so that individual kinetic parameters are not uniquely resolved while the temperature trajectory is; and (ii) an extrapolation gap between the directly data-constrained quantity — the realized temperature rise $\alpha_f \cdot \Delta T_{ad}$, where $\alpha_f \in (0,1)$ is the reaction fraction reached in the ARC window — and $\Delta T_{ad} = \Delta H / C_p$ itself, which is an extrapolation to full reaction completion ($\alpha \to 1$); the ratio of ΔT_{ad} to the directly observed temperature rise — close to, but not exactly, $1/\alpha_f$, since the fitted trajectory does not perfectly reproduce the observed rise — is approximately 2.1× for NMC and 1.6× for NCA-HEII. Both limits manifest in the collapse of per-parameter bootstrap confidence intervals to near-zero width. We provide the calibration code and benchmark results as a demonstration of an openly reproducible calibration workflow.

Keywords: thermal runaway; Arrhenius kinetics; accelerating rate calorimetry; lithium-ion battery; BESS safety; source term generation

---

1. Introduction

Battery Energy Storage Systems (BESS) have expanded rapidly in grid-scale installations, and with them the need for credible fire safety analysis. NFPA 855 [8] requires project-specific Hazard Mitigation Analysis (HMA) for most BESS installations; the 2026 edition (published 2025-09-09) extends this requirement to nearly all stationary energy storage system projects. UL 9540A [7] provides the standardized test method for thermal runaway fire propagation characterization that informs HMA inputs; HMA is mandated by NFPA 855, not by UL 9540A. HMA typically relies on CFD simulation of thermal runaway propagation using FLACS or KFX. The critical inputs to these simulations — mass flux, heat release rate, and flammable gas composition — must be derived from cell-level calorimetry data.

The standard thermochemical model for adiabatic thermal runaway is the Arrhenius reaction rate law. The multi-step kinetic formulations of Hatchard et al. [1] and Kim et al. [2] resolve several concurrent decomposition reactions (SEI, anode–electrolyte, and cathode–electrolyte); the single-step first-order ($n = 1$) form adopted in this work is an engineering reduction of those models, preferred in practice for its tractability and because per-cell calibration data typically contain insufficient dynamic range to resolve multi-step parameters. The model requires four parameters per cell: Ea, A, ΔH, and Cp.

Despite decades of application, to our knowledge no openly reproducible benchmark of Arrhenius calibration — with released code and raw results — against modern 21700 cells exists in the public literature. Published parameters scatter across three orders of magnitude in A and 50–100 kJ/mol in Ea, in part because calibration procedures differ in optimizer choice, normalization convention, and handling of the ARC heat-wait-seek (HWS) protocol gaps. Practitioners must either trust vendor-supplied parameters (with opaque provenance) or spend significant effort re-implementing calibration from scratch.

This paper addresses the gap with three contributions:

1. **A reproducible calibration workflow** (Calor) implementing the adiabatic two-state ODE with three-stage optimization and an explicit treatment of bootstrap confidence-interval collapse under the compensation effect.
2. **A public benchmark** covering five chemistries of 21700 lithium-ion cells using a CC BY 4.0 ARC dataset.
3. **An honest characterization** of the Arrhenius compensation effect, which limits individual parameter identifiability but does not prevent accurate source-term generation.

---

2. Methods

2.1 Adiabatic Thermal Runaway Model

Under HWS adiabatic conditions, cell self-heating follows a two-state ordinary differential equation (ODE):

$$\frac{d\alpha}{dt} = A \cdot (1 - \alpha) \cdot \exp\left(\frac{-E_a}{R \cdot T_K}\right) \quad (1)$$

$$\frac{dT}{dt} = \left(\frac{\Delta H}{\phi \cdot C_p}\right) \cdot \frac{d\alpha}{dt} \quad (2)$$

where $\alpha \in [0, 1]$ is the normalized reaction progress variable ($\alpha = 0$: unreacted; $\alpha = 1$: fully reacted), TK is absolute temperature (K), $R = 8.314$ J mol^{-1} K^{-1}, and φ is the thermal inertia correction factor ($\varphi \geq 1$). Initial conditions are $\alpha(t_0) = 0$ and $T(t_0)$ equal to the measured temperature of the first ARC row at or above Tonset (which equals Tonset only when a sample sits exactly on the threshold).

The reaction-progress term $(1 - \alpha)$ is essential: without it, the ODE reduces to a pure Arrhenius rate that diverges monotonically, whereas ARC experiments exhibit a finite maximum temperature [3]. This single-step first-order model ($n = 1$) is adopted for v1 of the workflow; variable n introduces additional parameter correlation and is deferred to future work.

Numerical integration. The ODE is integrated with a 4th-order Runge-Kutta (RK4) method using 40 substeps per measurement interval. The 40-substep discretization was verified to be sufficient across the full range of step sizes encountered in the Zenodo dataset: measurement intervals span 0.003–29.15 min (the majority — 66% across the five cells — under 0.5 min, with a long tail toward the 30 min HWS gap threshold), giving substep sizes of 7.5×10^{-5}–0.73 min; integration error at this resolution is negligible relative to the NRMSE floor across the full range. Within each derivative evaluation a guard clamps the reaction-progress variable α to [0, 1] before forming the $(1 - \alpha)$ factor, so that temporary overshoot during stiff intervals cannot drive the rate negative or unbounded; the integrated state α is additionally hard-clipped to [0, 1] once per measurement interval. A temperature guard returns zero derivatives for state temperatures outside [200, 6000] K; this guard was not triggered for any benchmark cell. Kelvin conversion (TK = T°C + 273.15) is applied within the ODE; all input data are in °C.

2.2 Heat-Wait-Seek Gap Handling

ARC data collected under the HWS protocol contains forced heating segments between adiabatic exothermic phases. These appear as intervals with $\Delta t > 30$ min between consecutive measurements. During gap intervals, the ODE integration is paused, T is reset to the measured post-gap value, and α is carried forward unchanged (the reaction has continued during the gap, but the calorimeter has externally heated the cell). Gap-reset points are excluded from the NRMSE objective function, as they carry zero residual by construction and would artificially deflate the reported error. The ODE initialization point (index 0, where T_sim[0] = T_onset by construction) carries an identically zero residual and therefore contributes nothing to the optimization gradient; it is retained in the optimization objective but excluded from the range denominator when computing the reported NRMSE (see §2.3).

2.3 Inverse Analysis: Parameter Estimation

Optimization variables. The four parameters $\theta = \{E\text{a}, \log_{10}(A), \Delta H, C\text{p}\}$ are estimated by minimizing the range-normalized RMSE:

$$\text{NRMSE} = \sqrt{\left(\text{mean}\left((T_{\text{meas}} - T_{\text{sim}})^2\right)\right)} \,/\, (T_{\max} - T_{\min}) \qquad (3)$$

Range normalization is used in preference to mean normalization because temperature is an origin-arbitrary quantity; mean normalization inflates the denominator for absolute temperature in K, masking poor fits. During optimization, Tmax and Tmin are computed over all non-gap-reset points including the ODE initialization point (whose residual is identically zero by construction, so it contributes nothing to the gradient). For the NRMSE values reported in Table 1, the ODE initialization point is additionally excluded from the Tmin calculation in the range denominator: since the initialization point is set at the onset temperature Tonset ≈ Tmin of the exotherm, including it in the denominator artificially widens the range and deflates the reported NRMSE. The reported NRMSE is computed from the fitted Tsim using the optimal θ^* with the initialization point excluded from both the mean and the range denominator; θ^* itself is unchanged by this post-optimization correction. HWS gap-reset points are excluded from both the optimization objective and the reported NRMSE. A parameter is considered data-constrained when it deviates ≥5% from its starting value; the deviation threshold is approximate and intended as a practical heuristic.

Plausible pre-exponential factors span many orders of magnitude across reported single-step and component reactions, so we bound the search to $\log_{10}(A) \in [6, 20]$; optimizing in linear A-space over this range produces severe ill-conditioning, so we optimize $\log_{10}(A)$ and recover $A = 10\text{^}(\log_{10}(A))$ at reporting time.

Physical bounds. Parameters are constrained within physically meaningful ranges established from the LIB thermal runaway literature:

Parameter	Lower bound	Upper bound
E_a [J/mol]	50,000	200,000
$\log_{10}(A)$	6	20
ΔH [J/kg]	100,000	2,000,000
C_p [J/(kg·K)]	800	1,500

Three-stage optimizer. Stage 1 uses L-BFGS-B with 500 iterations (ftol = 10^{-10}, gtol = 10^{-8}) from a common fixed starting point θ_0 = (E_a = 120 kJ/mol, $\log_{10}A$ = 14, ΔH = 500 kJ/kg, C_p = 1000 J/(kg·K)). Stage 2 refines with Nelder-Mead (adaptive = True, maxiter = 3000, xatol = 10^{-6}, fatol = 10^{-9}) initialized from the Stage 1 result; a Stage 2 solution that violates the box constraints is rejected and the Stage 1 result retained. Stage 3 is a default-value fallback: the accepted estimate is whichever of Stage 1 or Stage 2 returns the lower finite objective, and only if both optimizers fail to return a finite objective (e.g., numerical divergence) does the fit revert to θ_0 (flagged `method = "fallback"`). The starting-point objective $f(\theta_0)$ is not evaluated as an acceptance threshold. This fallback did not trigger for any of the five benchmark cells.

Uncertainty quantification. We attempt LM re-fit followed by `lmfit.conf_interval()` for 95% CIs. For all five cells examined, the Arrhenius compensation effect (described in §3.3) produces a flat likelihood landscape, causing the sign-change condition required by `conf_interval()` to fail. We fall back to a non-parametric residual bootstrap (n = 200 resamples of the non-gap residuals; random seed 42 for reproducibility). However, because the optimizer is initialized at the optimal point θ, *all bootstrap samples converge back to (or extremely near) θ* — yielding near-zero CI widths for all parameters (e.g., Ea CI width < 10^{-3} J/mol for NMC; exactly zero for NCA-HEI). These CIs do not represent true parameter uncertainty; they reflect the optimizer's convergence behavior under compensation, not the actual breadth of the compensation ridge. Individual per-parameter CIs are therefore not reported in Table 1. The collapsed bootstrap CI values are retained in the released `w4_benchmark.json` only as evidence of this collapse and are explicitly flagged there as optimization-geometry artifacts (top-level `ci_disclaimer` field); they must not be consumed as parameter uncertainty bounds. The observable $\Delta T_{ad} = \Delta H / C_p$ is discussed in §3.3. For the two cells (NMC, NCA-HEII) whose fitted ΔT_{ad} deviates ≥5% from the default value (500 K), the quantity directly constrained by the ARC data is the realized temperature rise $\alpha_f \cdot \Delta T_{ad}$ (where $\alpha_f \in [0,1]$ is the reaction fraction at the end of the ARC measurement window); ΔT_{ad} itself — the rise to full reaction completion ($\alpha \to 1$) — is an extrapolation beyond the observed data. For NCA-HEII specifically, C_p shifted +8.4% from the starting value while ΔH shifted only −1.5%, so the data primarily constrained the ratio $\Delta T_{ad} = \Delta H / C_p$ (455 K, −9% from default), not the individual parameters. For the other three cells (LFP, NCA-HEI, NCA-HP) ΔH and/or C_p remain within ≤1.5% of their starting-point or bound values, so only the temperature trajectory and onset timing are data-constrained.

2.4 Dataset

ARC data were obtained from the open dataset published by Ohneseit et al. at Zenodo (DOI: 10.5281/zenodo.7707929, CC BY 4.0) [4]. The dataset contains HWS ARC experiments on 21700 lithium-ion cells at multiple states of charge. We use the SOC = 100% M1 trial for each chemistry, as this represents the maximum thermal runaway hazard scenario relevant to HMA.

Five chemistries were analyzed:

#	File	Chemistry	T_{onset} [°C]	n pts
1	NMC_SOC100_M1.txt	NMC	90.0	105
2	LFP_SOC100_M1.txt	LFP	120.0	228

#	File	Chemistry	*T*onset [°C]	n pts
3	NCA_HEI_SOC100_M1.txt	NCA-HEI	95.0	99
4	NCA_HEII_SOC100_M1.txt	NCA-HEII	90.0	101
5	NCA_HP_SOC100_M1.txt	NCA-HP	100.0	102

n pts is the number of ARC data rows (header excluded) in each SOC = 100% M1 file.

*T*onset values are nominal self-heating-onset thresholds set at or near the first recorded temperature of each cell's final HWS ramp (true first-point temperatures: 90.4 / 120.6 / 95.0 / 90.5 / 100.5 °C). The fit uses all rows with $T \geq T$onset and initializes the ODE at the first such row. For NCA-HEI the first recorded sample (94.991 °C) falls just below the 95.0 °C onset, so 98 of the 99 file rows enter the fit and the ODE initializes at the first row at or above onset (≈105.1 °C); for the other four cells *T*onset lies at or below the first sample, so all file rows are used.

All computations were performed with Python 3.11.15, scipy 1.17.1, lmfit 1.3.4, and numpy 2.4.4 on an Apple M4 Pro.

---

3. Results

3.1 Arrhenius Parameter Estimates

Table 1 presents the estimated parameters and corrected NRMSE for each cell. Four of the five cells achieve NRMSE ≤ 0.304; NMC produces the best fit (NRMSE = 0.076). NCA-HEI is a complete non-fit (§3.2); its reported NRMSE (0.257) reflects the large temperature-range denominator rather than fit quality. Figure 1 shows measured and simulated temperature profiles for all five chemistries.

Table 1. Arrhenius parameters estimated by inverse analysis of ARC data (Zenodo DOI: 10.5281/zenodo.7707929).

Chemistry	*E*a [kJ/mol]	*A* [s^{-1}]	Δ*H* [J/kg]	*C*p [J/(kg·K)]	α_f	Δ*T*ad [K]‡	NRMSE*	absRMSE [K]	Conv.§
NMC (21700)	**96.3**	4.55×10^{7}	617,500	903	0.51	684	0.076	25	✗
LFP (21700)	120.0¶	6.00×10^{8}	500,000 †	1,000 †	0.42	500 †	0.304	81	✗
NCA-HEI (21700)	121.1	1.02×10^{6}	500,000 †	1,500 † †	0.00 ‡‡	333 †	0.257 ‡‡	97 ‡‡	✓
NCA-HEII (21700)	120.5	2.99×10^{10}	492,700	1,084	0.61	455	0.110	29	✗
NCA-HP (21700)	120.0¶	2.25×10^{10}	500,000 †	1,000 †	0.55	500 †	0.093	26	✓

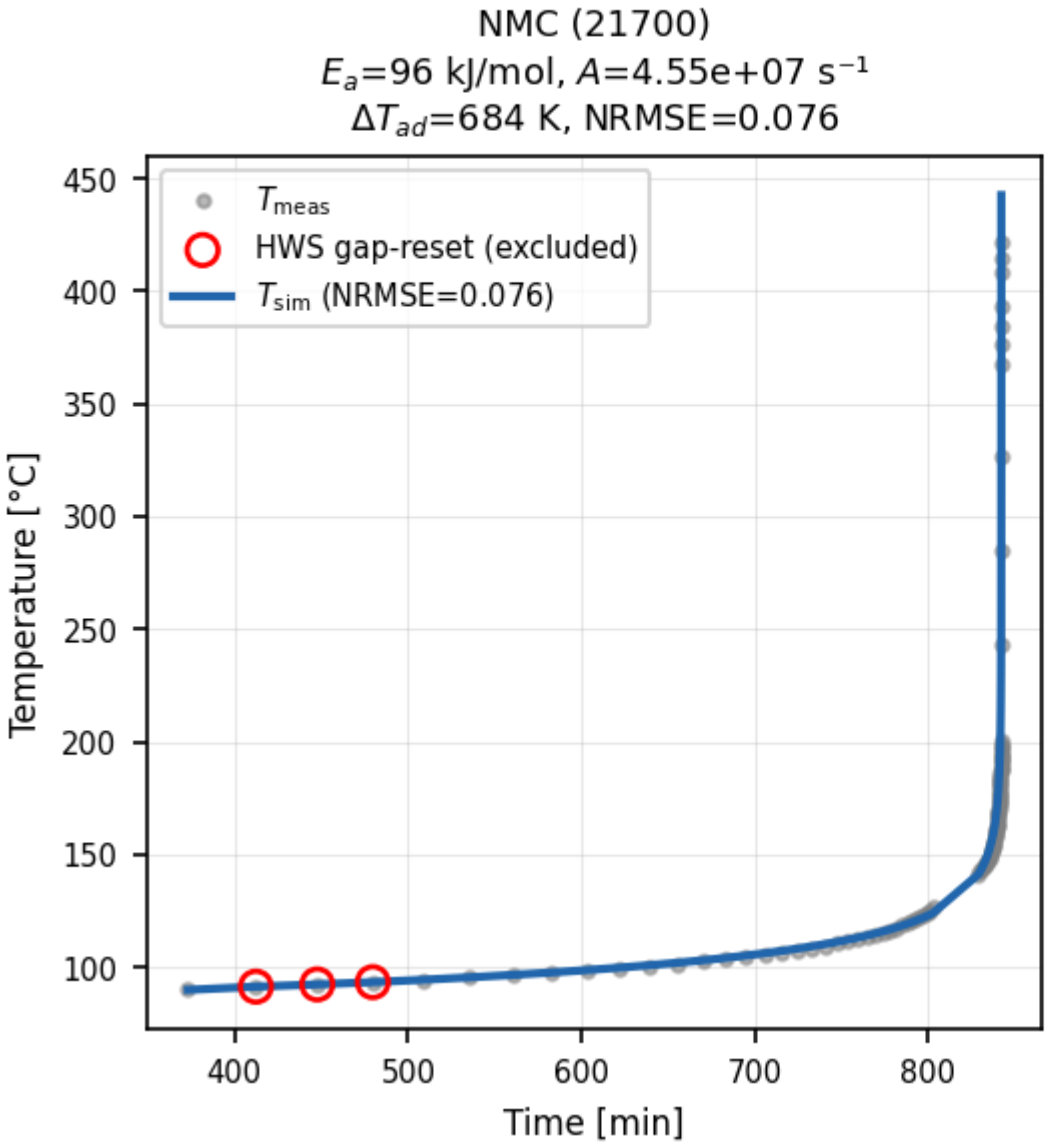
* NRMSE computed with HWS gap-reset points excluded from both the optimization objective and this reported metric (§2.2); the ODE initialization point remains in the optimization objective (its residual is identically zero, so it does not affect θ*) but is additionally excluded, post-hoc, from this reported NRMSE (§2.3). Raw fitted values and the objective-consistent NRMSE (initialization point included) are in `data/processed/w4_benchmark.json`.
† Δ*H* and/or *C*p deviating ≤1.5% from the optimizer starting value (≈500,000 J/kg, 1,000 J/

(kg·K)); these parameters are not data-constrained, so the corresponding Δ*T*ad (also marked †) is a literature-default placeholder, not data-derived.
† † *C*p at its upper physical bound (1,500 J/(kg·K)); likewise not data-constrained.
‡ Δ*T*ad = Δ*H* / *C*p [K]: adiabatic temperature rise to full reaction completion (α→1). The data-constrained quantity is the realized rise α_f·Δ*T*ad. For NMC: 0.51 × 684 ≈ 349 K vs. observed (measured *T*max − *T*min) 331 K (Δ*T*ad/obs ≈ 2.1×). For NCA-HEII: 0.61 × 455 ≈ 278 K vs. observed 289 K (≈1.6×). The 18 K (NMC) and 11 K (NCA-HEII) gaps between the model-implied realized rise (α_f·Δ*T*ad) and the directly observed temperature range are consistent with — and of the same order as — the reported absRMSE (25 K and 29 K respectively; Table 1), i.e. they reflect ordinary fit residual rather than a separate error source. Δ*T*ad values for NMC and NCA-HEII must be understood as extrapolations to α→1, not directly observed quantities. Per-parameter bootstrap 95% CIs collapse to near-zero for all cells and are not reported (§2.3, §3.3).
‡‡ NCA-HEI is a complete non-fit: α_f ≈ 0, simulated peak ≈139 °C vs. measured 484 °C, absRMSE ≈ 97 K. The range-normalized NRMSE (0.257) is misleading due to the large temperature-range denominator (≈379 K). **The NCA-HEI parameters in this table must not be used for CFD source-term generation.**
§ Optimizer convergence flag: ✓ = strict stopping tolerance met (NCA-HEI, NCA-HP, both L-BFGS-B); ✗ = terminated on the flat compensation ridge without meeting strict tolerance (NMC, NCA-HEII via Nelder-Mead; LFP via L-BFGS-B). Fit quality should be judged by NRMSE and absRMSE, not this flag (§3.3).
¶ *E*a coincides with the 120 kJ/mol starting value to within optimizer tolerance (>5 significant figures) and is not independently resolved; read as "consistent with ≈120 kJ/mol." NMC (96.3 kJ/mol) is the one cell with a strongly data-driven *E*a.

The NMC result stands out with a notably lower *E*a (96.3 kJ/mol) compared to the NCA and LFP chemistries (≈120 kJ/mol); it is the one strongly data-driven *E*a in the set (¶), whereas the NCA and LFP cells sit at or near the 120 kJ/mol start value. We attribute no specific mechanism given the single-trial (*n* = 1) scope — indeed NMC and NCA-HEII share the same nominal onset (90.0 °C) despite their differing *E*a, so onset does not track *E*a across this dataset. The NCA-HP chemistry achieves NRMSE = 0.093, comparable to NCA-HEII (0.110), despite being a high-power cell with different electrode thickness characteristics.

3.2 Fit Quality

Fig. 1 — ARC Calibra
for Five 21700 Cell Chemistrie



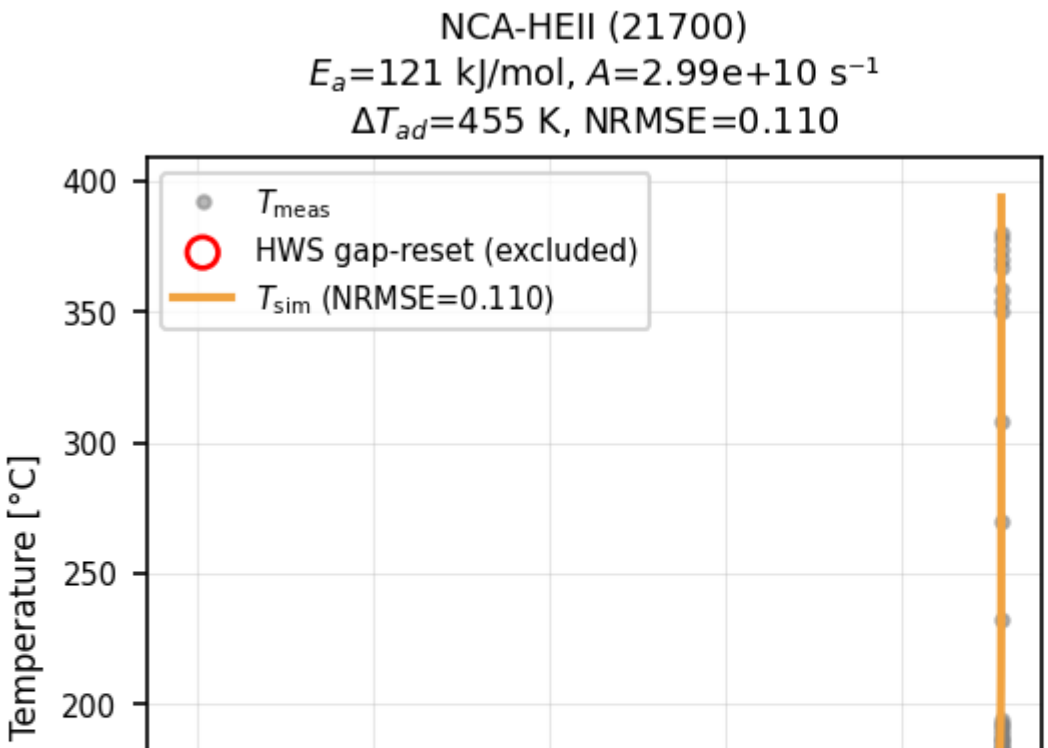


Figure 1. Measured (*T*meas, gray circles) and simulated (*T*sim, colored line) temperature profiles for five 21700 cell chemistries. Open red circles mark HWS gap-reset points, which are excluded from the NRMSE objective. Each panel title reports the fitted *E*a, *A*, Δ*T*ad and corrected NRMSE for that chemistry (cf. Table 1; Δ*T*ad values marked † in Table 1 are not data-constrained; the NCA-HEI panel shows a complete non-fit — see §3.2). ARC data from Ohneseit et al. [4].

Fit quality varies substantially across chemistries and must be assessed by both NRMSE and absolute RMSE (absRMSE); the range-normalized NRMSE alone can mislead when the temperature range varies across cells.

NMC (NRMSE = 0.076, absRMSE = 25 K) and **NCA-HP** (NRMSE = 0.093, absRMSE = 26 K) achieve the best trajectory reproduction; **NCA-HEII** (NRMSE = 0.110, absRMSE = 29 K) also fits well with α_f = 0.61. For these three cells, the single-step adiabatic model captures the dominant thermal runaway dynamics and the simulated trajectory is a credible starting point for PE/FPE review.

LFP (NRMSE = 0.304, absRMSE = 81 K, α_f = 0.42) shows significantly higher absolute error, consistent with the multi-step reaction behavior of LFP chemistry — staged anode-SEI and electrolyte-decomposition reactions rather than a single dominant cathode exotherm [6] — that the $n = 1$ model partially misrepresents. The simulated trajectory captures the runaway onset timing, but the absolute temperature error is substantial. Quantitative engineering-adequacy assessment for LFP is deferred to the multi-SOC/replicate v2 benchmark (§4.2).

NCA-HEI is a complete non-fit. Despite the optimizer reporting convergence (L-BFGS-B, ✓), the reaction progress variable α remains essentially zero throughout the ARC window (α_f ≈ 0.00), with a simulated peak temperature of approximately 139 °C versus the measured 484 °C peak — an absolute RMSE of 97 K. The range-normalized NRMSE = 0.257 is misleading: the large temperature range denominator (≈379 K) compresses the large absolute error into a deceptively moderate normalized value. The Calor three-stage optimizer failed to find any set of Arrhenius parameters consistent with the NCA-HEI ARC trajectory under the $n = 1$ single-step model; this chemistry likely requires a multi-step formulation, or the ARC data exhibits characteristics (e.g., multi-phase exotherm, measurement artefacts) outside the model's applicability domain. The NCA-HEI parameters in Table 1 must not be used for CFD source-term generation.

For the well-fit cells (NMC, NCA-HEII, NCA-HP), the key source-term outputs are the simulated temperature trajectory and the adiabatic rise α_f·Δ*T*ad (the data-constrained realized rise); Δ*T*ad itself requires extrapolation beyond the observed α_f and should be independently validated before use as a FLACS/KFX input. No published NRMSE threshold for source-term engineering adequacy is asserted here.

3.3 Arrhenius Compensation Effect

Multiple cells show Δ*H* and *C*p fixed at or near the optimizer initial values (500,000 J/kg and 1,000 J/(kg·K) respectively), or at a physical bound (NCA-HEI *C*p sits at its 1,500 J/(kg·K) upper bound), indicated by † / †† in Table 1. This is not a numerical failure but an expression of the **Arrhenius compensation effect**: along a compensation ridge, *E*a and $\log_{10}(A)$ co-vary with near-constant NRMSE, so the temperature trajectory can be reproduced without uniquely fixing them. Δ*H* and *C*p enter only through their ratio Δ*T*ad; where the fit does not fully resolve the temperature excursion, the optimizer leaves them at their default values (three of the five cells here, marked †). For the two cells with Δ*H*/*C*p deviating ≥5% from defaults (NMC, NCA-HEII), the quantity data-constrained by the ARC window is the realized rise α_f·Δ*T*ad (NMC: 0.51 × 684 ≈ 349 K vs. observed 331 K; NCA-HEII: 0.61 × 455 ≈ 278 K vs. observed 289 K), not Δ*T*ad itself; Δ*T*ad is an extrapolation to α→1 that exceeds the observed rise by approximately 2.1× (NMC) and 1.6× (NCA-HEII).

Concretely, for a fixed Δ*T*ad: - Increasing *E*a shifts the runaway onset to higher temperature - Increasing $\log_{10}(A)$ shortens the onset time - Within the experimental temperature window, pairs

(Ea, $\log_{10}(A)$) satisfying $\log_{10}(A) \approx E\text{a} / (2.303\ R\ T\text{c}) + \text{const.}$ — where Tc is the isokinetic (compensation) temperature of the ARC window, approximately the midpoint of the self-heating ramp — produce similar T(t) trajectories

This is a manifestation of the compensation effect, a well-documented identifiability limitation in thermal-decomposition kinetics [5]; here it is intrinsic to single-temperature-range ARC data, not a defect of the optimizer. For source-term generation, the compensation effect is benign: the temperature trajectory (and derived source terms) is well-constrained, even if individual Ea and A values are not uniquely identified.

CI collapse. In this benchmark, bootstrap CIs (n = 200) produce near-zero widths for all parameter estimates ($< 10^{-3}$ J/mol for Ea; exactly zero for NCA-HEI across all 200 samples). This is a direct consequence of single-start bootstrap: each resample is optimized from θ*, and the compensation ridge causes the optimizer to return to the same θ* regardless of the perturbation. The near-zero CIs are an artifact of the optimization geometry, not a credible uncertainty estimate. Quantifying the true breadth of the compensation ridge requires profile-likelihood mapping or multi-start exploration orthogonal to the ridge direction — this is deferred to a future benchmark release.

Practical implication for PE/FPE: For the cells whose $\Delta H/C$p deviate ≥5% from defaults (NMC and NCA-HEII in this benchmark), the realized rise $\alpha_f \cdot \Delta T$ad is directly constrained by the ARC data and provides a reproducible and transparent starting point for PE/FPE review. ΔTad itself — the extrapolation to full reaction completion — exceeds the observed rise by a factor close to, but not exactly, $1/\alpha_f$ (the two coincide only if the fitted trajectory exactly reproduces the observed rise; see §3.3, Table 1 footnote ‡) and must be independently validated against direct calorimetric measurements before use as a FLACS/KFX source-term input. For the cells whose ΔH and Cp remain at defaults (LFP, NCA-HEI, NCA-HP), ΔTad reflects a prior assumption rather than a data-constrained fit, and a measured or literature-sourced ΔTad must be substituted before use in source-term generation. NCA-HEI is a complete non-fit and its parameters must not be used. In all cases the simulated temperature trajectory is the well-constrained quantity; individual Ea and A values should be used only with explicit acknowledgment of the compensation effect. The engineer of record bears responsibility for validating these parameters in the context of their specific HMA.

---

4. Discussion

4.1 Comparison with Published Values

Published single-step Arrhenius parameters for 21700-class NMC cells are sparse; most literature targets 18650 cells or model compounds, and reported values span a wide range owing to differences in cell construction, state of charge, and calorimetry. The single-step whole-cell effective Ea recovered here (96–121 kJ/mol) is an aggregate that lumps SEI, anode, and cathode-electrolyte reactions, and is therefore not directly comparable to the individual component-reaction activation energies tabulated in multi-step abuse models such as those of Hatchard et al. [1] and Kim et al. [2], whose SEI-decomposition and positive-solvent terms differ from one another by tens of kJ/mol. Within this caveat, our NMC value (96.3 kJ/mol, 4.55×10^{7} s^{-1}) lies in the broad range of single-step NMC-class kinetics aggregated in review compilations [6].

Direct cell-for-cell comparison is precluded by differences in cell generation, SOC, C-rate preconditioning, and ARC apparatus, all of which affect the measured thermal runaway kinetics. The primary value of this benchmark is not absolute parameter values but **reproducible methodology and uncertainty quantification under transparent data provenance.**

4.2 Limitations

Single-step model. The adiabatic $n = 1$ model conflates multiple concurrent reactions (SEI decomposition, cathode-electrolyte reaction, lithium plating) into a single Arrhenius term. For FLACS/KFX source-term generation this approximation is typically accepted by AHJs; for detailed thermal propagation modeling, multi-step models (e.g., Kim et al. [2]) may be warranted.

Single SOC, single trial. This benchmark uses SOC = 100% M1 only. The Zenodo dataset contains data at SOC = 0, 30, 60, 80, and 100%, and two to four replicates per condition. Extending the benchmark to multi-SOC parameter surfaces and replicate consistency checks is planned for v2.

18650 dataset. The companion Zenodo dataset (DOI: 10.5281/zenodo.14956641) contains NMC811, LFP, and Na-ion 18650 cells in .EXO proprietary format. An .EXO parser is under development and will enable direct comparison of 18650 vs. 21700 parameters in a future release.

φ factor. All fits used φ = 1.0 (perfect adiabatic conditions). In a real ARC run φ is slightly above unity, so the measured adiabatic temperature rise ΔTad is suppressed by a factor 1/φ; fitting such data with the model fixed at φ = 1 forces the optimizer to absorb that suppression into the fitted ΔH, biasing the inferred ΔT_{ad} downward. PE/FPE practitioners should supply their facility's measured φ via the arc_conditions field in the calor-input v1 JSON schema so that the fit recovers the φ-corrected ΔT_{ad}. This correction is meaningful only for cells in which ΔH and C_p are independently resolved by the data (NMC, NCA-HEII here); where ΔH and C_p remain pinned at defaults, correcting φ does not change the fitted ΔT_{ad}.

---

5. Conclusion

We have demonstrated an openly reproducible Arrhenius calibration workflow against public ARC datasets for five 21700 cell chemistries. For the four chemistries with meaningful fits, NRMSE ranges from 0.076 (NMC) to 0.304 (LFP) and absolute RMSE from 25 K (NMC) to 81 K (LFP). One chemistry, NCA-HEI, is a complete non-fit under the $n = 1$ single-step model ($\alpha_f \approx 0$; absolute RMSE ≈ 97 K; simulated peak 139 °C vs. measured 484 °C); its parameters must not be used for CFD source-term generation.

The workflow documents two identifiability limits that practitioners must account for before using calibrated parameters in FLACS or KFX:

1. **Arrhenius compensation effect:** E_a and $\log_{10}(A)$ are not uniquely resolved from single-temperature-range ARC data; the temperature trajectory is well-constrained, but individual kinetic parameters are not.

2. **Extrapolation gap:** The quantity constrained by ARC data is the realized temperature rise $\alpha_f \cdot \Delta T_{ad}$ (where $\alpha_f \in (0,1)$ is the reaction fraction in the ARC window), not $\Delta T_{ad} = \Delta H / C_p$ itself. For NMC ($\alpha_f \approx 0.51$) and NCA-HEII ($\alpha_f \approx 0.61$), ΔT_{ad} exceeds the directly observed rise by approximately 2.1× and 1.6×, respectively. Independent validation of ΔT_{ad} against direct calorimetric measurements is recommended before use as a FLACS/KFX source-term input. For LFP and NCA-HP, ΔH and C_p remain at optimizer default values; a measured or literature-sourced ΔT_{ad} must be substituted.

The calibrated parameters provide a reproducible and transparent starting point for PE/FPE review; responsibility for validating their use in specific HMA contexts rests with the engineer of record.

The complete calibration code, input files, and results are publicly available on GitHub (https://github.com/kuro-tomo/calor-public) under the MIT License. ARC data are cited per the CC BY 4.0 terms of the source Zenodo dataset.

Data and Code Availability

- **ARC raw data:** Zenodo DOI: 10.5281/zenodo.7707929 (Ohneseit et al., CC BY 4.0)
- **Calibration code (Calor parser + benchmark):** https://github.com/kuro-tomo/calor-public (MIT License)
- **Results (w4_benchmark.json):** Zenodo DOI: 10.5281/zenodo.22104009
- **Reproduction steps:** `python -m tools.bench_w4` from project root, Python 3.11+, requirements in `requirements.txt`

Acknowledgments

ARC data provided open-access under CC BY 4.0 by Ohneseit et al. via Zenodo (DOI: 10.5281/zenodo.7707929). The calibration workflow reuses numerical infrastructure developed in-house at Shinonome Engineering LLC.

This work was developed with the assistance of an AI coding and writing assistant (Claude, Anthropic), used for code implementation, iterative technical review, reference verification, and drafting/copy-editing of this manuscript under human supervision. All scientific claims, benchmark results, and interpretations were verified and approved by the author.

Competing Interests

The author (K. Abe) is affiliated with Shinonome Engineering LLC, which develops Calor as a commercial calibration component for PE/FPE practitioners. This commercial interest did not influence the benchmark methodology or the reported results, which are reproducible from the cited open ARC dataset.

Funding

This work received no external funding; it was conducted internally by Shinonome Engineering LLC.

References

[1] Hatchard, T. D.; MacNeil, D. D.; Basu, A.; Dahn, J. R. Thermal Model of Cylindrical and Prismatic Lithium-Ion Cells. *J. Electrochem. Soc.* **2001**, 148, A755–A761. DOI: 10.1149/1.1377592.

[2] Kim, G.-H.; Pesaran, A.; Spotnitz, R. A Three-Dimensional Thermal Abuse Model for Lithium-Ion Cells. *J. Power Sources* **2007**, 170, 476–489. DOI: 10.1016/j.jpowsour.2007.04.018.

[3] Richard, M. N.; Dahn, J. R. Accelerating Rate Calorimetry Study on the Thermal Stability of Lithium Intercalated Graphite in Electrolyte. I. Experimental. *J. Electrochem. Soc.* **1999**, 146, 2068–2077. DOI: 10.1149/1.1391893.

[4] Ohneseit, S.; Finster, P.; Floras, C.; Lubenau, N.; Uhlmann, N.; Seifert, H. J.; Ziebert, C. *Exothermal data from thermal safety assessment of type 21700 lithium-ion batteries with NMC,*

NCA and LFP cathodes by means of Accelerating Rate Calorimetry (ARC). Karlsruhe Institute of Technology (KIT) / Queen's University, Zenodo, 2023. DOI: 10.5281/zenodo.7707929.

[5] Vyazovkin, S. *Isoconversional Kinetics of Thermally Stimulated Processes*; Springer, 2015. DOI: 10.1007/978-3-319-14175-6.

[6] Feng, X.; Ouyang, M.; Liu, X.; Lu, L.; Xia, Y.; He, X. Thermal Runaway Mechanism of Lithium Ion Battery for Electric Vehicles: A Review. *Energy Storage Mater.* **2018**, 10, 246–267.

[7] UL 9540A:2026. *Test Method for Evaluating Thermal Runaway Fire Propagation in Battery Energy Storage Systems*, 6th ed.; UL Standards & Engagement: Northbrook, IL, **2026** (published 2026-03-13).

[8] NFPA 855. *Standard for the Installation of Stationary Energy Storage Systems*, 2026 ed.; National Fire Protection Association: Quincy, MA, **2025** (published 2025-09-09).

Appendix A — calor-input v1 JSON Schema (excerpt)

```
{
  "schema_version": "calor-input-v1",
  "arc_data": [
    {"time_min": 0.0, "T_degC": 90.0, "dTdt_degC_min": null}
 ],
  "cell_spec": {
    "cell_id": "NMC_SOC100_M1",
    "chemistry": "NMC",
    "capacity_ah": 4.9,
    "mass_g": 68.0,
    "format": "21700"
 },
  "arc_conditions": {
    "T_onset_degC": 90.0,
    "phi": 1.0
 }
}
```

The `capacity_ah` and `mass_g` fields are nominal/representative 21700 values recorded as cell metadata; they are not inputs to the Arrhenius fit (which operates on the $T(t)$ time series, with ΔH and Cp in per-kilogram units) and therefore do not affect any Table 1 result.

Corresponding author: Kazuo Abe, Shinonome Engineering LLC — info@shinonome-giken.com · ORCID: 0009-0005-0557-0779
Preprint. Not peer-reviewed.