

# SpectraQuery: A Hybrid Retrieval-Augmented Conversational Assistant for Battery Science

Anonymous ACL submission

## Abstract

Scientific reasoning increasingly requires linking structured experimental data with the unstructured literature that explains it, yet most large language model (LLM) assistants cannot reason jointly across these modalities. We introduce **SpectraQuery**, a hybrid natural-language query framework that integrates a relational Raman spectroscopy database with a vector-indexed scientific literature corpus using a Structured and Unstructured Query Language (SUQL)-inspired design. By combining semantic parsing with retrieval-augmented generation, SpectraQuery translates open-ended questions into coordinated SQL and literature retrieval operations, producing cited answers that unify numerical evidence with mechanistic explanation. Across SQL correctness, answer groundedness, retrieval effectiveness, and expert evaluation, SpectraQuery demonstrates strong performance: approximately **80%** of generated SQL queries are fully correct, synthesized answers reach **93–97%** groundedness with 10–15 retrieved passages, and battery scientists rate responses highly across accuracy, relevance, grounding, and clarity (**4.1–4.6/5**). These results show that hybrid retrieval architectures can meaningfully support scientific workflows by bridging data and discourse for high-volume experimental datasets.

## 1 Introduction

### 1.1 Background and Motivation

Large Language Models (LLMs) have revolutionized natural-language understanding and reasoning, driving rapid progress in information retrieval, dialogue systems, and autonomous research assistants (Zhao et al., 2024). However, advances in battery science increasingly depend on correlating large volumes of quantitative measurements with qualitative mechanistic insights buried in the literature. For example, operando spectroscopy, electrochemical impedance measurements, and op-

tical microscopy generate rich structured datasets describing evolving electrode materials and cell performance (Xue et al., 2024). 3D operando Raman spectroscopy in particular is an emerging and promising technique which provides real-time tracking of structural and chemical transformations during charge-discharge cycles (Zhu et al., 2018). A visualization of what Raman spectroscopy captures for battery materials is demonstrated in Figure 1. Characteristic peaks such as the A1g mode of transition-metal-oxide vibrations and the D/G carbon bands reflect changes in redox state and structural disorder (Ferrari and Robertson, 2000). Each experiment yields tens of thousands of spectra across spatial and temporal coordinates, and recent initiatives in battery science have also begun establishing open and FAIR Raman databases to increase access to this valuable data (Coca-Lopez et al., 2025).

Unfortunately, though, drawing chemical meaning from these spectroscopic features typically demands cross-referencing dozens of prior literature that describe oxygen redox reactions, side product formation, or crystalline disordering, phase changes, and more (Heber and Hess, 2021). The result is an intensive manual workflow that separates numerical observation from conceptual interpretation. Current LLM-based tools rarely integrate both structured measurement data and unstructured text so that a researcher can pose a single query combining the two. As a result, the workflow remains divided: analysts query numerical datasets in isolation and separately review literature for context, a process that is time-consuming, error-prone, and may overlook cross-modal patterns.

### 1.2 Contributions of Work

SpectraQuery unifies structured experimental data and unstructured text by providing an end-to-end system that performs structured database reasoning, semantic text retrieval, and grounded synthesis in

response to natural language questions. Inspired by the SUQL framework of Liu et al. (2023) and built upon the retrieval-augmented generation paradigm of Lewis et al. (2020), the system demonstrates that hybrid querying and generative reasoning can be applied effectively to real scientific datasets (Liu et al., 2023; Lewis et al., 2020). By adapting hybrid query architectures originally developed for general-purpose knowledge bases, SpectraQuery extends structured-unstructured retrieval into a scientific setting, enabling automated yet interpretable generation of insights.

The key contributions of this work are: (1) a **scientific hybrid QA architecture** that couples executable spatiotemporal SQL programs over spectroscopy-derived peak parameters with literature retrieval and citation-grounded synthesis in a single interaction; (2) a **SUQL-style planner** that decomposes open-ended scientific questions into coordinated structured (SQL) and unstructured (vector search) operations, enabling multi-step phenomena such as cross-timestep comparisons and derived metrics (i.e., D/G) without requiring users to write queries; and (3) an **evaluation suite** for hybrid scientific assistants that combines SQL correctness, groundedness under varying retrieval depth, retrieval effectiveness (precision/recall/diversity), and expert ratings from domain scientists.

## 2 Related Work

### 2.1 Raman Spectroscopy in Battery Research

Raman spectroscopy has long been used to characterize battery materials. It can probe lattice structures, detect phase changes, and identify surface species via characteristic vibrational peaks (Baddour-Hadjean and Pereira-Ramos, 2010). For instance, the D ( $1350\text{ cm}^{-1}$ ) and G ( $1580\text{ cm}^{-1}$ ) bands of carbon report on relative disorder; the ratio of  $I(D)/I(G) > 1$  ( $I = \text{intensity}$ ) typically indicates significant disorder or defects in carbon (Ferrari and Robertson, 2001). In battery electrodes, an increasing  $I(D)/I(G)$  ratio may signal carbon black degradation or binder decomposition. Meanwhile, metal-oxide cathodes exhibit Raman-active modes such as A1g and Eg; changes and shifts in the A1g peak intensity or position can reflect lithium de-intercalation or oxygen lattice distortion (Flores et al., 2018). Operando Raman studies have observed phenomenon like spatially non-uniform lithiation, indicating heterogeneous battery material behavior and performance and hinting at under-

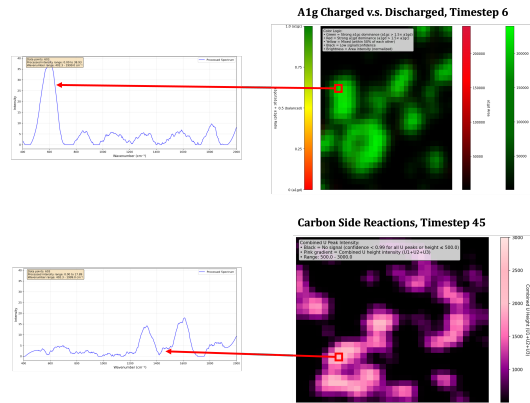


Figure 1: Each spatial pixel in the Raman maps (right panels) represents the intensity distribution of characteristic vibrational modes identified by automated peak detection. Representative spectra from individual pixels (left panels) show the corresponding local features—(top) A1g charged transition-metal–oxygen vibration and (bottom) combined unknown carbon peaks

lying drivers of degradation modes (Hiraoka et al., 2025). To date, most analysis of operando Raman is manual or relies on static plotting; SpectraQuery is, to our knowledge, the first system to enable interactive querying of operando Raman data integrated with literature insight.

### 2.2 Structured + Unstructured Data Querying

Traditional text-to-SQL research has tackled converting natural language questions into SQL on relational databases (Yu et al., 2018). Recently, hybrid Question Answering (QA) datasets like HybridQA explicitly require reasoning over both tabular data and text (Chen et al., 2020). Models that combine a knowledge base with a text corpus have shown improved coverage on such QA tasks. In particular, the Structured and Unstructured Query Language (SUQL) pioneered at Stanford, which achieved impressive performance on the HybridQA dataset, has introduced a paradigm for LLM in-context semantic parsers that can handle hybrid data (Liu et al., 2023). SpectraQuery follows this paradigm, treating the Raman spectra as a structured SQL database, and the battery literature as unstructured text.

While SpectraQuery uses text-to-SQL internally, our target setting differs from classical text-to-SQL benchmarks: many queries require derived scientific quantities (i.e., peak ratios), spatiotemporal comparisons across timesteps, and mechanistic interpretation that is not present in the database. A single-shot SQL query may retrieve the correct

164 numbers but cannot supply literature-grounded ex- 212  
165 planations, whereas literature QA alone cannot reli- 213  
166 ably compute joins, aggregations, and comparisons. 214  
167 SUQL provides an explicit planning abstraction to 215  
168 coordinate these operations and expose inspectable 216  
169 intermediate evidence. 217

## 170 2.3 Retrieval-Augmented Generation (RAG) 218

171 LLMs augmented with retrieval have reached 219  
172 state-of-the-art results in knowledge-intensive tasks 220  
173 (Lewis et al., 2020). By fetching relevant text 221  
174 from an external body, RAG models avoid hallu- 222  
175 cinations by providing verified information and 223  
176 citing sources. Prior RAG frameworks have fo- 224  
177 cused on open QA over extremely large external 225  
178 corpora, such as the ATLAS model with open QA 226  
179 over Wikipedia (Izacard et al., 2023). Spectra- 227  
180 Query instead builds a tailored literature index of 228  
181 domain-specific works; scientific QA systems such 229  
182 as SciQA and BioGPT have shown that incorporat- 230  
183 ing this type of tailored literature corpus can mean- 231  
184 ingfully enrich explanations (Auer et al., 2023; Luo 232  
185 et al., 2022). 233

## 186 2.4 LLM Planning and Execution 234

187 The idea of using an LLM as a controller to decide 235  
188 and execute actions has gained traction with ap- 236  
189 proaches like ReAct and OpenAI’s function-calling 237  
190 API, which allow the LLM to produce a program- 238  
191 matic plan as intermediate steps in reaching a final 239  
192 answer (Yao et al., 2022). SpectraQuery leverages 240  
193 this concept via a two stage process: in stage one, 241  
194 the SUQL planner (powered by an LLM prompt) 242  
195 generates a SQL query and a separate literature 243  
196 search query from the user’s question. In step two, 244  
197 after retrieving numerical data and relevant textual 245  
198 passages, the LLM produces the final answer with 246  
199 the secured evidence. 247

# 200 3 Core Ideas and Methodology 248

## 201 3.1 System Overview 249

202 SpectraQuery is a hybrid QA system that combines 250  
203 a relational database of Raman spectroscopy results 251  
204 with a vector-indexed corpus of literature. The user 252  
205 interacts via a chat-style interface, asking ques- 253  
206 tions in natural language. Internally, the system 254  
207 comprises several components: a query planner 255  
208 (SUQL) that interprets the question, a SQL query 256  
209 executor, a semantic literature search, and an LLM 257  
210 answer synthesizer. The design follows a plan- 258  
211 execute-synthesize paradigm, enabling complex 259

212 multi-hop reasoning in a single seamless interac- 213  
214 tion (Liu et al., 2023). SpectraQuery’s web inter- 215  
216 face, currently implemented with the Python toolkit 217  
218 Streamlit, allows the user to have a multi-turn con- 219  
220 versation. A full receipt of all queries, intermediate 220  
221 steps and information retrieval, and final synthe- 222  
223 sized answers is also easily downloadable in a PDF, 224  
225 allowing for accessible sharing of results. 226

## 220 3.2 Structured Database 220

### 221 3.2.1 Raman Data Preprocessing 221

222 The raw Raman dataset used for validation of Spec- 223  
224 traQuery comes from line-scan operando exper- 224  
225 iments on a lithium-ion layered transition metal 225  
226 oxide (LTMO) cathode sampled at 114 timesteps 226  
227 during charge and discharge cycling (experiment 227  
228 performed at SLAC National Accelerator Labora- 228  
229 tory). At each timestep, a linear spatial scan was 229  
230 performed, yielding a 30x30 two-dimensional grid 230  
231 of Raman spectra at each (x, y) coordinate (900 231  
232 spectra per timestep). Each spectrum spans 100- 232  
233 2700  $\text{cm}^{-1}$ , and contains many peaks of interest. 233

234 Raw spectra underwent a standard preprocess- 234  
235 ing pipeline of spike removal, smoothing with a 235  
236 Savitzky-Golay filter (window of 31 points), and 236  
237 Asymmetric Least Squares baseline subtraction. 237  
238 The resulting smooth, baseline-corrected spectrum 238  
239 was then passed to a Bayesian peak fitting algo- 239  
240 rithm, yielding a set of detected peaks with fitted 240  
241 parameters. 241

242 Each detected peak was assigned to one of eight 242  
243 canonical families by matching its fitted center 243  
244 to expected wavenumber ranges: TM–O lattice 244  
245 modes ( $E_g \approx 476 \text{ cm}^{-1}$ ,  $A_{1g,d} \approx 534.5 \text{ cm}^{-1}$ , 245  
246  $A_{1g,c} \approx 595.5 \text{ cm}^{-1}$ ), carbon bands ( $D \approx$  246  
247  $1330.5 \text{ cm}^{-1}$  indicating disordered/ $\text{sp}^3$ -rich car- 247  
248 bon and  $G \approx 1596.8 \text{ cm}^{-1}$  indicating graphitic/ $\text{sp}^2$  248  
249 carbon), and three additional carbon features 249  
250 ( $U_1 \approx 1173.3 \text{ cm}^{-1}$ ,  $U_2 \approx 1508.1 \text{ cm}^{-1}$ ,  $U_3 \approx$  250  
251  $1564.0 \text{ cm}^{-1}$ ); peaks with confidence  $< 0.98$  were 251  
discarded.

### 252 3.2.2 Database Schema 252

253 The processed data is stored in a SQLite relational 253  
254 database (via SQLAlchemy in our implementation). 254  
255 The process of building this Raman database is 255  
256 outlined on the left-hand side of Figure 2. The 256  
257 schema has two main tables: samples and peaks. 257

258 samples(id, ts, x, y): Each row represents a mea- 258  
259 surement at a specific timestep (ts) and spatial po- 259  
260 sition (x,y). We assign a unique ID to each (ts, x, 260  
261 y). 261

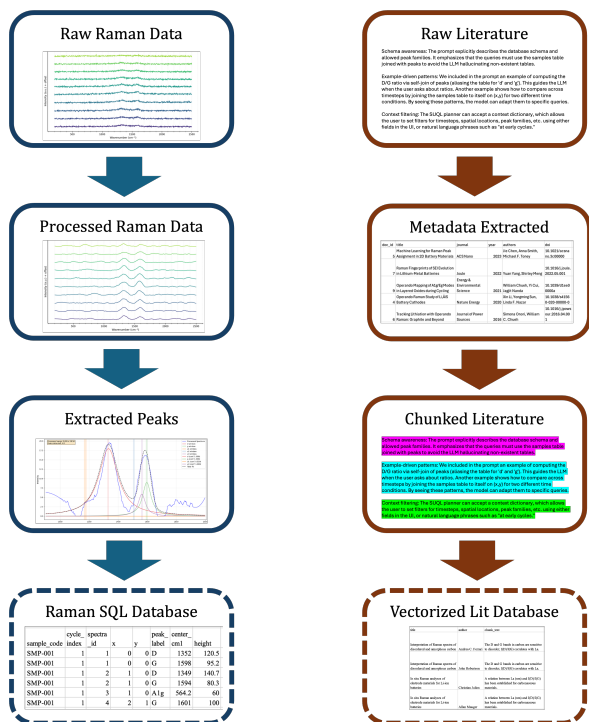


Figure 2: Data and literature pipelines. (Left) Raman preprocessing  $\rightarrow$  peak fitting  $\rightarrow$  relational tables. (Right) Literature ingestion  $\rightarrow$  metadata embeddings  $\rightarrow$  chunking  $\rightarrow$  vector search.

peaks(id, sample\_id, family, center, height, width): Each detected peak is a row linked via sample\_id to a specific sample. The family field is a categorical string  $\in \{\text{alg\_c, alg\_d, eg, d, g, u1, u2, u3}\}$ . If a peak family was not detected or filtered out at that sample, no row exists for that particular combination.

### 3.3 Unstructured Database

In parallel to building the SQL database, we constructed an unstructured text corpus of battery literature. The process of building this literature database is outlined on the right-hand side of Figure 2. 50 PDF documents were selected, including review articles on Raman in batteries, research papers on cathodes, papers discussing battery degradation, foundational texts on Raman signatures, etc. These PDFs were extracted into text using PyMuPDF, and then split into overlapping chunks (1000 tokens each with 150-token overlaps).

OpenAI’s text-embedding-ada-002 model, through the LiteLLM embedding API, was then applied to generate embeddings for each chunk, and these resulting embeddings and chunks were stored in a ChromaDB persistent index. Metadata, such as source paper title, page number, and

section header, were also stored for each chunk. The final vector store allows for fast similarity search: given a query embedding, it returns top-N chunks with cosine similarity scores.

## 3.4 SUQL Query Planner

### 3.4.1 Core Functionality

The core of SpectraQuery is the Structured-Unstructured Query Language (SUQL) planner. The process of this SUQL planner is outlined in Figure 3. The SUQL planner uses an LLM (this implementation used OpenAI GPT-4 via the LiteLLM API interface) to transform the user’s natural language question into two outputs: 1) a SQL SELECT query with optional parameters for the structured data, and 2) a concise keyword query for the literature search. This essentially decomposes the question into "What data do we need?" and "What background do we need?" We crafted a system prompt for the planner with detailed instructions and provided a few-shot examples for complex query types. Key aspects of the SUQL planner design include:

**Schema awareness:** The prompt explicitly describes the database schema and allowed peak families. It emphasizes that the queries must use the samples table joined with peaks to avoid the LLM hallucinating non-existent tables.

**Example-driven patterns:** We included in the prompt an example of computing the D/G ratio via self-join of peaks (aliasing the table for 'd' and 'g'). This guides the LLM when the user asks about ratios. Another example shows how to compare across timesteps by joining the samples table to itself on (x,y) for two different time conditions. By seeing these patterns, the model can adapt them to specific queries.

**Context filtering:** The SUQL planner can accept a context dictionary, which allows the user to set filters for timesteps, spatial locations, peak families, etc. using either fields in the UI, or natural language phrases such as "at early cycles."

**Safety checks:** After the LLM drafts the SQL, a validation function is run to ensure it is a safe SELECT-only query and only references the two allowed tables. Disallowed keywords such as DROP or DELETE trigger a rejection and error.

### 3.4.2 Parallel Query Execution

Once the SUQL planner returns the SQL and literature query, SpectraQuery executes them in parallel.

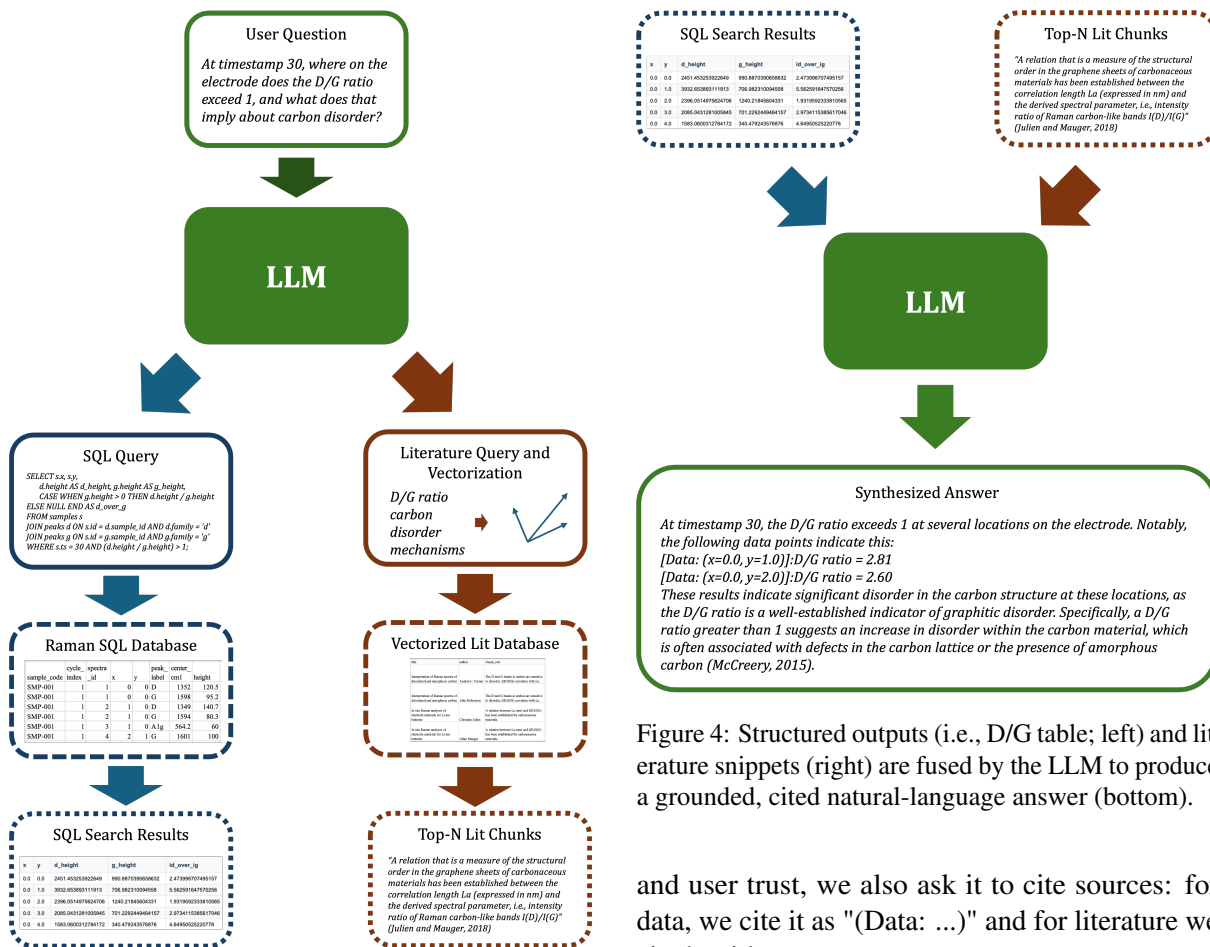


Figure 3: The planner parses a natural-language question, emits coordinated SQL over the Raman database (left) and a literature query over the vector index (right), and returns harmonized intermediate tables/snippets for generation.

The SQL query is run on the SQLite database, resulting in a set of rows which are formatted and possibly truncated for easier viewing. The literature query is vectorized using the aforementioned OpenAI text-embedding-ada-002 model, and a similarity search is performed in ChromaDB. The top k passages are retrieved, with text snippets and source metadata. Each passage comes with its text snippet and source metadata.

### 3.5 Answer Synthesis

The final step is to synthesize an answer that combines the structured data results and literature context. The process of this synthesis is outlined in Figure 4. We use the LLM in a second stage prompt for this. The prompt template instructs: summarize the relevant findings from the data, incorporate relevant points from the literature snippets, and produce a coherent answer. To maintain transparency

Figure 4: Structured outputs (i.e., D/G table; left) and literature snippets (right) are fused by the LLM to produce a grounded, cited natural-language answer (bottom).

and user trust, we also ask it to cite sources: for data, we cite it as "(Data: ...)" and for literature we cite by title.

## 4 Experimental Results

We evaluate SpectraQuery against three baseline configurations that isolate key components of the system: RAG-only (literature retrieval + synthesis, no SQL), SQL-only (execute generated SQL, no literature or synthesis), and text-to-SQL (a single SQL query generated directly from the question). These baselines reflect existing approaches commonly used in scientific question answering and data analysis. We assess SpectraQuery along four axes: (1) the correctness of SUQL-generated SQL queries, (2) the factual groundedness of synthesized answers, (3) literature retrieval effectiveness, and (4) expert perceptions of the final responses. Unless otherwise noted, all evaluations are performed on a benchmark set of 30 expert-curated Raman–battery questions derived from the operando LTMO cathode dataset (full list in Appendix A).

### 4.1 LLM-As-A-Judge Evaluation

#### 4.1.1 SQL Query Accuracy

We evaluated the correctness of SUQL-generated SQL using an LLM-as-a-judge protocol (Zheng et al., 2023). For each benchmark question, we

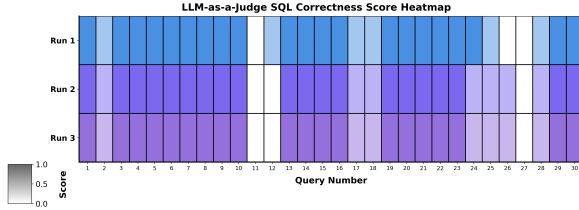


Figure 5: LLM-as-a-judge SQL correctness scores (0, 0.5, 1.0) for three independent runs across the 30 benchmark questions. Darker cells indicate higher correctness.

logged the natural-language prompt, the generated SQL, execution status, and returned rows, and asked GPT-5 to score query correctness on a three-level rubric: **1.0** if the SQL executes and fully satisfies the question (correct tables/joins/filters/aggregations; results match the requested information), **0.5** if it executes but only partially satisfies the question (i.e., missing a required condition or returning an incomplete slice), and **0.0** if it fails to execute or fundamentally misinterprets the question. To assess robustness, we repeated scoring three times with independent GPT-5 calls; across 30 questions, the fraction of queries scored fully correct (1.0) was **80.0%** (Run 1), **78.3%** (Run 2), and **78.3%** (Run 3). The remaining errors were split between partial and incorrect cases, typically due to missing one of several requested conditions (i.e., checking  $A_{1g}$  change but not D/G), flipped inequality directions, or incorrect joins between samples and peaks for cross-timestep comparisons. Figure 5 shows that these failures are concentrated in a small subset of more complex questions.

#### 4.1.2 Synthesized Answer Groundedness

We next measured how well final answers are grounded in retrieved evidence (Zheng et al., 2023). For each question, we concatenated the SQL result table with the top- $k$  retrieved literature passages and asked GPT-5 to assign a groundedness score on a discrete rubric: **1.0** if all claims are supported by the provided context, **0.5** if the answer is partially supported but includes unsupported or speculative statements, and **0.0** if key claims cannot be justified from the context. Varying the number of retrieved passages shows a clear retrieval-grounding trade-off: **83.3%** of answers are fully grounded with top-5 passages, rising to **93.3%** with top-10 and **96.7%** with top-15. Gains from 10 to 15 passages are smaller, suggesting diminishing returns beyond roughly 10 passages. Residual non-grounded cases

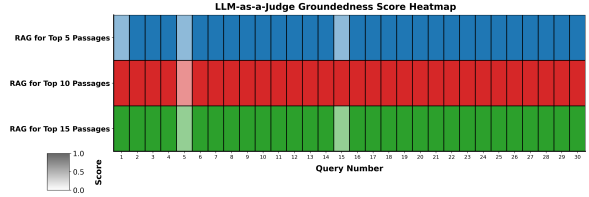


Figure 6: LLM-as-a-judge groundedness scores for synthesized answers when conditioning on the top-5, top-10, and top-15 retrieved passages.

| Metric      | @1    | @3    | @5    |
|-------------|-------|-------|-------|
| Precision   | 0.567 | 0.556 | 0.580 |
| Recall      | 0.433 | 0.533 | 0.600 |
| Unique Docs | –     | 1.80  | 2.30  |

Table 1: Retrieval performance aggregated over 30 benchmark queries.

are primarily due to missing retrieval (relevant mechanisms absent from the retrieved set) or mild over-generalization beyond the evidence. Figure 6 summarizes per-query groundedness across  $k$ .

#### 4.2 Retrieval Effectiveness Metrics

To better understand the retrieval component itself, we measured standard information retrieval metrics against a “gold” set of relevant papers (Singhal et al., 2001). For each of the 30 benchmark questions, we formed an expert-curated ranked list of the top five relevant ground-truth papers from our corpus that best addressed the phenomenon of interest. We then ran SpectraQuery’s literature retriever and computed paper-level **Precision@ $k$**  (the fraction of retrieved papers in the top  $k$  that appear in the ground-truth set) and **Recall@ $k$**  (the fraction of ground-truth papers that appear within the top  $k$  retrieved papers) (Manning, 2008), along with **UniqueDocs@ $k$**  (the number of distinct source papers represented among the top  $k$  retrieved passages, as a proxy for diversity) (Clarke et al., 2008).

Averaged across queries, we obtain the results in Table 1. These values indicate that when the retriever retrieves a paper, it is frequently relevant (precision around 0.56–0.58). Recall is more modest: by  $k = 5$ , we recover on average 60% of the manually identified relevant papers. UniqueDocs@ $k$  reveals that the top- $k$  passages often cluster on a small subset of documents (i.e., only 1.8 distinct papers among the top 3 passages).

Figure 7 provides a more fine-grained view across all 30 questions. Panel (a) visualizes passage-level Precision@ $k$ , which largely mirrors

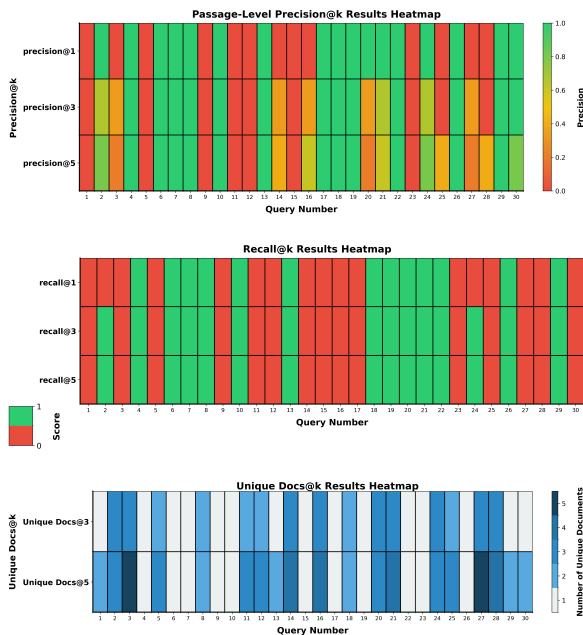


Figure 7: Per-query retrieval effectiveness visualized via Precision@k (top), Recall@k (middle), and Unique-Docs@k (bottom) across the 30 benchmark questions. Together, these heatmaps show that SpectraQuery retrieves highly relevant but somewhat redundant papers, with recall failures concentrated on a small subset of harder questions.

the paper-level trends and confirms that most retrieved passages are either clearly relevant (score 1.0) or clearly irrelevant (0.0). Panel (b) shows that recall failures are concentrated on a handful of queries whose information need is broad or abstract. Panel (c) highlights that many queries have UniqueDocs@3 = 1 and UniqueDocs@5 close to 2, indicating limited diversity: the retriever tends to return multiple passages from a single favored paper instead of surfacing complementary sources. This behavior explains some of the groundedness errors, since missing diversity can deprive the LLM of alternative mechanistic explanations.

### 4.3 Expert Feedback Evaluation

To complement automated metrics, we collected expert feedback from three battery scientists at the SLAC National Accelerator Laboratory (Van Schaik and Pugh, 2024; Semnani et al., 2025). Ten representative SpectraQuery answers were rated on a 1–5 Likert scale across eight dimensions, including scientific accuracy, grounding, relevance, clarity, depth, completeness, citation interpretability, and overall usefulness. Two exact query–answer pairs shown to experts are reproduced in Appendix B. Table 2 reports mean scores

| Metric                        | Mean Score (1–5) |
|-------------------------------|------------------|
| Scientific Accuracy           | 4.17             |
| Grounding in Evidence         | 4.37             |
| Relevance                     | 4.57             |
| Clarity                       | 4.33             |
| Depth of Insight              | 4.03             |
| Completeness                  | 4.13             |
| Interpretability of Citations | 3.27             |
| Overall Usefulness            | 4.13             |

Table 2: Average expert ratings for 10 SpectraQuery answers. Each score is averaged across 3 reviewers (30 ratings total).

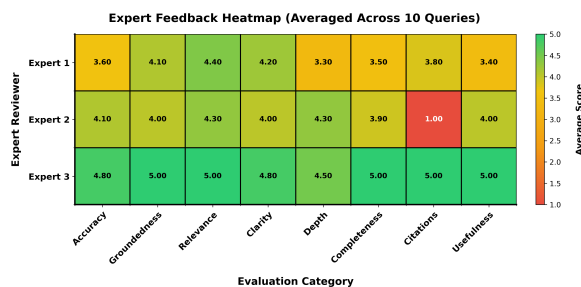


Figure 8: Expert feedback heatmap (10 questions averaged per cell). Each row corresponds to one reviewer and each column to one evaluation category, with numerical Likert scores overlaid.

over 30 judgments (10 answers  $\times$  3 reviewers), while Figure 8 visualizes per-reviewer scoring patterns.

Overall, experts rated the system favorably. Core quality metrics—accuracy, grounding, relevance, and clarity—are all above 4.0, indicating that responses are generally correct, well-supported, and directly address the questions. Depth and completeness scores near 4.0 suggest that the system typically captures the key mechanisms experts expect. The lowest-scoring dimension is citation interpretability (3.27), reflecting difficulty tracing references to specific sources or sections, and some variability across reviewers. Taken together, expert feedback validates SpectraQuery’s scientific credibility while highlighting retrieval completeness and citation presentation as primary areas for improvement.

## 5 Insights and Discussion

Across automated and expert evaluations, SpectraQuery is most reliable when the retriever surfaces the right mechanisms and diverse supporting sources. When key evidence is missing, answers

500 become incomplete or over-generalize despite cor- 550  
501 rect SQL execution. This coupling is consistent 551  
502 with our retrieval metrics (Recall@5 = 0.60; low di- 552  
503 versity with  $\sim 1$ –2 unique documents in the top- $k$ ), 553  
504 which often return multiple passages from a single 554  
505 favored paper. Improving recall and diversity via 555  
506 reranking or multi-stage retrieval is therefore the 556  
507 most direct lever for improving scientific reliability. 557

508 Furthermore, we observed that LLM-as-a-judge 558  
509 provides a scalable diagnostic for SQL correct- 559  
510 ness and groundedness, but experts apply stricter 560  
511 standards for completeness and citation usability. 561  
512 In practice, automatic judging is effective for ab- 562  
513 lations and error-finding, while expert review re- 563  
514 mains necessary to assess whether responses meet 564  
515 domain expectations for interpretability and scien- 565  
516 tific nuance. 566

## 517 6 Conclusions and Future Work 567

518 We presented SpectraQuery, a hybrid retrieval- 568  
519 augmented conversational system that allows re- 569  
520 searchers to pose natural-language questions over 570  
521 operando Raman datasets and scientific literature 571  
522 simultaneously. By combining a SUQL-inspired 572  
523 planner, a structured SQL backend over peak- 573  
524 parameter databases, and a domain-specific litera- 574  
525 ture retriever, SpectraQuery produces cited answers 575  
526 that unify numerical evidence with mechanistic 576  
527 explanations. To our knowledge, this is one of 577  
528 the first end-to-end systems that treats operando 578  
529 spectroscopy data and battery papers as a single 579  
530 queryable space. 580

531 Our evaluation shows that the approach is 581  
532 both feasible and useful. LLM-as-a-judge experi- 582  
533 ments indicate that roughly 80% of generated SQL 583  
534 queries are fully correct, and groundedness scores 584  
535 exceed 90% when the model is conditioned on suf- 585  
536 ficiently many retrieved passages. Expert battery 586  
537 scientists rate SpectraQuery’s answers highly on 587  
538 accuracy, grounding, relevance, and clarity, with 588  
539 somewhat lower scores on citation interpretability 589  
540 and completeness. These results suggest that hy- 590  
541 brid RAG architectures can meaningfully support 591  
542 real scientific workflows in niche domains, like 592  
543 battery spectroscopy. At the same time, our experi- 593  
544 ments highlight clear limitations; several avenues 594  
545 for improvement and extension are available. 595

546 First, SpectraQuery can be broadened beyond 596  
547 Raman spectra and text to incorporate additional 597  
548 battery-relevant modalities such as electrochemi- 598  
549 cal time-series (voltage profiles, impedance), opti- 599

550 cal microscopy images, and other characterization 550  
551 datasets (i.e., X-ray diffraction), enabling queries 551  
552 like “Compare Raman observations with capacity 552  
553 loss” or “What side reactions are occurring in dark 553  
554 cathode regions,” with the planner orchestrating 554  
555 retrieval across structured metrics, image-analysis 555  
556 outputs, and literature. Second, the system could 556  
557 expand from a local literature index to external 557  
558 knowledge bases—for example, integrating open 558  
559 resources like the Materials Project to retrieve re- 559  
560 ference spectra and expected peaks (i.e., lithium 560  
561 carbonate) directly from aggregated community 561  
562 data (Horton et al., 2025). Third, the LLM stack 562  
563 can be improved for cost and reliability: while 563  
564 GPT-5 is strong, smaller open models such as 564  
565 LLaMA-2 could be fine-tuned on domain-specific 565  
566 QA pairs to improve accuracy and reduce halluci- 566  
567 nations (Touvron et al., 2023), and complemented 567  
568 with a lightweight verification pass that checks each 568  
569 generated claim against retrieved evidence. Fourth, 569  
570 SpectraQuery could expose more user-facing tool- 570  
571 ing, including an optional interactive “debug mode” 571  
572 that allows expert users to inspect and edit inter- 572  
573 mediate SQL and retrieval results before synthesis, 573  
574 and richer visualization capabilities such as spectra 574  
575 overlays across timesteps, coordinate-linked plots, 575  
576 and spatial heatmaps (i.e., “Show me the D/G ratio 576  
577 map at  $t=30$ ”) to make outputs more interpretable 577  
578 and actionable. Finally, our current implementation 578  
579 executes SQL and literature retrieval in parallel. A 579  
580 promising extension is an *iterative two-pass* strat- 580  
581 egy: preliminary SQL findings (i.e., unexpected 581  
582 spatial hotspots or peak shifts) can trigger a refined 582  
583 literature query, and retrieved mechanisms can in 583  
584 turn suggest additional structured features to com- 584  
585 pute. This closed-loop interaction could reduce 585  
586 missed mechanisms and improve completeness on 586  
587 hard queries. 587

588 In summary, this work contributes an evaluation 588  
589 methodology that combines automated LLM-as- 589  
590 a-judge scoring with targeted expert review. We 590  
591 find that strong LLM judges correlate well with 591  
592 experts on obvious factuality errors but miss finer- 592  
593 grained concerns about completeness and presen- 593  
594 tation. Future AI assistants for scientific domains 594  
595 will likely need both scalable automatic metrics 595  
596 and carefully designed human studies to ensure 596  
597 that they truly augment, rather than mislead, do- 597  
598 main experts. SpectraQuery is a first step in this 598  
599 direction, illustrating how structured and unstruc- 599  
600 tured data can be jointly leveraged to accelerate 600  
601 insight generation in battery research. 601

## 602 Limitations

603 SpectraQuery has several limitations that motivate  
604 future work. First, the system’s reliability depends  
605 on retrieval quality: while precision is relatively  
606 high, recall and document diversity are imperfect,  
607 meaning that relevant mechanisms may be absent  
608 from the retrieved context for complex or niche  
609 queries. In such cases, answers may be incomplete  
610 or rely too heavily on a single source.

611 Second, structured reasoning depends on the cor-  
612 rectness of SUQL-generated SQL. Although most  
613 queries are correct, errors such as missing condi-  
614 tions or incorrect joins can distort numerical sum-  
615 maries if undetected. The current system mitigates  
616 this risk through evaluation and filtering, but does  
617 not yet perform full semantic verification of query  
618 intent.

619 Third, SpectraQuery is evaluated on a single  
620 operando Raman dataset and a curated literature  
621 corpus. While the architecture is general, perfor-  
622 mance may vary across materials systems, experi-  
623 mental modalities, or scientific domains with dif-  
624 ferent data distributions or terminology.

625 Fourth, portions of the programming and soft-  
626 ware development were performed with the assis-  
627 tance of AI-based tools, including GPT-4 and the  
628 Cursor development environment, which were used  
629 as productivity aids during code implementation  
630 and debugging. GPT-4 was also used to support  
631 editing and refinement of the manuscript text. All  
632 experimental design choices, evaluations, and in-  
633 terpretations were authored and verified by the au-  
634 thors, but the use of such tools may introduce subtle  
635 biases that are difficult to fully quantify.

636 Fifth, we do not provide a full quantitative  
637 head-to-head comparison against proprietary long-  
638 context assistants or end-to-end text-to-SQL sys-  
639 tems, because many baselines do not expose com-  
640 parable intermediate artifacts (executed SQL, re-  
641 trieved passages) needed for our correctness and  
642 grounding evaluations. Instead, we provide concep-  
643 tual baselines aligned with the benchmark require-  
644 ments and emphasize reproducible component-  
645 level metrics.

646 Finally, expert evaluation involved three domain  
647 experts who are professional colleagues of the au-  
648 thors. All reviewers provided explicit consent for  
649 participation and for the use of their anonymized  
650 feedback in this study. The limited number of re-  
651 viewers and their professional proximity to the au-  
652 thors may constrain the generalizability of the qual-

itative findings. 653

## Ethical Considerations and Potential Risks 654

655 SpectraQuery is designed as a research assistant  
656 to support scientific interpretation of experimental  
657 battery data and literature, rather than as an au-  
658 tonomous decision-making system. As such, its  
659 primary risks arise from misinterpretation or over-  
660 reliance on generated explanations. If used without  
661 expert oversight, incomplete retrieval or incorrect  
662 SQL could lead to misleading conclusions about  
663 material behavior or degradation mechanisms.

664 The system does not generate new experimen-  
665 tal data, access personal information, or operate  
666 on sensitive human data. All inputs consist of lab-  
667 oratory measurements and publicly available sci-  
668 entific literature. Nevertheless, users should treat  
669 generated explanations as hypotheses supported by  
670 retrieved evidence, not as definitive conclusions.

671 A secondary risk concerns overgeneralization.  
672 The system is evaluated on a specific materials sys-  
673 tem and literature corpus, and its outputs may not  
674 transfer reliably to other chemistries or experimen-  
675 tal settings. We mitigate these risks by emphasizing  
676 interpretability, explicit citation of sources, and po-  
677 sitioning SpectraQuery as a decision-support tool  
678 rather than a substitute for domain expertise.

679 Finally, while AI-based tools were used to assist  
680 code development and writing, all scientific claims  
681 and evaluations were reviewed and validated by the  
682 author. The system is intended for research use  
683 only, and not for deployment in safety-critical or  
684 operational battery management settings.

## References 685

- 686 Sören Auer, Dante AC Barone, Cassiano Bartz, Ed-  
687 uardo G Cortes, Mohamad Yaser Jaradeh, Oliver  
688 Karras, Manolis Koubarakis, Dmitry Mouromt-  
689 sev, Dmitrii Pliukhin, Daniil Radyush, and others.  
690 2023. The sciqa scientific question answering bench-  
691 mark for scholarly knowledge. *Scientific Reports*,  
692 13(1):7240. Publisher: Nature Publishing Group UK  
693 London.
- 694 Rita Baddour-Hadjean and Jean-Pierre Pereira-Ramos.  
695 2010. Raman microspectrometry applied to the study  
696 of electrode materials for lithium batteries. *Chem-  
697 ical reviews*, 110(3):1278–1319. Publisher: ACS  
698 Publications.
- 699 Wenhu Chen, Hanwen Zha, Zhiyu Chen, Wenhan Xiong,  
700 Hong Wang, and William Wang. 2020. HybridQA: A  
701 dataset of multi-hop question answering over tabular  
702 and textual data. *arXiv preprint arXiv:2004.07347*.

|     |   |  |     |
|-----|---|--|-----|
| 703 | Charles LA Clarke, Maheedhar Kolla, Gordon V                  | Shicheng Liu, Jialiang Xu, Wesley Tjangnaka, Sina J          | 759 |
| 704 | Cormack, Olga Vechtomova, Azin Ashkan, Stefan                 | Semnani, Chen Jie Yu, and Monica S Lam. 2023.                | 760 |
| 705 | Büttcher, and Ian MacKinnon. 2008. Novelty and                | SUQL: Conversational search over structured and              | 761 |
| 706 | diversity in information retrieval evaluation. In <i>Pro-</i> | unstructured data with large language models. <i>arXiv</i>   | 762 |
| 707 | <i>ceedings of the 31st annual international ACM SIGIR</i>    | <i>preprint arXiv:2311.09818</i> .                           | 763 |
| 708 | <i>conference on Research and development in informa-</i>     |  |     |
| 709 | <i>tion retrieval</i> , pages 659–666.                        |  |     |
| 710 | Nicolas Coca-Lopez, Victor Alcolea-Rodriguez,                 | Renqian Luo, Liai Sun, Yingce Xia, Tao Qin, Sheng            | 764 |
| 711 | Miguel A Bañares, Sandor Brockhauser, Julien                  | Zhang, Hoifung Poon, and Tie-Yan Liu. 2022.                  | 765 |
| 712 | Gorenflot, Alex Henderson, Ron Hildebrandt, Nina              | BioGPT: generative pre-trained transformer for               | 766 |
| 713 | Jeliazkova, Nikolay Kochev, Enrique Lozano-Diz,               | biomedical text generation and mining. <i>Briefings</i>      | 767 |
| 714 | Zdenek Pilat, Dario Polli, Philip Strömert, Chris             | <i>in bioinformatics</i> , 23(6):bbac409. Publisher: Oxford  | 768 |
| 715 | Sturm, Renzo Vanna, and Raquel Portela. 2025.                 | University Press.  | 769 |
| 716 | Open and FAIR Raman spectroscopy: Paving the                  |  |     |
| 717 | way for artificial intelligence.                              | Christopher D Manning. 2008. <i>Introduction to informa-</i> | 770 |
| 718 | Andrea C Ferrari and John Robertson. 2000. Interpreta-        | <i>tion retrieval</i> . Syngress Publishing.,                | 771 |
| 719 | tion of Raman spectra of disordered and amorphous             |  |     |
| 720 | carbon. <i>Physical Review B</i> , 61(20):14095–14107.        | Sina Semnani, Jirayu Burapachep, Arpandeeep Khatua,          | 772 |
| 721 | Andrea Carlo Ferrari and John Robertson. 2001. Reso-          | Thanawan Atcharyachanvanit, Zheng Wang, and                  | 773 |
| 722 | nant Raman spectroscopy of disordered, amor-                  | Monica Lam. 2025. Detecting corpus-level knowl-              | 774 |
| 723 | phous, and diamondlike carbon. <i>Physical review</i>         | edge inconsistencies in wikipedia with large language        | 775 |
| 724 | <i>B</i> , 64(7):075414. Publisher: APS.                      | models. In <i>Proceedings of the 2025 Conference on</i>      | 776 |
| 725 | Eibar Flores, Petr Novák, and Erik J Berg. 2018. In           | <i>Empirical Methods in Natural Language Processing</i> ,    | 777 |
| 726 | situ and operando Raman spectroscopy of layered               | pages 34827–34854.   | 778 |
| 727 | transition metal oxides for Li-ion battery cathodes.          | Amit Singhal and 1 others. 2001. Modern information          | 779 |
| 728 | <i>Frontiers in Energy Research</i> , 6:82. Publisher: Fron-  | retrieval: A brief overview. <i>IEEE Data Eng. Bull.</i> ,   | 780 |
| 729 | tiers Media SA.   | 24(4):35–43.   | 781 |
| 730 | Marcel Heber and Christian Hess. 2021. Monitoring             | Hugo Touvron, Louis Martin, Kevin Stone, Peter Al-           | 782 |
| 731 | electrode/electrolyte interfaces of Li-ion batteries un-      | bert, Amjad Almahairi, Yasmine Babaei, Nikolay               | 783 |
| 732 | der working conditions: A surface-enhanced Raman              | Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti           | 784 |
| 733 | spectroscopic study on LiCoO <sub>2</sub> composite cathodes. | Bhosale, and others. 2023. Llama 2: Open foun-               | 785 |
| 734 | Koji Hiraoka, Yoshiki Yokoyama, Sarina Mine, Kazuo            | dation and fine-tuned chat models. <i>arXiv preprint</i>     | 786 |
| 735 | Yamamoto, and Shiro Seki. 2025. Advanced Ra-                  | <i>arXiv:2307.09288</i> .                                    | 787 |
| 736 | man spectroscopy for battery applications: Materials          | Tempest A Van Schaik and Brittany Pugh. 2024. A              | 788 |
| 737 | characterization and operando measurements. <i>APL</i>        | field guide to automatic evaluation of llm-generated         | 789 |
| 738 | <i>Energy</i> , 3(2). Publisher: AIP Publishing.              | summaries. In <i>Proceedings of the 47th International</i>   | 790 |
| 739 | Matthew K Horton, Patrick Huck, Ruo Xi Yang, Ja-              | <i>ACM SIGIR Conference on Research and Develop-</i>         | 791 |
| 740 | son M Munro, Shyam Dwaraknath, Alex M Ganose,                 | <i>ment in Information Retrieval</i> , pages 2832–2836.      | 792 |
| 741 | Ryan S Kingsbury, Mingjian Wen, Jimmy X Shen,                 | Pengcheng Xue, Rui Qiu, Chuchuan Peng, Zehang Peng,          | 793 |
| 742 | Tyler S Mathis, and others. 2025. Accelerated data-           | Kui Ding, Rui Long, Liang Ma, and Qifeng Zheng.              | 794 |
| 743 | driven materials science with the Materials Project.          | 2024. Solutions for lithium battery materials data           | 795 |
| 744 | <i>Nature Materials</i> , pages 1–11. Publisher: Nature       | issues in machine learning: Overview and future out-         | 796 |
| 745 | Publishing Group UK London.                                   | look. <i>Advanced Science</i> , 11(48):2410065. Publisher:   | 797 |
| 746 | Gautier Izacard, Patrick Lewis, Maria Lomeli, Lucas           | Wiley Online Library.  | 798 |
| 747 | Hosseini, Fabio Petroni, Timo Schick, Jane Dwivedi-           | Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak             | 799 |
| 748 | Yu, Armand Joulin, Sebastian Riedel, and Edouard              | Shafran, Karthik R Narasimhan, and Yuan Cao. 2022.           | 800 |
| 749 | Grave. 2023. Atlas: Few-shot learning with retrieval          | React: Synergizing reasoning and acting in language          | 801 |
| 750 | augmented language models. <i>Journal of Machine</i>          | models. In <i>The eleventh international conference on</i>   | 802 |
| 751 | <i>Learning Research</i> , 24(251):1–43.                      | <i>learning representations</i> .                            | 803 |
| 752 | Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio          | Tao Yu, Zifan Li, Zilin Zhang, Rui Zhang, and Dragomir       | 804 |
| 753 | Petroni, Vladimir Karpukhin, Naman Goyal, Hein-               | Radev. 2018. Typesql: Knowledge-based type-                  | 805 |
| 754 | rich Küttler, Mike Lewis, Wen-tau Yih, Tim Rock-              | aware neural text-to-sql generation. <i>arXiv preprint</i>   | 806 |
| 755 | täschel, and Sebastian Riedel. 2020. Retrieval-               | <i>arXiv:1804.09769</i> .                                    | 807 |
| 756 | augmented generation for knowledge-intensive NLP              | Shuo Zhao, Sihui Chen, Jiayi Zhou, Chao Li, Tan Tang,        | 808 |
| 757 | tasks. In <i>Advances in Neural Information Processing</i>    | Stephen J Harris, Yang Liu, Jiayu Wan, and Xin Li.           | 809 |
| 758 | <i>Systems (NeurIPS)</i> , volume 33, pages 9459–9474.        | 2024. Potential to transform words to watts with             | 810 |
|     |   | large language models in battery research. <i>Cell Re-</i>   | 811 |
|     |   | <i>ports Physical Science</i> , 5(3). Publisher: Elsevier.   | 812 |

813 Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan  
814 Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin,  
815 Zhuohan Li, Dacheng Li, Eric Xing, and 1 others.  
816 2023. Judging llm-as-a-judge with mt-bench and  
817 chatbot arena. *Advances in neural information pro-*  
818 *cessing systems*, 36:46595–46623.

819 Wen Zhu, Dongqiang Liu, Andrea Paoletta, Catherine  
820 Gagnon, Vincent Gariépy, Ashok Vijh, and Karim  
821 Zaghbi. 2018. [Application of operando X-ray diffraction and Raman spectroscopies in elucidating the behavior of cathode in lithium-ion batteries.](#) *Frontiers in Energy Research*, 6:66.

|     |   |  |  |     |
|-----|---|--|--|-----|
| 825 | <b>A Benchmark Question Set</b>                           |  |  |     |
| 826 | To evaluate SpectraQuery’s SQL correctness, re-           |  | mean about behavior on the cathode edge ver-                             | 869 |
| 827 | trieval quality, groundedness, and expert-rated use-      |  | sus in the middle?   | 870 |
| 828 | fulness, we curated 30 Raman–battery analysis             |  |  |     |
| 829 | questions representing real scientific workflows of       |  | 10. Find timesteps and coordinates where the u3                          | 871 |
| 830 | existing collaborators. Questions probe peak evolu-       |  | height > 200, and discuss what this means                                | 872 |
| 831 | tion, degradation signatures, spatial heterogeneity,      |  | with respect to side reactions and lithium loss.                         | 873 |
| 832 | unknown-mode behavior, and mechanistic interpre-          |  |  |     |
| 833 | tation.   |  | <b>A.2 B. Peak Shape, Lattice Disorder, and</b>                          | 874 |
| 834 | The benchmark set is divided below into the-              |  | <b>Structural Transition Questions</b>                                   | 875 |
| 835 | matic categories.   |  |  |     |
| 836 | <b>A.1 A. Expert Evaluation Subset (Questions</b>         |  | 11. Compare the A1g <sub>c</sub> FWHM at the first and last              | 876 |
| 837 | <b>1–10)</b>  |  | timesteps. What does the broadening tell us                              | 877 |
| 838 | These ten questions span core Raman diagnostics           |  | about lattice disorder accumulation?                                     | 878 |
| 839 | used by domain scientists and were selected for           |  |  |     |
| 840 | blinded expert scoring of accuracy, clarity, ground-      |  | 12. Compute the A1g <sub>c</sub> /E <sub>g</sub> intensity ratio at each | 879 |
| 841 | ing, and usefulness.                                      |  | timestep. At which timestep is the ratio lowest,                         | 880 |
|     |   |  | and what structural transition is associated                             | 881 |
|     |   |  | with this?   | 882 |
| 842 | 1. Which timestep has the highest average A1g             |  |  |     |
| 843 | charged height, and what does that imply                  |  | 13. Find any coordinates where the A1g <sub>c</sub> center               | 883 |
| 844 | about the state of charge at that timestep?               |  | > 595 cm <sup>-1</sup> . Summarize literature describing                 | 884 |
|     |   |  | A1g blue-shifts during high-voltage oxygen                               | 885 |
| 845 | 2. Which timestep has the highest D/G ratio, and          |  | activity.  | 886 |
| 846 | what does that imply about carbon disorder?               |  |  |     |
|     |   |  | 14. Determine where the A1g <sub>c</sub> and A1g <sub>d</sub> peaks      | 887 |
| 847 | 3. What is the single highest A1g discharged              |  | disappear entirely. What failure mechanism                               | 888 |
| 848 | peak, and what does that mean?                            |  | does complete A1g loss correspond to?                                    | 889 |
|     |   |  |  |     |
| 849 | 4. Find coordinates/timestamps where u2 height            |  | 15. At timestep 75, which 10 coordinates have                            | 890 |
| 850 | > 200 and discuss literature connecting such              |  | the highest A1g <sub>c</sub> intensity, and what might                   | 891 |
| 851 | carbon-region unknowns to side products and               |  | this say about spatially selective retention of                          | 892 |
| 852 | capacity fade.  |  | TM–O structural integrity?   | 893 |
|     |   |  |  |     |
| 853 | 5. What is the ratio of the average A1g charged           |  | 16. At which timestep does the average G-band                            | 894 |
| 854 | height to the average A1g discharged height               |  | height reach a maximum? What does this                                   | 895 |
| 855 | at timestep 60? What does this ratio mean?                |  | mean about graphitic ordering?   | 896 |
|     |   |  |  |     |
| 856 | 6. Which (x, y) coordinate has the highest av-            |  | <b>A.3 C. Spatial Heterogeneity and</b>                                  | 897 |
| 857 | erage D height, and what does that mean for               |  | <b>Electrode-Asymmetry Questions</b>                                     | 898 |
| 858 | carbon disorder?  |  |  |     |
|     |   |  | 17. Compare the D/G ratios for the left vs. right                        | 899 |
| 859 | 7. Which timestep has the single coordinate with          |  | halves of the electrode at timestep 40. What                             | 900 |
| 860 | the largest sum of u1, u2, and u3 peaks, and              |  | spatial asymmetry does this reveal?                                      | 901 |
| 861 | what does that indicate about side reactions?             |  |  |     |
|     |   |  | 18. Which timestep has the lowest D/G ratio, and                         | 902 |
| 862 | 8. What is the difference between the average             |  | what does that indicate about carbon structural                          | 903 |
| 863 | u3 height at the last timestep and at the first           |  | recovery (if any)?   | 904 |
| 864 | timestep, and what does this tell us about                |  |  |     |
| 865 | degradation due to cycling?                               |  | 19. At coordinates where D > 400, retrieve lit-                          | 905 |
|     |   |  | erature linking high-D features to electrolyte                           | 906 |
| 866 | 9. At timestep 30, what is the A1g <sub>c</sub> height at |  | attack or carbon amorphization.  | 907 |
| 867 | location (0,0)? What is the A1g <sub>c</sub> height at    |  |  |     |
| 868 | location (15,15)? What does this difference               |  | 20. Identify any coordinates where the G-band                            | 908 |
|     |   |  | disappears but A1g persists. What edge-case                              | 909 |
|     |   |  | failure or delamination does this pattern map                            | 910 |
|     |   |  | onto?  | 911 |

- 912 21. Determine all points where  $u_2$  and  $u_3$  are si-  
 913 multaneously  $>150$ . What multipeak signa-  
 914 tures are associated with multi-step side reac-  
 915 tions?
- 916 22. Find coordinates where  $u_3$  height  $> A1g_c$   
 917 height. What does dominance of un-  
 918 known/byproduct modes imply about end-of-  
 919 life chemistry?
- 920 23. Compare  $A1g$  intensity at coordinates (0,0),  
 921 (15,15), and (29,29). Which region ages  
 922 fastest?
- 923 24. Determine which quadrant of the electrode  
 924 has the highest mean  $u_2$  height at the last  
 925 timestep. What could be causing quadrant-  
 926 specific side reactions?
- 927 25. Find edge vs. center contrast in  $A1g$  FWHM  
 928 at the last timestep. Why does the literature  
 929 say edges degrade faster?

930 **A.4 D. Degradation Kinetics and Mechanistic**  
 931 **Interpretation Questions**

- 932 26. Compute the percentage loss of  $A1g$  intensity  
 933 from timestep 0  $\rightarrow$  final. Interpret this loss in  
 934 the context of typical TM–O bond destabiliza-  
 935 tion percentages.
- 936 27. Determine when the  $A1g$  discharged peak first  
 937 falls below 50% of its initial value, and ex-  
 938 plain the mechanistic significance.
- 939 28. Compute the  $A1g_{charged}/A1g_{discharged}$  ratio  
 940 across all timesteps and find when it deviates  
 941 most from unity. How does this ratio map to  
 942 irreversibility?
- 943 29. For all timesteps where  $u_2 > 150$ , retrieve  
 944 literature that links mid-frequency Raman fea-  
 945 tures in LTMO cathodes to parasitic oxygen  
 946 reactions.
- 947 30. Return all timesteps where average  $A1g_c <$   
 948 average D. What scenario does “carbon domi-  
 949 nating over lattice” represent?

950  
951  
952  
953  
954  
955

## **B Expert Evaluation Packet**

The following pages contain the first two Spectra-Query query–answer examples provided to expert evaluators. These were used to assess scientific accuracy, grounding, clarity, completeness, and overall usefulness of the system.

## SpectraQuery - Query History

Generated: 2025-11-17 16:52:38

### Query #1

#### Question:

Which timestep has the highest average A1g charged height, and what does that imply about the state of charge at that timestep?

#### SQL Query:

```
SELECT s.ts, AVG(a1g_height) AS avg_a1g_height FROM samples s JOIN peaks a1g ON s.id = a1g.sample_id AND a1g.family IN ('a1g_c', 'a1g_d') GROUP BY s.ts ORDER BY avg_a1g_height DESC LIMIT 1;
```

#### SQL Results:

| ts  | avg_a1g_height   |
|-----|------------------|
| 0.0 | 707.692353122218 |

#### Literature Results:

[1] Advanced Raman spectroscopy for battery applications- Materials characterization and operando measurements (similarity: 0.249)

Authors: Koji Hiraoka et al.

Journal: APL Energy

Year: 2025

Citation: (Hiraoka et al., 2025)

Source file: Advanced Raman spectroscopy for battery applications- Materials characterization and operando measurements.pdf

Passage: The charge-discharge CV profile is presented in Fig. 6(c), showing current peaks at 3.0 and 3.5 V during charging and 3.0 V during discharging. These current peaks indicate significant irreversible capacity during the first cycle and degradation with cycling, as evidenced by broader and decreasing current values, especially for the first cycle. However, it was difficult to understand the degradation mechanism based on only the CV profile, and Raman spectra were acquired from the center of the positive electrode, negative electrode, and solid electrolyte layers in the ASSB cross section during charge-discharge cycles, as shown in Fig. 6(d). The solid electrolyte maintained a pristine state over the three charge-discharge cycles, indicating a stable crystal structure in the central region during Na<sup>+</sup> transport between electrodes. In contrast, the Raman spectra of the positive and negative electrode layers showed irreversible changes from their

Figure 9: Expert evaluation packet, page 1.

pristine states during the first charge-discharge cycle. These results imply that the crystal structures of NNPP and NTP transition to metastable states, contributing to decreased electrochemical activity and capacity. Furthermore, the Raman spectra during the third cycle did not change, suggesting the formation of unreactive crystal structures in NNPP and NTP. These irreversible crystal structures directly influence capacity degradation during cycling. Overall, operando Raman spectroscopy is effective for evaluating degradation mechanisms in ASSBs by observing crystal structure changes at the sub-micrometer scale. C. Carbon structural changes in sulfide-based all-solid-state batteries In next-generation secondary batteries, carbon materials play important roles as conductive additives with disordered structures [e.g., AB, ketjen black, carbon black (CB), and carbon nanotubes] in APL Energy 3, 021502 (2025); doi: 10.1063/1.5272588 3, 021502-9 © Author(s) 2025 04 November 2025 21:38:34

[2] In situ Raman analyses of electrode materials for Li-ion batteries (similarity: 0.242)

Authors: Christian M Julien & Alain Mauger

Journal: AIMS Materials Science

Year: 2018

Citation: (Julien et al., 2018)

Source file: In situ Raman analyses of electrode materials for Li-ion batteries.pdf

Passage: It is claimed that Raman spectral changes are correlated with the potential plateaus recorded on the charging reaction as follows: a phase transition from dilute stage-1 to stage-4 followed by the stage-2' (LiC18) → stage-2 (LiC12) and stage-2' (LiC12) → stage-1 (golden LiC6) transitions. This model will be modified in further investigations of Lonza KS-44 and KS-6 graphite materials [136]. Pantz et al. [44] presented in situ Raman patterns and imaging of electrode made of TIMREX SFG 44 synthetic graphite with PVDF binder, which reveal a new band at ca. 1850 cm<sup>-1</sup> at potential negative to 0.18 V vs. Li/Li<sup>+</sup> assigned to the complex between Li ions and decomposition of the aprotic solution. This electrolyte/graphite interaction was later reconsidered by Nova k et al. [137] in terms of SEI. The so-called dilute stage-1 phase of GICs was deeply studied on KS-44 carbon particles (2–50 μm in diameter) embedded in heat-treated Ni foils [138]. Kostock and McLarnon [139] investigated the structural degradation of graphite electrodes by poring the graphite D/G band ratio of the anode via a 50 μm × 75 μm cross-section Raman image.

[3] In situ Raman analyses of electrode materials for Li-ion batteries (similarity: 0.232)

Authors: Christian M Julien & Alain Mauger

Journal: AIMS Materials Science

Year: 2018

Citation: (Julien et al., 2018)

Source file: In situ Raman analyses of electrode materials for Li-ion batteries.pdf

Passage: 684 AIMS Materials Science Volume 5, Issue 4, 650–698, Figure 19. (a) Discharge-charge profile of the Li xSb anode. (b) In situ Raman spectra collected in the potential between 0.0 V and 0.02 V vs. Li/Li<sup>+</sup>. Reprinted by permission from Ref. [194]. 4.5.2. ZnM2O4 (M = Mn, Fe) Among the ternary Zn-based oxides, the spinel ZnMn2O4 material reacts with lithium with both conversion and alloying process, which results in a specific capacity 1024 mAh g<sup>-1</sup>. In situ Raman spectra of mesoporous ZnMn2O4 microspheres were collected in the potential range 2.0–0.01 V vs. Li/Li<sup>+</sup> (Figure 20) [19 5]. The structural changes were monitored by following the

Figure 10: Expert evaluation packet, page 2.

evolution of the three Raman bands located at 681 cm<sup>-1</sup> (A1g symmetry involving motion of oxygen in A0 4 tetrahedra), 386 and 324 cm<sup>-1</sup> (involving the vibration of BO 6 octahedra) according to the usual notations for compounds of generic form AB2O4. When the Mn 3+ ions are reduced to Mn 2+ during the lithiation process at 1.21 V, the A1g band begins to wane and disappears at 0.16 V, while the strength of Raman bands at 386 and 324 cm<sup>-1</sup> decreases. At this potential Zn2+ and Mn2+ are reduced to Zn0 and Mn0. During the charge process (delithiation) at 1.25 V, a band grows at 694 cm<sup>-1</sup> which is characteristic of the asymmetric Mn–O stretch for R-Mn2O3 associated with the oxidation of Mn 0 to Mn3+. Above 1.52 V, a band located at 407 cm<sup>-1</sup> assigned to the E2 mode of ZnO is the fingerprint of oxidized Zn0. Therefore, the two weak Raman bands at 324 and 386 cm<sup>-1</sup> for ZnMn2O4 reappear at the end of the charge process. Cabo-Fernandez et al. [196] investigated the delithiation process of carbon-coated ZnFe2O4 (ZFO) nanoparticles. Like ZnMn2O4, ZFO is a low-cost anode material with a theoretical specific capacity ~1000 mAh g<sup>-1</sup> that reacts with lithium via a conversion/alloying process. The first-order Raman spectrum displays 5 bands located at 647 cm<sup>-1</sup> (A1g), 467 cm<sup>-1</sup> (T2g), 340 cm<sup>-1</sup> (T2g), 246 cm<sup>-1</sup> (Eg) and 221 cm<sup>-1</sup> (T2g). The conversion reaction that occurs at ~0.8 V leads the formation of Zn 0 particles, which enhance the Raman signal of the SEI components. In the electrode potential between 0.80 and 0.69 V, the Raman spectrum is dominated by the reduction products of electrolyte carbonate solvents which form the SEI.

[4] Capacity Decay Mechanism for Lithium-rich Layered Oxides under 2.0-4.5V; Speculation and Experimental Study (similarity: 0.228)

Authors: Jiahui Xiong et al.

Journal: Capacity Decay Mechanism for Lithium-Rich Layered Oxides Under

Citation: (Xiong et al.)

Source file: Capacity Decay Mechanism for Lithium-rich Layered Oxides under 2.0-4.5V; Speculation and Experimental Study.pdf

Passage: To understand the effect of the cross-linking process during cycling, the cathodes of disassembled pouch cells after 400 cycles are characterized by XPS, XRD, and Raman spectroscopy. Furthermore, the anodes of disassembled pouch cells after 400 cycles are characterized by ICP. Fig. 5 (a-f) shows the obtained XPS patterns for both cathodes after 400 cycles (BC400 and AC400). Figs. 8 a and b show the line spectra of O 1s belonging to the BC400 and AC400. The obtained peaks at 529.5, 531.2, 532.1, and 533.5 cm<sup>-1</sup> correspond to structural oxygen, oxidized oxygen, oxygen deposit, and electrolyte oxide, respectively [30]. Comparing both patterns of Fig. 8 a and b, the intensity of the structural oxygen peak at 529.5 cm<sup>-1</sup> are almost similar, indicating that the cross-linking reaction could not destroy the internal structure of the cathodes. However, the significant reduction that occurred in the peak intensity at 531.2 cm<sup>-1</sup> is related to O 2- in the oxidation process [31]. The oxidation process of O 2- into O n- can be reduced due to the occupied links between the surface lattice oxygen and COOH-CNT. The peak at 532.1 cm<sup>-1</sup> that belongs to the oxygen deposit is mainly composed of Li2CO3[32], where the surface lattice oxygen reacts with the EC to form lithium carbonate during cycling [21]. The smaller XPS peak of the lithium carbonate after cross-linking process indicates that the cross-linking decreases the formation of lithium carbonate during cycling, which is in line with the speculation already proposed. Figs. 5c and d show the XPS patterns of Ni 2p belonging to the BC400 and This preprint research paper has not been peer reviewed. Electronic copy available at: https://arxiv.org/abs/2011.15281 Preprint not peer reviewed

[5] In situ Raman analyses of electrode materials for Li-ion batteries (similarity: 0.228)

Figure 11: Expert evaluation packet, page 3.

Authors: Christian M Julien & Alain Mauger

Journal: AIMS Materials Science

Year: 2018

Citation: (Julien et al., 2018)

Source file: In situ Raman analyses of electrode materials for Li-ion batteries.pdf

Passage: 671 AIMS Materials Science Volume 5, Issue 4, 650–698, Figure 14. Evolution of in situ Raman spectra in one charge-discharge cycle at C/2 rate for (a) LiMn1.45Ni0.45Co0.1O4 and (b) LiMn1.45Ni0.45Co0.1O4. Reprinted by permission from Ref. [103]. 3.3. Vanadium pentoxide (V2O5) V2O5 is an attractive electrode material which can accommodate 3 Li per mole of oxide to deliver a specific capacity of 450 mAh g<sup>-1</sup> within the potential range 1.5–4.5 V vs. Li/Li<sup>+</sup>[104]. Vanadium pentoxide, V2O5, with high-valent V 5+ cations, are layered structures which consist of edge and corner-sharing VO5 square pyramids attached by weak chemical bonds favourable to Li insertion process. The phase diagram of Li xV2O5 is rather complex with reversible and irreversible phases across the composition 0 < x < 3. Over the composition 0 ≤ x ≤ 3, Li xV2O5 undergoes the phase transitions with the corresponding potential plateaus α → ε → δ and β → γ at 3.4, 3.2 and 2.3, respectively. The γ phase is irreversibly transformed into the α-phase at the potential 2.05 V (x > 2). From XRD patterns, it was shown that, upon lithiation from α to ε-phase, the lattice parameter, a, decreases continuously, while the lattice parameter b remains almost unchanged. In situ Raman spectra of Li intercalated V2O5 have been measured on crystal [19,30,105] and thin films [106,107] to characterize structural changes in Li xV2O5 according to different C-rates, but it also to study the abnormal stress modification upon phase transitions [107,108]. The first in situ Raman spectroscopy studies of electrochemically lithiated and delithiated V2O5 materials were reported by Frech et al. [10,30,105]. Findings of Raman patterns of Li xV2O5 electrodes are as follows: (i) a shift of the low-frequency B1g(B2g) Raman peak (translational mode) from 144 to 154 cm<sup>-1</sup>, (ii) a progressive intensity decay of the spectral response in the range 195–715 cm<sup>-1</sup> attributed to the reduction of V 5+ to V 4+ generating negative charge carriers, (iii) the subsequent reduction of the optical-spin depth due to the metallic nature of the Li intercalated V2O5 electrode, (iv) the disappearance of the vanadyl stretching mode at 996 cm<sup>-1</sup> at electrode potential of ~3.27 V vs. Li/Li<sup>+</sup>, replaced by a new peak at 983 cm<sup>-1</sup> observed at ~3.40 V related to the coexistence of the α- and ε-LiV2O5 phases, (v) the new peak at 983 cm<sup>-1</sup> shifts down to 972 cm<sup>-1</sup> for further degree of Li

#### Answer:

The timestep with the highest average A1g charged height is at ts = 0, with an average height of approximately 707.69 (Data: ts=0, (x=0, y=707.692353)). This significant value suggests that at this initial timestep, the state of charge is relatively high, indicating that the battery is likely in a fully charged state or very close to it.

In the context of battery behavior, a high average A1g charged height can be associated with the structural integrity and stability of the electrode materials during the charging process. As noted in the literature, Raman spectral changes are often correlated with the potential plateaus observed during charging reactions, which can indicate phase transitions in the electrode materials (Julien et al., 2018). Specifically, the transition from a dilute stage to a more concentrated stage in lithium-ion batteries can reflect changes in the material's electronic and ionic conductivity, which are crucial for efficient charge

Figure 12: Expert evaluation packet, page 4.

storage and transfer.

Moreover, the presence of current peaks during the charge-discharge cycles, as highlighted in the literature, indicates significant electrochemical activity at specific voltage levels (Hiraoka et al., 2025). These peaks suggest that the battery is actively engaging in charge transfer processes, which aligns with the high average A1g height observed at this timestep.

Overall, the high average A1g charged height at  $t_s = 0$  implies that the battery is in a favorable state for charge storage, potentially leading to enhanced performance and longevity, provided that the structural integrity of the electrode materials is maintained throughout the cycling process (Xiong et al.).

Year: 2025  
 Citation: (Mele et al., 2025)  
 Source file: Raman Spectroscopy of Practical LIB Cathodes- A Study of Humidity-Induced Degradation.pdf  
 Passage: Molecules 2025, 30, 3448 7 of 13 shift in the G band due to stiffening of the C-C bonds, corresponding to an alteration of the electronic structure of the carbon [41]. Finally, SEI formation has been reported to lead to changes in the local bonding environment of the carbon surface, that yields blue shifts [42]. In the Raman shift range corresponding to the PO 43- stretching modes (Figure 4c), the R-LFP electrodes exhibit the typical spectral pattern of LFP. Conversely, the H-LFP sample displays, in addition to the band at 993 cm<sup>-1</sup>, indicative of LFP lithiation, a shoulder at 957 cm<sup>-1</sup>, corresponding to the symmetric PO 43- stretching vibration, characteristic of delithiated LFP [43]. Finally, in the O-P-O bending region (Figure 4e), a weak band at 441 cm<sup>-1</sup> appears in the spectral pattern of the water-exposed electrode, which is absent in the spectra of pristine LFP, that is attributable to imperfect lithiation [44]. Figure 4. (a) Raman spectra of pristine (R-LFP, blue curves) and humidity-damaged (H-LFP, red curves) LFP electrodes. Details of (b) the carbon bands; (c) the PO43- stretching bands; (d) the O-P-O bending modes and vibrations related to Li- and Fe-related bonds. SEM micrographs of pristine (R-LFP) and humidity-damaged (H-LFP) LFP cathodes are reported in panels (a-f) of Figure 5. The low-magnification micrographs of panels (a) and (d) highlight the formation of damaging patterns at the H-LFP mesoscopic scale. Moreover, images at a higher magnification (panels (b), (c), (e) and (f)) highlight that exposure to humidity leads to a less defined micro texture of the LFP clusters. Specifically, H-LFP (panels (e) and (f)) exhibits an etched morphology characterized by looser and spheroidized particles, suggesting some degree of corrosive attack and the formation of a surface layer, in-keeping with the literature [16,45], that reports relative LFP tolerance to water traces in the electrolyte. These features are the morphological counterpart of the structural alterations highlighted by Raman spectroscopy [46–48].

[3] Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries (similarity: 0.223)

Authors: Koji Hiraoka et al.  
 Journal: Journal of Materials Chemistry A  
 Year: 2025  
 Citation: (Hiraoka et al., 2025)  
 Source file: Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries.pdf  
 Passage: process. We also calculated the changes in the peak positions of the D- and G-bands from the curve fitting analysis, as shown in Fig. 4(c) and (d). At both measurement points, the position of the D-band, especially at point 2, increased up to approximately 1.5 V, then slightly decreased up to approximately 2.4 V, and finally increased further during the charge process, while the position increased continuously during the discharge process. The position of the G-band slightly decreased continuously below approximately 3.0 V and then decreased considerably at a higher voltage during the charge process, whereas a substantial increase in all voltage regions was observed during the discharge process. The gradual changes (especially for the G-band position) should correspond to the partial intercalation/deintercalation reaction of Li+. There is a possibility that both measurement points changed continuously during the charge-discharge process because of the volume changes associated with the deposition and dissolution of metallic Li. However, the changes in aspect ratios,

Figure 13: Expert evaluation packet, page 5.

Figure 15: Expert evaluation packet, page 7.

Query #2

Question:  
 Which timestep has the highest D/G ratio, and what does that imply about carbon disorder?

```
SQL Query:
SELECT s.ts, CASE WHEN g.height > 0 THEN d.height / g.height ELSE NULL END AS
id_over_ig FROM samples s JOIN peaks d ON s.id = d.sample_id AND d.family = 'd'
JOIN peaks g ON s.id = g.sample_id AND g.family = 'g' ORDER BY id_over_ig DESC
LIMIT 1;
```

SQL Results:

| ts  | id_over_ig       |
|-----|------------------|
| 5.0 | 29.2726068105224 |

Literature Results:  
 [1] Raman Spectroscopy of Practical LIB Cathodes- A Study of Humidity-Induced Degradation (similarity: 0.273)

Authors: Claudio Mele et al.  
 Journal: Molecules  
 Year: 2025  
 Citation: (Mele et al., 2025)  
 Source file: Raman Spectroscopy of Practical LIB Cathodes- A Study of Humidity-Induced Degradation.pdf  
 Passage: Similar comments apply to band ratio modifications: in particular, variations in the D/G intensity ratio denote differences in structural disorder or defect concentration within the carbon matrix, that in the literature have been referred to as different types of damaging modes [40,54–56]. In addition, a broad peak at ca. 1077 cm<sup>-1</sup> is observed only in R-NMC-LMO. This vibrational feature is characteristic of carbonates, like NiCO<sub>3</sub>, Li<sub>2</sub>CO<sub>3</sub>, CoCO<sub>3</sub>, and MnCO<sub>3</sub>, that can be related to CEI formation [4]. The absence of these vibration in the H-NMC-LMO can be explained with their leaching by HF produced in the reaction of the electrolyte with humidity. Panels (c) and (d) of Figure 6 report the fitting of the M-O vibrational band. The band of R-NMC-LMO can be fitted with three peaks centered at 497, 577, and 631 cm<sup>-1</sup>, while that of H-NMC-LMO has three components at 496, 579, and 612 cm<sup>-1</sup>. For the R-NMC-LMO cathode, the peaks at 631 and 577 cm<sup>-1</sup> can be assigned to Mn-O vibrations

[2] Raman Spectroscopy of Practical LIB Cathodes- A Study of Humidity-Induced Degradation (similarity: 0.234)  
 Authors: Claudio Mele et al.  
 Journal: Molecules

Figure 14: Expert evaluation packet, page 6.

Rs, and positions of the D- and G- bands suggest that the structure/regularity of CB (e.g., bond angle/length, bond disorder, and clustering order) was affected by electrochemical reactions, such as the small amount of intercalation/deintercalation in also intercalation processes regardless of the disordered carbon material. A similar peak shift has been reported for a graphite negative electrode active material, which was interpreted as the formation of a stage structure of the Li-graphite intercalated compounds (Li-GICs). 44,48–51 In several studies, the Li-GIC structure in the liquid electrolyte has been investigated, and the phase transition from stage 4 to stage 1 in the high SOC region (low potential regions, Li metal) has been explained by operando Raman spectroscopies. Here, the stage structure is often described as changing the number of graphene layers that separate Li+ during the intercalation reaction. For instance, in stage 4, Li+ is separated by 4 graphene layers (LC36). By transitioning to stage 1, the G-band begins to shift to a lower wavenumber and split of peak shapes, and then to a broader shape, associated with the twist of the graphene layer and strain by Li+ intercalation. In this study, as similar trends were observed, CB was also expected to intercalate Li and change the structural regularity in graphene layers, resulting in the changes in the R values and peak positions in the D- and G-bands. In addition to the fluorescence effect and changes in the D- and G-bands in the Raman spectra, three additional peaks approximately 1087, 1854, and 3648 cm<sup>-1</sup> (green regions in Fig. 3) appeared during charging, and that at 1854 cm<sup>-1</sup> also remained after the discharging.

[4] Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries (similarity: 0.223)

Authors: Koji Hiraoka et al.  
 Journal: Journal of Materials Chemistry A  
 Year: 2025  
 Citation: (Hiraoka et al., 2025)  
 Source file: Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries.pdf  
 Passage: In this study, as similar trends were observed, CB was also expected to intercalate Li and change the structural regularity in graphene layers, resulting in the changes in the R values and peak positions in the D- and G-bands. In addition to the fluorescence effect and changes in the D- and G-bands in the Raman spectra, three additional peaks approximately 1087, 1854, and 3648 cm<sup>-1</sup> (green regions in Fig. 3) appeared during charging, and that at 1854 cm<sup>-1</sup> also remained after the discharging. In a previous report, transmission electron microscope and electron diffraction pattern observations indicated that an Ag-Li alloying reaction with charging occurred in the Ag/C layer of an AF-ASSB system with the same cell configuration. Therefore, the Ag-Li alloying reaction also occurred in this AF-ASSB. This charging mechanism may affect the obtained Raman spectra, such as the observed additional peaks. Microscopic images and Raman spectra of the cross section of the AF-ASSB were acquired after short-circuit (–3.9 V) and in the discharged state. Fig. 5(a) shows cross-sectional images of the AF-ASSB in the pristine state, at 3.9 V (after short-circuit), and at 2.1 V (discharged) and Raman spectra of the AF-ASSB in these states obtained by operando Raman spectroscopy. The cross-sectional image of the SE layer clearly exhibited line-like cracks and small precipitates with metallic luster close to the Ag/C layer after the short-circuit; these features mostly remained after the discharge process. Moreover, the microscopic image revealed that the PE layer had a slight metallic luster after a short-circuit compared to that of the pristine state. We consider that these cracks and precipitates induced the short-circuit in the case with the cross-section because the laminate-type cell before the cut exhibited stable charge-discharge behavior and cycle

Figure 16: Expert evaluation packet, page 8.

capability (Fig. 2(b)). The Raman spectra acquired from the precipitate area also exhibited a peak of around 1848  $\text{cm}^{-1}$  (the green region in Fig. 5(b)), which was not observed in a pristine state. Since this peak appears at a wavenumber similar to the new peak that formed during the charge process (Fig. 3), it is attributed to some Raman active species induced from the precipitate with a metallic luster. In general, metals and alloys hardly exhibit Raman scattering owing to the dominance of Rayleigh scattering and high back-ground signals. However, since the formed precipitates were predicted to be small similar to the laser spot size, the obtained Raman spectra were expected to include both information on Fig.

[5] Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries (similarity: 0.223)

Authors: Koji Hiraoka et al.

Journal: Journal of Materials Chemistry A

Year: 2025

Citation: (Hiraoka et al., 2025)

Source file: Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries.pdf

Passage: 2(b)). The Raman spectra acquired from the precipitate area also exhibited a peak of around 1848  $\text{cm}^{-1}$  (the green region in Fig. 5(b)), which was not observed in a pristine state. Since this peak appears at a wavenumber similar to the new peak that formed during the charge process (Fig. 3), it is attributed to some Raman active species induced from the precipitate with a metallic luster. In general, metals and alloys hardly exhibit Raman scattering owing to the dominance of Rayleigh scattering and high back-ground signals. However, since the formed precipitates were predicted to be small similar to the laser spot size, the obtained Raman spectra were expected to include both information on Fig. 5 Microscopic images of the cross-sectional AF-ASSB at pristine, 3.9 V (charge) and 2.1 V (discharge) states during operando Raman spectroscopy (a), and their Raman spectra (b) acquired from the observed precipitates in the SE layer after the short-circuit and at points 1 and 2 during potential sweep. This journal is © The Royal Society of Chemistry 2025. J. Mater. Chem. A, 2025, 13, 27960–27969 | 27965 Paper. Journal of Materials Chemistry A Published on 12 June 2025. Downloaded by Stanford University Libraries on 11/4/2025 9:37:45 PM. View Article Online

Answer:

The timestep with the highest D/G ratio is at  $t_s = 5$ , with a value of  $\text{id\_over\_ig} = 29.272661$ . This elevated D/G ratio indicates a significant level of structural disorder or defect concentration within the carbon matrix.

In Raman spectroscopy, the D band is associated with defects and disorder in the carbon structure, while the G band corresponds to the stretching of C-C bonds in graphitic materials. A high D/G ratio suggests that there are more defects present relative to the ordered carbon structure, which can imply a degradation of the material's structural integrity (Mele et al., 2025).

The implications of a high D/G ratio are critical for understanding the performance of

Figure 17: Expert evaluation packet, page 9.

carbon-based materials, particularly in applications such as lithium-ion batteries. Increased disorder can affect the electronic properties of the carbon, potentially leading to reduced conductivity and efficiency in energy storage systems (Hiraoka et al., 2025).

Moreover, the presence of defects can facilitate lithium intercalation, which may alter the structural regularity of the carbon layers, further impacting the material's performance (Hiraoka et al., 2025). Therefore, the high D/G ratio at this timestep suggests that the carbon material may be experiencing significant changes that could affect its functionality in practical applications.

Figure 18: Expert evaluation packet, page 10.

## C Literature Corpus: Full Citation List

- 956
- 957
- 958
- 959
- 960
- 961
- 962
- 963
- 964
- 965
- 966
- 967
- 968
- 969
- 970
- 971
- 972
- 973
- 974
- 975
- 976
- 977
- 978
- 979
- 980
- 981
- 982
- 983
- 984
- 985
- 986
- 987
- 988
- 989
- 990
- 991
- 992
- 993
- 994
- 995
- 996
- 997
- 998
- 999
- 1000
- Koji Hiraoka, Yoshiki Yokoyama, Sarina Mine, Kazuo Yamamoto, and Shiro Seki. 2025. Advanced Raman spectroscopy for battery applications: Materials characterization and operando measurements. *APL Energy*, 3(2).
  - Jiahui Xiong, Soroosh Mahmoodi, Zhimin Huang, and Shengwen Zhong. Capacity Decay Mechanism for Lithium-Rich Layered Oxides Under 2.0–4.5V; Speculation and Experimental Study. *SSRN* (preprint). DOI: 10.2139/ssrn.4631123.
  - Eibar Flores, Petr Novák, and Erik J. Berg. 2018. In situ and operando Raman spectroscopy of layered transition metal oxides for Li-ion battery cathodes. *Frontiers in Energy Research*, 6:82.
  - Christian M. Julien and Alain Mauger. 2018. In situ Raman analyses of electrode materials for Li-ion batteries. *AIMS Materials Science*, 5(4):650–698.
  - Jinglei Lei, Frank McLarnon, and Robert Kostecki. 2005. In situ Raman microscopy of individual  $\text{LiNi}_{0.8}\text{Co}_{0.15}\text{Al}_{0.05}\text{O}_2$  particles in a Li-ion battery composite cathode. *The Journal of Physical Chemistry B*, 109(2):952–957.
  - M. A. Cabañero, Markus Hagen, and E. Quiroga-González. 2021. In-operando Raman study of lithium plating on graphite electrodes of lithium ion batteries. *Electrochimica Acta*, 374:137487.
  - Marco-Tulio Fonseca Rodrigues, Victor A. Maroni, David J. Gosztola, Koffi P. C. Yao, Kaushik Kalaga, Ilya A. Shkrob, and Daniel P. Abraham. 2018. Lithium acetylide: A spectroscopic marker for lithium deposition during fast charging of Li-ion cells. *ACS Applied Energy Materials*, 2(1):873–881.
  - Koji Hiraoka, Junichi Sakabe, Naoki Suzuki, and Shiro Seki. 2025. Operando and ex situ Raman spectroscopies for evaluating carbon structural changes in anode-free-type sulfide-based all-solid-state Li-ion batteries. *Journal of Materials Chemistry A*.
  - Rose E. Ruther, Andrew F. Callender, Hui Zhou, Surendra K. Martha, and Jagjit Nanda. 2014. Raman microscopy of lithium-manganese-rich transition metal oxide cathodes. *Journal of The Electrochemical Society*, 162(1):A98.
  - Claudio Mele, Filippo Ravasio, Andrea Casalegno, Elisa Emanuele, Claudio Rabbisi, and Benedetto Bozzini. 2025. Raman Spectroscopy of Practical LIB Cathodes: A Study of Humidity-Induced Degradation. *Molecules*, 30(16):3448. DOI: 10.3390/molecules30163448.
  - Eibar Flores, Petr Novak, Ulrich Aschauer, and Erik J. Berg. 2019. Cation ordering and redox chemistry of layered Ni-rich  $\text{Li}_x\text{Ni}_{1-2y}\text{Co}_y\text{Mn}_y\text{O}_2$ : an operando Raman spectroscopy study. *Chemistry of Materials*, 32(1):186–194.
  - Vijay A. Sethuraman, Laurence J. Hardwick, Venkat Srinivasan, and Robert Kostecki. 2010. Surface structural disordering in graphite upon lithium intercalation/deintercalation. *Journal of Power Sources*, 195(11):3655–3660. DOI: 10.1016/j.jpowsour.2009.12.034.
  - Ermanno Miele, Wesley M. Dose, Ilya Manyakin, Michael H. Frosz, Zachary Ruff, Michael F. L. De Volder, Clare P. Grey, Jeremy J. Baumberg, and Tijmen G. Euser. 2022. Hollow-core optical fibre sensors for operando Raman spectroscopy investigation of Li-ion battery liquid electrolytes. *Nature Communications*, 13(1). DOI: 10.1038/s41467-022-29330-4.
  - Adrian Lindner, Hannes Radinger, Frieder Scheiba, and Helmut Ehrenberg. 2022. Structure–activity correlation of thermally activated graphite electrodes for vanadium flow batteries. *RSC Advances*, 12(22):14119–14126. DOI: 10.1039/D2RA02368G.
  - Gozde Oney, Federico Monaco, Saptarshree Mitra, Asma Medjahed, Manfred Burghammer, Dmitry Karpov, Marta Mirolo, Jakub Drnec, Isabelle C. Jolivet, Quentin Arnoux, Samuel Tardif, Quentin Jacquet, and Sandrine Lyonard. 2025. Dead, Slow, and Overworked Graphite: Operando X-Ray Microdiffraction Mapping of Aged Electrodes.

1049 *Advanced Energy Materials*, 15(38). DOI: 10.1002/aenm.202502032. 1097

1050

1051 • Dominika A. Buchberger, Bartosz 1100

1052 Hamankiewicz, Monika Michalska, Al- 1101

1053 icja Głaszczka, and Andrzej Czerwinski. 1102

1054 2024. Ex Situ Raman Mapping of  $\text{LiMn}_2\text{O}_4$  1103

1055 Electrodes Cycled in Lithium-Ion Batteries. 1104

1056 *ACS Omega*, 9(28):30381–30391. DOI: 1105

1057 10.1021/acsomega.4c01480.

1058 • Jun-Wei Yin, Yi-Meng Wu, Xin-Yu Liu, Jing 1106

1059 Li, Peng-Fei Wang, Zong-Lin Liu, Lin-Lin 1107

1060 Wang, Jie Shu, and Ting-Feng Yi. 2025. In- 1108

1061 sights into degradation mechanisms and en- 1109

1062 gineering strategies of layered manganese- 1110

1063 based oxide cathodes for sodium-ion bat- 1111

1064 tery. *Green Energy & Environment*. DOI: 1112

1065 10.1016/j.gee.2025.07.013.

1066 • Dominika A. Buchberger, Maciej Boczar, 1113

1067 Jacek B. Jasinski, and Andrzej Czer- 1114

1068 wiński. 2025. Raman spectroscopy com- 1115

1069 plemented with XRD and TEM for study- 1116

1070 ing structural evolution in initial cycles of 1117

1071  $\text{LiNi}_{1/3}\text{Mn}_{1/3}\text{Co}_{1/3}\text{O}_2$  cathode material. *Dis-* 1118

1072 *cover Nano*, 20(1). DOI: 10.1186/s11671- 1119

1073 025-04319-2.

1074 • Marcel Heber, Kathrin Hofmann, and Chris- 1120

1075 tian Hess. 2022. Raman Diagnostics of 1121

1076 Cathode Materials for Li-Ion Batteries Us- 1122

1077 ing Multi-Wavelength Excitation. *Batteries*, 1123

1078 8(2):10. DOI: 10.3390/batteries8020010.

1079 • Ruichuan Yuan, Yiwen Guo, Ilke Gurgan, 1124

1080 Nahian Siddique, Yu-Sheng Li, Seokhoon 1125

1081 Jang, Gina A. Noh, and Seong H. Kim. 2025. 1126

1082 Raman spectroscopy analysis of disordered 1127

1083 and amorphous carbon materials: A review of 1128

1084 empirical correlations. *Carbon*, 238:120214. 1129

1085 DOI: 10.1016/j.carbon.2025.120214.

1086 • Tianxun Cai, Mingzhi Cai, Jinxiao Mu, Siwei 1130

1087 Zhao, Hui Bi, Wei Zhao, Wujie Dong, and 1131

1088 Fuqiang Huang. 2023. High-Entropy Layered 1132

1089 Oxide Cathode Enabling High-Rate for Solid- 1133

1090 State Sodium-Ion Batteries. *Nano-Micro Let-* 1134

1091 *ters*, 16(1). DOI: 10.1007/s40820-023-01232- 1135

1092 0.

1093 • Xinyu Liu, Jaehoon Choi, Zhen Xu, Clare P. 1136

1094 Grey, Simon Fleischmann, and Alexander C. 1137

1095 Forse. 2024. Raman Spectroscopy Measure- 1138

1096 ments Support Disorder-Driven Capacitance 1139

in Nanoporous Carbons. *Journal of the Ameri-* 1140

*can Chemical Society*, 146(45):30748–30752. 1141

DOI: 10.1021/jacs.4c10214. 1142

• Sven Jovanovic, Peter Jakes, Steffen Merz, 1143

Rüdiger-A. Eichel, and Josef Granwehr. 2021. 1144

Lithium intercalation into graphite: In 1145

operando analysis of Raman signal widths. 1146

*Electrochemical Science Advances*, 2(4). DOI: 1147

10.1002/elsa.202100068.

• M. J. Madito. 2025. Revisiting the 1148

Raman disorder band in graphene-based 1149

materials: A critical review. *Vibra-* 1150

*tional Spectroscopy*, 139:103814. DOI: 1151

10.1016/j.vibspec.2025.103814.

• Debbie Zhuang and Martin Z. Bazant. 2022. 1152

Theory of Layered-Oxide Cathode Degrada- 1153

tion in Li-ion Batteries by Oxidation-Induced 1154

Cation Disorder. *Journal of The Electro-* 1155

*chemical Society*, 169(10):100536. DOI: 1156

10.1149/1945-7111/ac9a09.

• Yasutaka Matsuda, Naoaki Kuwata, Tat- 1157

sunori Okawa, Arunkumar Dorai, Osamu 1158

Kamishima, and Junichi Kawamura. 2019. In 1159

situ Raman spectroscopy of  $\text{LiCoO}_2$  cathode 1160

in  $\text{Li/Li}_3\text{PO}_4/\text{LiCoO}_2$  all-solid-state thin-film 1161

lithium battery. *Solid State Ionics*, 335:7–14. 1162

DOI: 10.1016/j.ssi.2019.02.010.

• Lukas Karapınar-Springorum, Asia Sarycheva, 1163

Andrew Dopilka, Hyungyeon Cha, Muham- 1164

ammad Ihsan-UI-Haq, Jonathan M. Larson, and 1165

Robert Kostecki. 2025. An infrared, Ra- 1166

man, and X-ray database of battery interphase 1167

components. *Scientific Data*, 12(1). DOI: 1168

10.1038/s41597-024-04236-6. 1169

• R. Hausbrand, G. Cherkashinin, H. Ehren- 1170

berg, M. Gröting, M. Albe, C. Hess, and 1171

W. Jaegermann. 2015. Fundamental degra- 1172

dation mechanisms of layered oxide Li-ion 1173

battery cathode materials: Methodology, in- 1174

sights and novel approaches. *Materials Sci-* 1175

*ence and Engineering: B*, 192:3–25. DOI: 1176

10.1016/j.mseb.2014.11.014. 1177

• Slaheddine Jabri, Luciana Pitta Bauermann, 1178

and Matthias Vetter. 2023. Raman spectrom- 1179

etry measurements for the 2D mapping of the 1180

degradation products on aged graphite anodes 1181

of cylindrical Li-ion battery cells. *AIP Ad-* 1182

*vances*, 13(11). DOI: 10.1063/5.0171158. 1183

|      |   |      |
|------|---|------|
| 1145 | • Ray H. Baughman, Anvar A. Zakhidov, and   | 1193 |
| 1146 | Walt A. de Heer. 2007. Charge transfer  | 1194 |
| 1147 | in carbon nanotube actuators investigated   | 1195 |
| 1148 | using in situ Raman spectroscopy. <i>Ad-</i>  | 1196 |
| 1149 | <i>vanced Materials</i> , 19(19):3213–3218. DOI:  | 1197 |
| 1150 | 10.1002/adma.200602660.   |      |
| 1151 | • Yong Soo Cho, Younghwa Kim, Jae-  | 1198 |
| 1152 | jin Kim, and others. 2019. Detec-   | 1199 |
| 1153 | tion of secondary phase in NMC811(OH) <sub>2</sub>  | 1200 |
| 1154 | precursor using X-ray powder diffraction  | 1201 |
| 1155 | and Raman spectroscopy. <i>Ceramics In-</i>   | 1202 |
| 1156 | <i>ternational</i> , 45(18):24036–24042. DOI:   | 1203 |
| 1157 | 10.1016/j.ceramint.2019.08.246.   |      |
| 1158 | • Robert J. Nemanich and S. A. Solin. 1979.   | 1204 |
| 1159 | Distinguishing disorder-induced bands from  | 1205 |
| 1160 | allowed Raman bands in graphite. <i>Phys-</i>   | 1206 |
| 1161 | <i>ical Review B</i> , 20(2):392–401. DOI:  | 1207 |
| 1162 | 10.1103/PhysRevB.20.392.  | 1208 |
| 1163 | • Rodolfo P. Vidano and David B. Fischbach.   | 1209 |
| 1164 | 1981. Fundamentals, overtones, and combi-   | 1210 |
| 1165 | nations in the Raman spectrum of graphite.  | 1211 |
| 1166 | <i>Solid State Communications</i> , 39(2):341–344.  | 1212 |
| 1167 | DOI: 10.1016/0038-1098(81)90686-4.  | 1213 |
| 1168 | • Hannes Radinger, Martin Petz, Helmut Ehren-   | 1214 |
| 1169 | berg, and others. 2022. Hierarchical structur-  | 1215 |
| 1170 | ing of NMC111-cathode materials in lithium-   | 1216 |
| 1171 | ion batteries: An in-depth study on the in-   | 1217 |
| 1172 | fluence of primary and secondary particle   | 1218 |
| 1173 | sizes on electrochemical performance. <i>Jour-</i>  | 1219 |
| 1174 | <i>nal of Power Sources</i> , 528:231210. DOI:  | 1220 |
| 1175 | 10.1016/j.jpowsour.2022.231210.   | 1221 |
| 1176 | • Yihan Zhu, Shanthi Murali, Wei Cai, and oth-  | 1222 |
| 1177 | ers. 2017. In situ electrochemical Raman  | 1223 |
| 1178 | investigation of charge storage in rGO and  | 1224 |
| 1179 | N-doped rGO. <i>Carbon</i> , 123:424–432. DOI:  | 1225 |
| 1180 | 10.1016/j.carbon.2017.07.065.   |      |
| 1181 | • Arne Sadezky, Heinrich Muckenhuber, Heiko   | 1226 |
| 1182 | Grothe, Reinhard Niessner, and Ulrich   | 1227 |
| 1183 | Pöschl. 2005. Raman spectroscopic   | 1228 |
| 1184 | investigations of activated carbon materi-  | 1229 |
| 1185 | als. <i>Carbon</i> , 43(8):1731–1742. DOI:  | 1230 |
| 1186 | 10.1016/j.carbon.2005.02.018.   | 1231 |
| 1187 | • Naoaki Kuwata, Junichi Kawamura, and  | 1232 |
| 1188 | others. 2017. Raman spectroscopy for  | 1233 |
| 1189 | LiNi <sub>1/3</sub> Mn <sub>1/3</sub> Co <sub>1/3</sub> O <sub>2</sub> composite positive             | 1234 |
| 1190 | electrodes in all-solid-state lithium bat-  | 1235 |
| 1191 | teries. <i>Solid State Ionics</i> , 304:71–76. DOI:   | 1236 |
| 1192 | 10.1016/j.ssi.2016.11.016.  | 1237 |
|      | • Mildred S. Dresselhaus, Ado Jorio, and Ri-  | 1238 |
|      | ichiro Saito. 2005. Raman spectroscopy of   | 1239 |
|      | carbon materials: structural basis of observed  |      |
|      | spectra. <i>Physics Reports</i> , 409(2):47–99. DOI:  |      |
|      | 10.1016/j.physrep.2004.10.006.  |      |
|      | • Xing Zhang and others. 2020. Sur-   |      |
|      | face changes of LiNi <sub>x</sub> Mn <sub>y</sub> Co <sub>1-x-y</sub> O <sub>2</sub> in Li-           |      |
|      | ion batteries using in situ surface-enhanced  |      |
|      | Raman spectroscopy. <i>Journal of Materi-</i>   |      |
|      | <i>als Chemistry A</i> , 8(9):4560–4569. DOI:   |      |
|      | 10.1039/C9TA13634A.   |      |
|      | • Wei Li and others. 2017. The carbon-  |      |
|      | -based 3D-hierarchical cathode architecture   |      |
|      | for Li-ion batteries. <i>Advanced Func-</i>   |      |
|      | <i>tional Materials</i> , 27(15):1606439. DOI:  |      |
|      | 10.1002/adfm.201606439.   |      |
|      | • Steven Sloop and others. 2021. Operando   |      |
|      | Raman spectroscopic analysis for electrolyte/-  |      |
|      | electrode interface reaction in lithium–sulfur  |      |
|      | batteries with sparingly solvating electrolyte.   |      |
|      | <i>Journal of Power Sources</i> , 507:230282. DOI:  |      |
|      | 10.1016/j.jpowsour.2021.230282.   |      |
|      | • Bruno Scrosati and others. 2013. Exploring  |      |
|      | carbon electrode parameters in Li–O <sub>2</sub> cells:   |      |
|      | Li <sub>2</sub> O <sub>2</sub> and Li <sub>2</sub> CO <sub>3</sub> formation. <i>Energy &amp; En-</i> |      |
|      | <i>vironmental Science</i> , 6(5):1440–1445. DOI:   |      |
|      | 10.1039/C3EE24284G.   |      |
|      | • Sungho Park and others. 2019. Charac-   |      |
|      | terization of pitch carbon coating proper-  |      |
|      | ties affecting the electrochemical behavior   |      |
|      | of silicon nanoparticle lithium-ion battery an-   |      |
|      | odes. <i>Electrochimica Acta</i> , 320:134557. DOI:   |      |
|      | 10.1016/j.electacta.2019.134557.  |      |
|      | • Masashi Inaba and others. 2016. In situ   |      |
|      | surface-enhanced Raman spectroelectrochem-  |      |
|      | istry reveals the molecular conformation of   |      |
|      | electrolyte additives in Li-ion batteries. <i>Jour-</i>   |      |
|      | <i>nal of Power Sources</i> , 307:504–512. DOI:   |      |
|      | 10.1016/j.jpowsour.2015.12.060.   |      |
|      | • M. Matsui and others. 2019. In  |      |
|      | operando Raman microscopy of the  |      |
|      | Cu/Li <sub>1.5</sub> Al <sub>0.5</sub> Ge <sub>1.5</sub> (PO <sub>4</sub> ) <sub>3</sub> solid elec-  |      |
|      | trolyte interphase. <i>Journal of Materials</i>   |      |
|      | <i>Chemistry A</i> , 7(27):16157–16166. DOI:  |      |
|      | 10.1039/C9TA03689G.   |      |
|      | • Markus Hagen and others. 2020. Operando   |      |
|      | Raman spectroscopy for investigating  |      |

- 1240 lithium deposition/dissolution and dif-  
1241 fusion at the microelectrode surface.  
1242 *Electrochimica Acta*, 354:136744. DOI:  
1243 10.1016/j.electacta.2020.136744.
- 1244 • Yutaka Yamada and others. 2021. In operando  
1245 Raman spectroscopy reveals Li-ion solvation  
1246 in lithium metal batteries. *Nature Communi-*  
1247 *cations*, 12(1):5403. DOI: 10.1038/s41467-  
1248 021-25669-2.
- 1249 • Walter Van Schalkwijk and Bruno Scrosati  
1250 (editors). 2002. *Materials for Lithium-*  
1251 *Ion Batteries*. Springer, Boston. DOI:  
1252 10.1007/978-1-4419-9269-3.

## D Expert Evaluation Form

Expert reviewers evaluated each SpectraQuery response using an ordinal 5-point Likert scale, where **1 = Strongly Disagree / Poor** and **5 = Strongly Agree / Excellent**. For each query–answer pair, reviewers provided ratings for the following criteria, along with an optional free-response comment.

1. **Scientific Accuracy:** “The response is scientifically correct and free of factual or interpretive errors.”
2. **Grounding in Evidence:** “The response is well-supported by both Raman data and cited literature excerpts.”
3. **Relevance to the Question:** “The response directly answers the specific scientific question posed.”
4. **Clarity and Readability:** “The explanation is clearly written, logically organized, and easy to follow.”
5. **Depth and Insight:** “The response demonstrates meaningful scientific reasoning or new insight beyond surface-level description.”
6. **Completeness:** “The response includes all key aspects or considerations relevant to the question.”
7. **Interpretability of Citations / References:** “I can easily trace the cited figures, tables, or literature passages to verify claims.”
8. **Usefulness for Expert Workflow:** “This output would meaningfully aid my own data analysis or interpretation.”
9. **Optional free response:** What did you like or dislike about this answer?