ISSN : 2311-5122

7universum.com

UNIVERSUM: ТЕХНИЧЕСКИЕ НАУКИ

9(150)

НАУЧНЫЙ ЖУРНАЛ

са́ндрович Попо́в (1821, Санкт-Петербург — Петербург) — русский флотоводец, роитель, адмирал (1891 год). Его имя носит Попова рядом с Владивостоком.
Родился 22 сентября (4 октября) 1821 года в Санкт-Петербурге в семье известного кораблестроителя, управляющего Охтинской верфью Александра Андреевича Попова.

С 1830 года учился на морском отделение Александровского корпуса; в 1838 году окончил Морской кадетский корпус с присвоением чина мичмана и зачислением в 32-й флотский экипаж (Черноморский флот).

был командиром парохода «Метеор» (вспомогательного крейсера Черноморского флота). В 1853 году был командирован в Константинополь для сбора сведений о вооружении Босфора и близлежащих к нему укреплённых мест по Чёрному морю и Дунаю до Рущука. Участвовал в Крымской кампании. В 1853—1854 годах командовал пароходами «Эльборус», «Андия», «Турок», потопил шесть турецких транспортов; в сентябре 1854 года командовал в чине капитан-лейтенанта пароходом «Тамань», прорвался из блокированного Севастополя в Одессу и вернулся обратно с грузом для оборонявшихся.

Руководил установкой морских орудий на укреплениях Севастополя и артиллерийским снабжением. Руководил снаряжением двух брандеров. Руководил созданием бона для преграждения севастопольского рейда. За боевые заслуги в этот период — награждён двумя орденами.

Произведен в капитаны 1-го ранга; с 1855 — флигель-адъютант. В 1855 году был командирован с театра военных действий в Кронштадт, Выборг, Свеаборг и другие укреплённые порты побережья Балтийского моря. В период 1856—1858 годов был начальником штаба Кронштадтского порта.

В 1858—1861 гг. — командующий отрядом из двух корветов («Рында» и «Гридень») и клипера «Опричник» (2-й Амурский отряд), приведя его из Кронштадта в Японское море, затем ходил у берегов Японии, исследовал побережья русского Приморья (один из заливов — у 45-й параллели — был назван в честь корвета Попова «Рында»). В 1859 году корабли под его командованием нанесли визит в порт Сан-Франциско.

В 1861 году был произведён в контр-адмиралы назначением в Свиту[2]. Избран действит Кораблестроительного и Морского уче ведал переделкой парусных судов Совершил плавание к берегам 1861,года был назначен Отрядом судов Балтийс (Первая независима капитана 1 ран январе 1862 Цусим

205 ЛЕТ СО ДНЯ РОЖДЕНИЯ
21 сентября 2026 г.

ПОПОВ
АНДРЕЙ АЛЕКСАНДРОВИЧ
1821-1898 гг.

Г. МОСКВА

РУССКИЙ ФЛОТОВОДЕЦ, КОРАБЛЕСТРОИТЕЛЬ, АДМИРАЛ

UNIVERSUM: ТЕХНИЧЕСКИЕ НАУКИ

Научный журнал

Издается ежемесячно с ноября 2013 года

Выпуск: 9(150)

Сентябрь 2026 г.

Часть 7

Новосибирск
2026

ББК 3
U55

U55 Universum: технические науки : электронный научный журнал. — № 9(150). — Часть 7. — Новосибирск : Изд-во «Юниверсум», 2026. — 96 с. — Текст : электронный. — URL: https://7universum.com/ru/tech/archive/category/9150 (дата публикации: 28.09.2026).

ISSN 2311-5122
DOI: 10.32743/UniTech.2026.150.9
Учредитель и издатель: ООО «Юниверсум»

Содержание

ARTICLES IN ENGLISH

INFORMATION AND DIGITAL TECHNOLOGIES

ARTIFICIAL INTELLIGENCE AND INTELLIGENT SYSTEMS

Статья поступила в редакцию: 03.08.2026
Статья принята к публикации: 02.09.2026
Статья опубликована: 28.09.2026

УДК 004.738.52

Engineering methods for improving search relevance in e-commerce systems based on Elasticsearch and OpenSearch

Perelekhov Ivan
Senior Web Developer,
Oggetto Web LLC,
Taganrog, Russia
ORCID ID: 0009-0008-8448-3092
E-mail: ivanperelekhov@gmail.com

Инженерные методы повышения релевантности поиска в e-commerce системах на основе Elasticsearch и OpenSearch

Перелехов Иван Владимирович
старший веб-разработчик, ООО Оджетто Веб
Российская Федерация, г. Таганрог

Abstract

The article presents an analytical review of engineering approaches to improving search relevance in e-commerce systems based on Elasticsearch and OpenSearch. The aim of the study is to systematise technical solutions applied at the stages of product-data preparation, index design, user-query processing, and result ranking. The review draws on recent scientific publications and industry studies addressing product-data quality, indexing, query interpretation and expansion, spelling correction, ranking mechanisms, user behaviour, and search-system performance. The reviewed approaches are grouped according to their engineering purpose, potential effect on search

Библиографическое описание: Перелехов И.В. Инженерные методы повышения релевантности поиска в e-commerce системах на основе Elasticsearch и OpenSearch // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23371

For citation: Perelekhov I. Engineering methods for improving search relevance in e-commerce systems based on Elasticsearch and OpenSearch // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23371

quality, and principal implementation limitations. The reviewed sources indicate that search relevance can be influenced by attribute normalisation, appropriate field-type selection, analyser configuration, synonym management, structured parameter extraction, fuzzy matching, and controlled query relaxation. The review also highlights a recurring trade-off between recall and precision when query-expansion mechanisms are applied. Textual, attribute-based, behavioural, commercial, contextual, and infrastructure-related signals used in product ranking are systematised. The findings of the review allow relevance improvement to be considered as a set of interconnected engineering tasks performed at different stages of the search process while accounting for data-quality limitations, computational cost, and search-system performance.

Аннотация

В статье представлен аналитический обзор инженерных подходов к повышению релевантности поиска в e-commerce системах на основе Elasticsearch и OpenSearch. Цель исследования состоит в систематизации технических решений, применяемых на этапах подготовки товарных данных, проектирования индекса, обработки пользовательских запросов и ранжирования результатов. Материалом для обзора послужили современные научные публикации и отраслевые исследования, посвященные качеству товарных данных, индексированию, интерпретации и расширению запросов, коррекции опечаток, ранжированию, пользовательскому поведению и производительности поисковых систем. Рассмотренные подходы сгруппированы по их инженерному назначению, потенциальному влиянию на качество поиска и основным ограничениям применения. Показано, что на релевантность поиска могут влиять нормализация атрибутов, выбор типов полей, настройка анализаторов, управление синонимами, извлечение структурированных параметров, нечеткое сопоставление и контролируемое ослабление условий поиска. Обзор также выявляет характерный компромисс между полнотой и точностью выдачи при использовании механизмов расширения запросов. Систематизированы текстовые, атрибутные, поведенческие, коммерческие, контекстные и инфраструктурные сигналы, учитываемые при ранжировании. Результаты обзора позволяют представить повышение релевантности как совокупность взаимосвязанных инженерных задач, решаемых на разных этапах поискового процесса с учетом ограничений по качеству данных, вычислительным затратам и производительности системы.

Keywords: e-commerce, product search, relevance, Elasticsearch, OpenSearch, indexing, ranking.

Ключевые слова: e-commerce, товарный поиск, релевантность, Elasticsearch, OpenSearch, индексация, ранжирование.

Introduction

Search in e-commerce systems is one of the key mechanisms of user interaction with a product catalog. Unlike general information retrieval, product search is aimed not only at identifying textual matches but also at aligning user intent with specific product characteristics, category, attributes, availability, and commercial relevance. Therefore, the quality of search results is directly associated with navigation convenience, catalog browsing depth, bounce rate, and the probability of purchase.

The main challenge in improving product search accuracy is the diversity of product information, user input variability, synonymy, misspellings, vague query formulation, and inconsistencies in naming conventions.

The practical importance of product-data quality is illustrated by Salsify's 2025 consumer research conducted among 1,910 respondents in the USA and UK. The survey indicates that inconsistent or inaccurate product information can affect consumer trust and purchasing decisions (fig. 1).

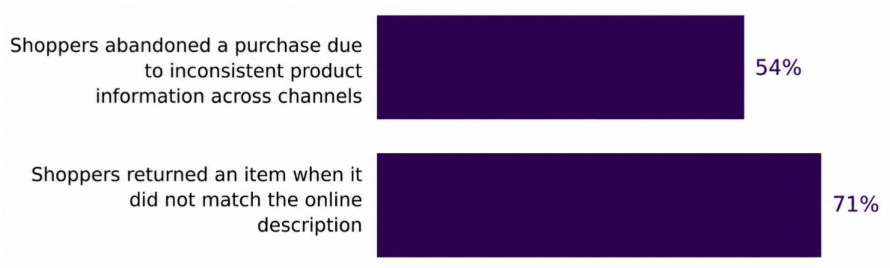

Figure 1. Consumer response to inconsistent and inaccurate product information in e-commerce according to the Salsify survey [1]

These findings support the need to address product-data quality at the early stages of the search pipeline through data normalisation, attribute structuring, and consistent representation of product information before query-processing and ranking mechanisms are applied.

Despite the implementation of state-of-the-art search solutions such as Elasticsearch and OpenSearch, search effectiveness depends on engineering configuration, analysis of catalog structures and query flows, and adaptation of ranking rules to the characteristics of a particular platform.

Recent research on e-commerce search demonstrates a shift from isolated lexical matching techniques towards more complex approaches that integrate semantic processing, user behaviour, and relevance evaluation. Ren et al. [2] systematised information discovery mechanisms in e-commerce and showed that contemporary solutions combine information retrieval, recommender systems, natural language processing, and user interaction data. Nguyen et al. [3] focused on query rewriting and proposed a hybrid framework combining offline knowledge distillation with online reinforcement learning; experiments on the Amazon ESCI dataset demonstrated improvements in query relevance, diversity, and adaptability while addressing the latency and computational costs associated with large generative models. More recent work by Ou et al. [4] extends query rewriting towards alignment with human cognition and user preferences, reflecting the transition from generic query reformulation to preference-aware search optimisation. Jain et al. [5] addressed the evaluation stage by proposing the PROBES framework, which distinguishes exact, substitute, complement, and irrelevant results and identifies both precision and missed-recall errors; its evaluation across 100 product categories achieved an Actionable Error Rate of 76 %.

Overall, the reviewed studies address different components of e-commerce search, including product-data preparation, query processing, ranking, and relevance evaluation. Their combined consideration across the main stages of the product-search process is less common in the reviewed literature.

The aim of this analytical review is to systematise engineering approaches to improving search relevance in e-commerce systems based on Elasticsearch and OpenSearch and to examine their engineering roles and principal limitations at different stages of the product-search pipeline. The review objectives are to summarise the principal sources of relevance loss in heterogeneous product catalogues; examine approaches to data preprocessing, index design, and analyser configuration; systematise query-processing mechanisms and their limitations; and examine how textual, attribute-based, behavioural, commercial, contextual, and infrastructure-related signals are used in product ranking.

Materials and methods

The study was conducted as an analytical review of engineering approaches to improving product-search relevance in e-commerce systems based on Elasticsearch and OpenSearch. The source base included recent scientific publications and industry studies addressing product-data quality and preprocessing, search indexing, query interpretation and expansion, spelling correction, ranking and recommendation mechanisms, user behaviour, and search-system performance.

The reviewed approaches were organised according to the principal stages of the product-search pipeline: product-data preparation, index design, query processing, and ranking of retrieved items. Within each stage, the mechanisms described in the reviewed sources were grouped according to their engineering purpose, potential effect on search quality, and principal implementation limitations. Particular attention was given to the precision-recall trade-off, zero-result and irrelevant-result queries, filtering accuracy, computational cost, and other limitations associated with the use of individual search mechanisms.

The reviewed materials were then descriptively classified and summarised across the identified stages of the search pipeline. Search-quality indicators, including precision, recall, zero-result rate, click-through rate, conversion-related metrics, query latency, and computational cost, were considered as criteria for discussing the reviewed approaches rather than as empirically measured variables. The study did not include benchmarking or experimental comparison on a common dataset. Accordingly, its purpose was not to establish the quantitative superiority of individual methods, but to systematise their engineering functions, potential effects, and limitations in e-commerce search.

Results and discussion

Problems of search relevance in product catalogs

Based on the sources considered in this review, four recurring groups of factors affecting relevance in e-commerce product search can be distinguished: product-data heterogeneity, variability of user queries, semantic mismatch between user language and product descriptions, and commercial constraints. These factors occur at different stages of the search process and are addressed by different engineering mechanisms described in the literature. In the e-commerce environment, the same product may be described through its brand, model, category, functional purpose, technical specification, color, size, material, or consumer use case. Therefore, a user query is rarely a simple set of keywords. It represents a compressed expression of demand that must be matched with both structured and unstructured elements of a product card [6]. If such queries are processed insufficiently, the search system may return results that formally match the query but are practically irrelevant, potentially reducing the usefulness of search results for the user.

One recurring source of relevance problems is the heterogeneity of product data. Information from different sources, such as suppliers, marketplace sellers, inventory management systems, and external databases, can result in inconsistencies in attribute names, measurement units, value formats, categories, and the completeness of product descriptions. These differences create difficulties when creating mappings, analyzers, and normalization rules for Elasticsearch and OpenSearch. If attributes are not separated into distinct fields and normalised during indexing, a larger proportion of product information may have to be processed as unstructured text. In this way, the effectiveness of filters, faceted browsing, and ranking would be negatively impacted, especially for complex searches using structured query elements. Thus, product-data heterogeneity represents a relevance problem that originates before query processing and ranking are performed. However, even a well-structured product catalog does not eliminate relevance loss, since search quality also depends on how the system interprets user input.

The second problem category is connected to the variability of user queries. Users may enter specific names, common categories, descriptions, abbreviations, professional terminology, jargon, transcribed words, and even misspelled entries. Given that the search window still continues to be one of the key gateways to a product catalog, it is necessary for the system to understand both standardized names of products and their unstandardized variations. This problem is confirmed by Baymard Institute’s Search UX 2026 benchmark, which analyzed more than 10,000 assessments across over 170 e-commerce websites and applications. According to the benchmark, 56 % of platforms did not provide sufficient support for users’ search needs, while approximately half of usability test participants preferred to start product search through the search bar rather than through navigation menus [7]. Thus, errors in query processing affect one of the central scenarios of user interaction with a product catalog.

A separate problem is synonymy and the semantic gap between user language and product-card terminology. For example, a customer may search for “running sneakers,” while relevant products are categorized as “training shoes,” or enter “laptop for design,” although suitable items are described through processor, graphics card, and display specifications. In such cases, simple token matching does not reflect actual relevance. Approaches described for addressing this problem include synonym dictionaries, morphological analysis, query expansion, field boosting for categories and attributes, and controlled use of fuzzy search. Without appropriate configuration, relevant items may be excluded from the retrieved set, while unrelated items may receive relatively high scores because of accidental term matches.

Last but not least, in e-commerce, relevance is constrained by the problem that search results that are correct might be useless at the same time. In particular, a product may fully match the user’s query but still have a low rating, be unavailable, or be unsuitable for purchase. Conversely, a product with lower lexical similarity may be more appropriate for the user’s actual purchase intent than an item with a stronger textual match. Thus, relevance in e-commerce extends beyond lexical matching and may require a combination of textual, attribute-based, behavioural, and commercial criteria. Therefore, the problem of relevance for product catalogs arises in the confluence of data quality, query variation, semantics, and business aspects. Accordingly, approaches to improving search quality are not limited to ranking adjustments at the final stage and may also involve product-data preparation, structuring of searchable fields, and index design.

Indexing and data preprocessing as the basis of relevant search results

The reviewed approaches indicate that relevance problems observed at the ranking stage may originate earlier, during product-data preparation and index construction. For this reason, preprocessing and indexing were examined as interconnected components of the search pipeline rather than as isolated technical operations. In e-commerce search, indexing is not only the technical loading of product cards into Elasticsearch or OpenSearch but also an engineering stage of data modelling that can influence the accuracy of search, filtering, and ranking [8]. The product card generally holds heterogeneous data that includes the name of the product, its brand, category, description, price, model, size, color, specifications, and availability. The lack of normalization and separation of these data into appropriate field types can lead to incorrect matching, product exclusion, or irrelevant product placement in search results.

Data preprocessing can influence how accurately catalogue attributes are matched with user queries. In e-commerce systems, this stage is especially important because queries often combine several product features rather than reproduce a full product name. If these features are not represented in the index as searchable and filterable fields, the system becomes overly dependent on textual matching. The main methods of preparing product data for search indexing are presented in table 1.

Table 1.

Methods of preparing product data for search indexing

Data preparation method	Engineering role	Potential effect on search relevance
Text cleaning	Removes markup, noise, service symbols, and redundant fragments.	Can reduce accidental matches and irrelevant results.
Case and spelling normalization	Unifies variants of brands, categories, and attributes.	Can reduce fragmentation of search results and facets.
Unit standardization	Converts values to a single measurement format.	Can improve numeric search and attribute-based filtering.
Attribute extraction	Separates color, size, material, model, and compatibility into fields.	Can improve filtering and category search.
Brand and model normalization	Links alternative spellings and product-line variants.	Can improve consistency of exact search and aggregations.
Product card deduplication	Detects duplicate or near-duplicate product records.	Can reduce information noise in search results.
Data enrichment	Adds tags, categories, synonyms, compatibility, and behavioral features.	Can expand search context and support intent matching.

After data preprocessing, the next step is to define how each product element should be represented in the index. At this stage, the same logical entity may need to be stored differently depending on the search purpose. The name of the product will be used for searching, whereas the brand will be used for matching. The price is required for sorting and filtering by ranges, and the specifications are used in faceted navigation. The use of inappropriate field types may lead to structured data being treated as regular text or vice versa. Therefore, mappings should take into account both the internal structure of the product card and the characteristics used for attribute extraction and query matching [9, 10]. The main indexed field types and their functions in product search are presented in table 2.

Table 2.

Indexed field types and their role in product search

Field group	Data examples	Preferred indexing method	Search function
Full-text fields	Product name, description, short description.	Text fields with language analyzers.	Natural-language product search.
Exact values	Brand, SKU, model, product ID.	Keyword fields with normalizer.	Exact matching, aggregation, filtering.
Category data	Category, subcategory, product type.	Keyword fields or hierarchical structures.	Category search and faceted navigation.
Numeric attributes	Price, screen size, memory, weight, power.	Integer, float, scaled_float.	Range filtering and sorting.
Nested attributes	Product specifications, variation parameters.	Nested or object fields.	Search by complex attribute combinations.
Location and stock data	Delivery region, warehouse, availability.	Keyword, boolean, geo fields if needed.	Regional availability and stock-aware search.
Behavioral signals	Views, clicks, purchases, add-to-cart events.	Numeric or aggregated fields.	Additional ranking based on user behavior.

Configuring analyzers is the next level of index engineering optimization. An analyzer determines how the source text is transformed into searchable tokens. It usually includes character filters, a tokenizer, and token filters. In product search, this makes it possible to convert words to lowercase, remove insignificant symbols, process morphological forms, and apply stemming, lemmatization, n-grams, or synonym dictionaries.

Multi-field usage becomes particularly beneficial when one and the same field is indexable by means of multiple formats. For instance, the name of a product will be available in a full-text field, a keyword field, and an n-gram field for implementing search queries for the purpose of completion. Multi-fields with normalisers for keyword fields can be useful for values that require normalisation without tokenisation, such as brand names, SKU numbers, product codes, categories, colour designations, size indications, and inventory status.

This is particularly important for large marketplaces, where product data is supplied by multiple sellers and providers. In such environments, incomplete attributes, inconsistent categories, and different naming conventions become a systemic issue. Elasticsearch and OpenSearch may not fully compensate for low-quality source content if product data is not validated and normalized before indexing. Therefore, indexing should be integrated into a broader data management process, while well-prepared data and a correctly designed index reduce the load on ranking mechanisms and make search more stable as the catalog grows.

Technical mechanisms for processing user queries

The reviewed query-processing mechanisms illustrate a recurring trade-off between retrieval precision and recall. Mechanisms intended to broaden matching or compensate for query variation include query expansion, spelling correction, fuzzy search, and query relaxation [11, 12]. Recent e-commerce studies also consider learned query rewriting as an approach to adapting user input before product retrieval [13]. These approaches can reduce the number of zero-result queries but may simultaneously increase the probability of retrieving irrelevant products. In e-commerce, this is especially important because queries often contain typos, abbreviations, incomplete product names, colloquial terms, mixed brand and attribute references, and implicit purchase intent. Therefore, query processing must balance recall and precision: excessive expansion may return irrelevant results, while overly strict matching may lead to zero-result searches and lost conversions. The main query processing mechanisms are presented in table 3.

Table 3.

Query processing mechanisms and their potential engineering effects in Elasticsearch and OpenSearch

Query processing mechanism	Technical implementation	Potential engineering effect	Potential risk
Query normalization	Lowercasing, whitespace removal, symbol unification, ASCII-folding.	Can improve consistency of query-index matching.	May remove meaningful symbols in model names or SKUs.
Morphological processing	Language analyzers, stemming, lemmatization, token filters.	Can expand matching across grammatical forms.	May merge semantically close but distinct terms.
Synonym handling	Synonym token filters and controlled synonym dictionaries.	May improve recall for alternative product formulations.	May over-expand queries and mix product categories.
Typo correction	Fuzzy query, edit distance, suggesters.	Can reduce zero-result searches caused by input errors.	May return irrelevant items for short or ambiguous queries.
Autocomplete	Edge n-gram, completion suggester, search-as-you-type fields.	Supports incomplete input and accelerates search.	May increase index size and generate irrelevant suggestions.
Structured attribute parsing	Query parsing and extraction of brand, model, size, color, and capacity.	Converts raw input into text, term, and range conditions.	May produce parsing errors for non-standard product names.
Field boosting	Multi-match query, field boosting, dis_max.	Improves ranking of matches in high-value fields.	May penalize relevant items with incomplete product cards.
Zero-result handling	Relaxation logic, fallback queries, partial filter removal.	Preserves the search scenario when exact matches are unavailable.	May show products that do not satisfy critical constraints.

The mechanisms presented above show that user query processing in e-commerce search cannot be reduced to simply passing the input string to full-text search. In practical implementations, a query may undergo several processing steps, including normalisation, linguistic processing, typo handling, structured-attribute extraction, synonym-based expansion, and field weighting. These mechanisms can be organised as a managed pipeline and evaluated using search logs, frequent queries, zero-result cases, and user-behaviour metrics.

Ranking search results using textual, behavioral, and business factors

For analytical purposes, the ranking mechanisms discussed in the reviewed sources can be grouped into six categories of signals: textual, attribute-based, behavioural, commercial, contextual, and infrastructure-related. Their functions differ: some determine query-product correspondence, whereas others reflect product utility, user behaviour, commercial conditions, personal context, or the computational constraints of the search system. In product search, ranking may account not only for lexical similarity to the query but also for alignment with user intent [14]. In e-commerce contexts, the practical usefulness of a retrieved item may additionally depend on its availability, required attributes, price, rating, and completeness of the product description. In Elasticsearch and OpenSearch, this logic can be implemented through field boosting, multi-match queries, dis_max, function score, script score, rescore, and combinations of full-text conditions with filters. Accordingly, textual relevance may be supplemented with attribute-based, behavioural, commercial, and contextual signals, while field weights should be adjusted according to product category and search scenario (table 4).

Table 4.

Signal groups used in product search ranking

Signal group	Feature examples	Technical implementation	Potential ranking role
Textual signals	Product name, description, category, brand, model.	BM25, multi-match, dis_max, field boosting.	Represent basic textual proximity between the query and product card.
Attribute signals	Color, size, memory capacity, material, technical parameters.	Term, range, nested queries, filters, facets.	Can refine matching by specific product characteristics.
Behavioral signals	Clicks, views, add-to-cart events, purchases.	Aggregated fields, feature store, function score, ML models.	Can reflect observed patterns of user interaction with products.
Commercial signals	Margin, stock availability, delivery time, price, rating.	Script score, rank_feature, business rules.	Can incorporate commercial characteristics into ranking.
Contextual signals	Region, device, user history, seasonality.	Personalized boost rules, separate indices, hybrid ranking.	Can account for the current user or request context.
Infrastructure signals	Response time, cluster load, index size, computation cost.	Rescore window, caching, shard routing, script limits.	Can help manage trade-offs between ranking complexity and system performance.

Behavioural and commercial signals may become particularly relevant when product search extends beyond lexical matching. If products have similar textual relevance, higher positions may be assigned to items with stronger evidence of user interest, such as clicks, add-to-cart events, or purchases. Short-term interaction history can also be incorporated into ranking to account for the buyer’s current search intent [15]. However, these signals require careful interpretation, since high click-through rates may reflect visual appeal, low price, or temporary promotion rather than actual satisfaction. These observations indicate that behavioural signals can be interpreted together with commercial and outcome-related metrics, such as conversion rate, revenue per search, purchase rate, and returns to search results.

Hybrid ranking architectures represent one of the approaches discussed for combining lexical retrieval with additional ranking signals. Elasticsearch and OpenSearch can be used for candidate retrieval, filtering, and primary scoring, while an external machine-learning component performs additional reranking of a limited set of retrieved products. Such architectures can be used to combine lexical and attribute-based matching with behavioural signals, user context, product availability, and other features that are difficult to incorporate into a static scoring formula.

Taken together, the reviewed approaches show that product-search ranking can incorporate textual and attribute-based matching together with behavioural, commercial, and contextual signals. In Elasticsearch and OpenSearch, built-in scoring mechanisms may also be combined with analytics, feature engineering, machine-learning components, and experimental evaluation where these are appropriate to the search scenario.

For the purposes of this review, the discussed approaches can be summarised according to the main stages of the product-search process, each addressing different engineering tasks and limitations. A summary of the reviewed methods, their potential effect on search quality, and their principal risks is presented in table 5.

Table 5.

Summary of engineering methods for improving product-search relevance

Search stage	Main methods	Potential effect	Main limitations
Product data preparation	Normalisation, cleaning, attribute extraction, deduplication.	Potentially more consistent data and more accurate filtering.	Dependence on source-data quality.
Index design	Text, keyword, numeric and nested fields; analysers; multi-fields.	Potential improvement in query-product matching and filtering.	Larger and more complex indices.
User query processing	Synonyms, typo correction, attribute extraction, query relaxation.	Potentially higher recall and fewer zero-result queries.	Possible loss of precision.
Result ranking	BM25, field boosting, function score, behavioural and commercial signals.	Potential improvement in the ordering of retrieved products.	Weight calibration and ranking bias.

Thus, individual mechanisms should not be considered in isolation. Their practical effects may depend on decisions made at the data, index, query-processing, and ranking levels, as well as on trade-offs between search quality and computational cost.

The sources considered in this review suggest that no single Elasticsearch or OpenSearch mechanism is sufficient to address all aspects of product-search relevance. Data normalisation and index design can support more accurate retrieval, whereas synonym expansion, fuzzy matching, and query relaxation may increase recall at the cost of additional irrelevant results. Behavioural and commercial signals can provide additional information for ranking but require careful weighting to avoid excessive preference for popular or promoted products. At scale, these mechanisms may also introduce technical constraints: extensive use of n-grams and enriched fields increases index size, complex scoring and reranking may increase query latency, and behavioural features require regular updating and additional data-processing infrastructure. As an analytical review, the study does not provide experimental validation or a quantitative comparison of the discussed mechanisms on a common dataset. Accordingly, the effects described above should be interpreted as literature-based engineering tendencies and potential trade-offs rather than as experimentally verified performance gains.

Conclusion

The review indicates that search relevance in Elasticsearch- and OpenSearch-based e-commerce systems can be addressed at several interconnected stages of the search pipeline. At the data and indexing level, the key engineering measures include normalisation of product attributes and measurement units, separation of searchable and filterable characteristics into appropriate field types, and the use of suitable analysers and structured fields. These measures can reduce dependence on unstructured textual matching and support more structured filtering and ranking.

At the query-processing and ranking stages, relevance depends on the controlled combination of query normalisation, synonym handling, typo correction, structured attribute extraction, field boosting, and fallback logic for zero-result queries. The reviewed mechanisms illustrate a recurring precision-recall trade-off: broader matching may improve recall but can also increase the number of irrelevant results, whereas stricter filtering may improve precision while increasing the risk of excluding potentially relevant products. Final ranking can therefore combine textual matching with attribute-based, behavioural, commercial, and contextual signals using mechanisms such as multi-match, dis_max, function score, rank_feature, and rescore.

Based on the review, product-data preparation, index design, query transformation, ranking, and operational monitoring can be considered as interconnected stages of the product-search process. The performance of such search configurations can be assessed using retrieval, behavioural, commercial, and infrastructure indicators, including precision, recall, zero-result rate, click-through rate, conversion-related metrics, query latency, and computational cost. The study is limited by its analytical design and does not include an experimental comparison of the identified mechanisms on a common dataset, which defines a direction for further research.

References

1. Salsify. 2025 Consumer Research [Electronic resource]. – URL: https://www.salsify.com/resources/report/2025-consumer-research (дата обращения: 04.08.2026).
2. Ren Z., He X., Yin D. de Rijke M. Information discovery in e-commerce // Foundations and Trends in Information Retrieval. – 2024. – Vol. 18, No. 4–5. – P. 417–690. – DOI: 10.1561/1500000097.
3. Nguyen D.A., Mohan R.K., Yang S., Akash P.S., Chang K.C.-C. MiniELM: A lightweight and adaptive query rewriting framework for e-commerce search optimization // Findings of the Association for Computational Linguistics: ACL. – 2025. – P. 6952–6964. – DOI: 10.18653/v1/2025.findings-acl.363.

4. Ou R., Wang K., Shao J., Zhang T., Huo C. Aligning query rewriting with human cognition and preference in e-commerce search // Proceedings of the ACM Web Conference. – 2026. – P. 7933–7942. – DOI: 10.1145/3774904.3792806.
5. Jain S. DSouza C.A., Rana J.B., Joshi A., Yenigalla P. PROBES: Performance and relevance observation for better search // Proceedings of the 19th Conference of the European Chapter of the Association for Computational Linguistics (Volume 5: Industry Track). – 2026. – Vol. 5. – P. 625–635. – DOI: 10.18653/v1/2026.eacl-industry.48.
6. Zhenyu L., Haoran C., Mingkai W. Personalizing search, navigation, and product discovery experiences in B2C e-commerce using machine learning models // Chronicle of Emerging Scientific Paradigms. – 2025. – No. 10.
7. Baymard Institute. Ecommerce Search UX 2026: 8 Search “Query Types” UX Best Practices [Electronic resource]. – URL: https://baymard.com/blog/ecommerce-search-query-types (дата обращения: 04.08.2026).
8. Sharma C., Kaur A., Datta P., Gulzar Y. Optimizing eCommerce data: effective approaches for data collection, cleansing, and preprocessing // Strategic Innovations of AI and ML for E-Commerce Data Security. IGI Global Scientific Publishing. – 2025.
9. Loughnane R., Liu J., Chen Z., Wang Z., Giroux J., Du T., Schroeder B., Sun W. Explicit attribute extraction in e-commerce search // Proceedings of the Seventh Workshop on e-Commerce and NLP @ LREC-COLING. – 2024. – P. 125–135.
10. Moulton M., Ng Y.-K. Boolean interpretation, matching, and ranking of natural language queries in product selection systems // Discover Computing. – 2024. – Vol. 27. – DOI: 10.1007/s10791-024-09432-x.
11. Allahim A., Cherif A., Imine A. Semantic approaches for query expansion: taxonomy, challenges, and future research directions // PeerJ Computer Science. – 2025. – P. 1–28. – DOI: 10.7717/peerj-cs.2664.
12. Güzel Turhan C. Leveraging large language models for spelling correction in Turkish // PeerJ Computer Science. – 2025. – P. 1–26. – DOI: 10.7717/peerj-cs.2889.
13. Dai A., Zhu Z., Hu H., Tang G., Liu L., Xu S. Enhancing E-Commerce Query Rewriting: A Large Language Model Approach with Domain-Specific Pre-Training and Reinforcement Learning // Proceedings of the 33rd ACM International Conference on Information and Knowledge Management. – 2024. – P. 4439–4445. – DOI: 10.1145/3627673.3680109.
14. Saadany H., Bhosale S., Agrawal S., Kanojia D., Orasan C., Wu Z. Centrality-aware product retrieval and ranking // Proceedings of the. – 2024. – P. 215–224. – DOI: 10.18653/v1/2024.emnlp-industry.17.
15. Sheng T., Muthiah S., Islam A., Feng J. Progressive refinement of e-commerce search ranking based on short-term activities of the buyer // Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval. – 2025. – P. 4370–4374. – DOI: 10.1145/3726302.3731966.

Статья поступила в редакцию: 14.09.2026
Статья принята к публикации: 19.09.2026
Статья опубликована: 28.09.2026

УДК 004.9+004.75+621.3.019.3

DOI: 10.32743/UniTech.2026.150.9.23417

Improving microservice failure diagnosis with llms and observability evidence

Zazhogin Stanislav
Senior Software Engineer, Digital IQ
USA, Saratoga
E-mail: stanislav.zazhogin@digital-iq.com

Улучшение диагностики сбоев микросервисов с использованием больших языковых моделей и данных наблюдаемости

Зажогин Станислав
Старший инженер-программист
Digital IQ
США, г. Саратога

Abstract

Large language model agents can synthesize heterogeneous operational telemetry, but a plausible explanation may still cite the wrong entity, confuse a downstream symptom with the initiating fault, or rely on observations that do not support a valid propagation path. The objective of this study is to develop and evaluate EG-RCA (Evidence Guided Root Cause Analysis), an architecture separating LLM hypothesis selection from deterministic evidence verification. The method normalizes metrics, logs, traces, events, and alerts; ranks observations with semantic, temporal, anomaly, topology, modality, and alert-relevance signals; exposes a bounded set of competing hypotheses to the LLM; and checks the selected explanation against propagation paths, timestamps, cited evidence, and numeric constraints. The evaluation methodology includes a reproducible controlled comparison of five deterministic strategies on 480 incidents in a 15-entity microservice topology, with 12 fault types, six operational categories, and 31,855 observations. Under a matched 12-observation evidence budget, the controlled candidate-ranking mechanism reaches an overall score of 67.5, compared with 60.9 for temporal retrieval and 55.9 for flat retrieval, with comparable estimated prompt contexts of 403–428 tokens. A separate production-cluster evaluation uses 500 logical entities, 2,066 runtime instances, and 24 incident windows containing 2,296,172 observations in total. A frozen LLM-in-the-loop pilot additionally records

---

Библиографическое описание: Зажогин С. Улучшение диагностики сбоев микросервисов с использованием больших языковых моделей и данных наблюдаемости // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23417

For citation: Zazhogin S. Improving Microservice Failure Diagnosis with LLMs and Observability Evidence // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23417

single-pass GPT-5.6 Sol Pro outputs on 12 held-out incidents. In that pilot, Evidence F1 is 75.4 with candidate-aware context and 57.0 with flat retrieval, while localization and fault-type accuracy are unchanged. The production-cluster evaluation complements the controlled study; broader validation across additional clusters and expert-labeled datasets remains necessary.

Аннотация

Агенты на основе больших языковых моделей способны объединять разнородную эксплуатационную телеметрию, однако правдоподобное объяснение может ссылаться на неверный объект, путать последующий симптом с исходной неисправностью или опираться на наблюдения, не подтверждающие допустимый путь распространения сбоя. Цель исследования, разработать и оценить архитектуру EG-RCA (Evidence Guided Root Cause Analysis), отделяющую выбор гипотезы языковой моделью от детерминированной проверки свидетельств. Метод нормализует метрики, журналы, трассировки, события и оповещения; ранжирует наблюдения по семантической, временной, аномальной, топологической и модальной релевантности, а также связи с оповещением; предоставляет модели ограниченный набор конкурирующих гипотез и проверяет выбранное объяснение по путям распространения, временным меткам, использованным свидетельствам и численным ограничениям. Методология оценки включает воспроизводимое контролируемое сравнение пяти детерминированных стратегий на наборе из 480 инцидентов в микросервисной топологии из 15 объектов; набор охватывает 12 типов неисправностей, шесть эксплуатационных категорий и 31 855 наблюдений. При одинаковом бюджете в 12 наблюдений механизм ранжирования достигает общей оценки 67,5 против 60,9 для временного поиска и 55,9 для плоского поиска при сопоставимом оценочном контексте в 403–428 токенов. Отдельная оценка на производственном кластере включает 500 логических объектов, 2066 экземпляров времени выполнения и 24 окна инцидентов, содержащих суммарно 2 296 172 наблюдения. В зафиксированном пилотном исследовании с участием языковой модели записаны однопроходные ответы GPT-5.6 Sol Pro для 12 отложенных инцидентов. F1-мера свидетельств составляет 75,4 при контексте с учётом кандидатов и 57,0 при плоском поиске, тогда как точность локализации и определения типа неисправности не изменяется. Оценка производственного кластера дополняет контролируемое исследование; необходима дальнейшая проверка на других кластерах и наборах данных с экспертной разметкой.

Keywords: root cause analysis, microservices, large language models, observability, evidence verification, AIOps, DevOps and site reliability engineering.

Ключевые слова: анализ первопричин, микросервисы, большие языковые модели, наблюдаемость, проверка свидетельств, AIOps, DevOps и обеспечение надёжности систем.

Introduction

Modern cloud applications combine independently deployed services, managed data stores, message brokers, container schedulers, and shared infrastructure. This decomposition improves delivery velocity, but it makes failure diagnosis a cross-layer reasoning task. A frontend alert may originate in a database lock, a failed DNS lookup, a saturated worker node, a stalled runtime, or an application defect. Relevant observations are distributed across metrics, logs, traces, events, alerts, and topology, and they often appear at different times as the failure propagates [1, 32]. Common observability stacks combine time-series monitoring and distributed tracing, as illustrated by Prometheus and Jaeger [10, 23].

Traditional root cause analysis (RCA) methods use causal graphs, change-point detection, Bayesian models, and multimodal representations [9, 15, 21, 28, 31]. Such methods can be effective under stable telemetry schemas and known failure models. Production investigations also require open-ended interpretation: operators translate an alert into queries, select a subsystem, reconcile contradictory signals, and produce an explanation that another engineer can audit.

Large language models (LLMs) are relevant because they can interpret incident descriptions, call operational tools, and summarize heterogeneous evidence. Chain-of-thought prompting, ReAct, and API-oriented instruction tuning have expanded the range of tool-using agents [24, 27, 30]. AIOps research has applied LLMs to cloud incidents, database diagnosis, and collaborative troubleshooting [5, 20, 26, 33, 34]. A fluent explanation is not necessarily a correct diagnosis. An LLM may favor a frequently mentioned entity, cite a strong anomaly that occurred after the alert, or identify a downstream symptom as the root cause.

Cai et al. make this weakness measurable with AgenticOpsEval, a benchmark that separates *Localization*, *Identification*, and *Reason* [2]. Its AIOps2025 and RCA100 datasets contain 503 expert-labeled incidents and approximately 15.3 GB of multimodal observability data. Localization identifies the faulty entity, Identification names the fault, and Reason checks whether the conclusion follows from key evidence or a causal chain. We therefore make evidence selection and verification explicit stages of the diagnostic pipeline rather than leaving them as unstructured prompt context.

The objective of this study is to develop and evaluate EG-RCA, an evidence-guided architecture for microservice failure diagnosis that separates probabilistic hypothesis selection from deterministic evidence validation. Telemetry is mapped to canonical entities and timestamps. Candidate fault/entity pairs are ranked using semantic, temporal, anomaly, topology, modality, and alert-relevance signals. The LLM receives a small set of competing candidates and must return a typed explanation with evidence IDs and an ordered propagation path. A verifier then checks graph connectivity, temporal order, modality coverage, citations, and numeric claims. An explanation that fails these checks is rejected or returned for additional evidence collection.

The paper makes four contributions:

• A formal model for evidence-guided RCA over fault hypotheses, operational entities, multimodal observations, and propagation paths.

• A two-stage agent architecture in which an LLM selects among bounded hypotheses and a deterministic verifier evaluates the resulting explanation.

• A reproducible evaluation combining a 480-incident controlled corpus with a production-cluster evaluation spanning 500 logical entities, 2,066 runtime instances, and 24 incident windows.

• A recorded LLM-in-the-loop pilot on 12 held-out incidents, including the frozen prompts, candidate contexts, responses, and scoring artifacts.

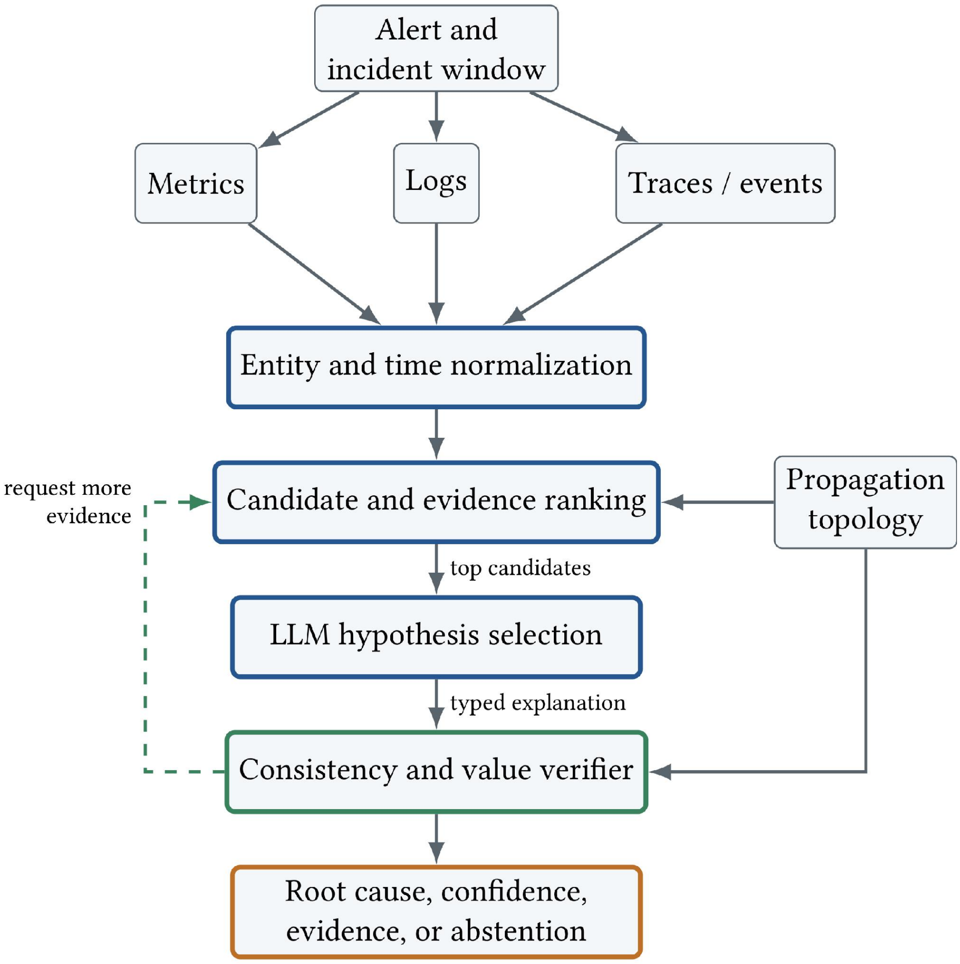

Figure 1. EG-RCA architecture. A bounded set of candidate explanations is presented to the LLM. The selected explanation is accepted only after its path, evidence IDs, temporal order, and numeric claims pass deterministic checks

Background and Related Work

Microservice Failure Diagnosis

Microservice failures are usually observed downstream from their initiating event. MicroScope and MicroRCA combine service dependencies with causal signals to localize performance problems [15, 28]. RCD applies causal discovery to failure localization [9], while BARO combines Bayesian online change-point detection with robust ranking [21]. DiagFusion learns joint representations over metrics, logs, and traces [31]. RCAEval provides a benchmark for telemetry-based microservice RCA [22], and Zhang et al. survey the broader design space [32].

These systems show why time and topology must be part of the diagnostic model. Metrics often expose resource exhaustion, logs provide semantic error context, traces identify request propagation, and lifecycle events explain scheduler or container behavior. Storage systems add replication, routing, and resource layers; TiDB is one representative distributed data platform used in microservice research [8].

LLM Agents for Operations

LLM-based diagnosis extends entity ranking with planning, query generation, and tool use. Chen et al. study automatic RCA of cloud incidents [5]. D-Bot applies an LLM-centered architecture to database diagnosis [34]. Flow-of-Action incorporates standard operating procedures into a multi-agent RCA workflow [20]; mABC and TrioXpert study collaborative and specialized agents [26, 33].

Retrieval-augmented generation reduces prompt size by selecting external evidence before generation [14]. Dense retrieval and sentence embeddings improve semantic matching beyond exact keywords [13, 25]. In operational telemetry, semantic similarity alone is insufficient. A matching error from the wrong replica, an earlier deployment, or an unrelated service may be lexically convincing while being topologically or temporally incompatible with the incident.

Benchmarks and Evaluation

OpsEval evaluates LLM capability on IT operations tasks [16]. OpenRCA studies whether LLMs can locate software failure root causes from incident evidence [29]. AIOpsLab, ITBench, and SREGym place agents in interactive environments where they inspect and manipulate realistic systems [4, 6, 11].

The closest reference is the benchmark of Cai et al. [2]. AIOps2025 contains 400 incidents over metrics, logs, and traces. RCA100 contains 103 incidents with metrics, logs, traces, events, alerts, and topology. The datasets were validated through competitions involving 6,093 teams. Their contribution is an evaluation framework and expert-labeled data. The present work addresses the method side by defining an agent architecture whose intermediate evidence and path structures can be evaluated by such benchmarks.

Table 1.

Positioning relative to representative RCA resources. “M/L/T” denotes metrics, logs, and traces. “Tools” denotes evaluated external tool interaction; “eval.” indicates evaluation support rather than integrated tool execution. “Process” denotes explicit evaluation or enforcement of the diagnostic evidence path

Resource	M/L/T	Topology	Tools	Process
RCAEval [22]	✓	–	–	–
OpenRCA [29]	✓	partial	✓	–
AIOpsLab [4]	✓	✓	✓	partial
ITBench [11]	✓	✓	✓	partial
AgenticOpsEval [2]	✓	✓	eval.	✓
EG-RCA	✓	✓	–	enforced

Materials and methods

Problem Formulation

Let the propagation structure of a cloud-native system be represented by a directed graph

$$G = (V, A), \tag{1}$$

where V contains services, pods, nodes, databases, caches, and other operational entities. An edge $(u, v) \in A$ means that a failure at u can affect v. For a caller/dependency relation, this direction is opposite to the request edge: a failed dependency can affect its caller. For a placement relation, a failed worker node can affect a hosted service.

An incident is initiated by an unknown root entity $v^\star \in V$ and fault type $h^\star \in H$. An alert is observed at entity $a \in V$ at time t_a within a window $W = [t_0, t_1]$. The telemetry is normalized into evidence objects

$$E = \{e_i\}_{i=1}^{N}, \qquad e_i = (m_i, v_i, t_i, x_i, z_i), \tag{2}$$

where m_i is modality, v_i is entity, t_i is timestamp, x_i is text or structured content, and $z_i \in [0, 1]$ is anomaly magnitude.

The agent returns

$$\hat{y} = \left(\hat{v}, \hat{h}, \widehat{P}, \widehat{E}, c\right), \tag{3}$$

where $\hat{v}$ is the root entity, $\hat{h}$ the fault type, $\widehat{P}$ an ordered path from $\hat{v}$ to a, $\widehat{E} \subseteq E$ the cited evidence, and $c \in [0, 1]$ confidence. A valid result must localize the initiating entity, identify the initiating fault rather than a symptom, cite relevant evidence, and provide a topologically valid and temporally plausible propagation path.

Diagnostic quality and context cost are represented by

$$\max_{\hat{y}} \quad Q\left(\hat{y}\right) - \lambda_b B\left(\widehat{E}\right) - \lambda_u U\left(\hat{y}\right), \tag{4}$$

where Q measures diagnosis quality, B is evidence-budget cost, and U penalizes unsupported claims.

The EG-RCA Method

Normalization and Entity Alignment

Operational identifiers differ across tools. A trace may use a service name, a Kubernetes event a pod UID, and a cloud metric a provider-specific resource ID. EG-RCA maps these identifiers to canonical entities before ranking or reasoning, following the broader direction of agent-ready observability models such as UModel [19]. Numeric values are retained in structured form so that comparisons can be checked without asking the LLM to reproduce them from text.

Evidence Relevance Scoring

For candidate fault h, alert entity a, and observation e_i, the relevance score is

$$\begin{aligned} R(e_i, h, a) &= \alpha S_{sem}(e_i, h) + \beta S_{temp}(e_i, a) \\ &+ \gamma S_{anom}(e_i) + \delta S_{topo}(e_i, a) \\ &+ \eta S_{mod}(e_i, h) + \zeta S_{alert}(e_i, a), \end{aligned} \tag{5}$$

with nonnegative weights satisfying

$$\alpha + \beta + \gamma + \delta + \eta + \zeta = 1. \tag{6}$$

S_{sem} measures compatibility between an observation and the fault description. $S_{anom} = z_i$. S_{mod} expresses fault-specific modality reliability; for example, lifecycle events are informative for crash loops and traces for request propagation. S_{alert} measures similarity to the incident description and alert symptoms. In the controlled evaluation, these two text scores are deterministic rather than embedding-based. Let $T(x)$ be the set of lowercase alphanumeric tokens in text x, and let K_h be the fault-keyword set. Then

$$S_{sem}(e_i, h) = min\left(1, \frac{|T(x_i) \cap K_h|}{\max(2, 0.55\,|K_h|)}\right), \tag{7}$$

$$S_{alert}(e_i, a) = \frac{|T(x_i) \cap T(q_a)|}{\sqrt{|T(x_i)|\ |T(q_a)|}}, \tag{8}$$

where q_a is the incident description. S_{alert} is defined as zero if either token set is empty. No learned embedding model is used in the controlled comparison.

For an observation whose entity has no directed propagation path to the alert, $S_{temp}(e_i, a) = 0$. For connected entities, temporal relevance uses an expected evidence time

$$\hat{t}(v_i, a) = t_a - \widehat{\Delta}(v_i, a), \tag{9}$$

with

$$\widehat{\Delta}(v_i, a) = \{0,\ d(v_i, a) = 0,\ \mu_0 + \mu_1 d(v_i, a),\ d(v_i, a) > 0. \tag{10}$$

Evidence on the alert entity is therefore expected near the alert, while evidence on upstream entities is expected earlier according to propagation distance. The temporal score is where

$$S_{temp}(e_i, a) = exp\left(-\frac{\left|t_i - \hat{t}(v_i, a)\right|}{\tau}\right)\kappa(t_i, t_a), \tag{11}$$

$$\kappa(t_i, t_a) = \{1,\ t_i \le t_a,\ \chi,\ t_i > t_a, \qquad 0 \le \chi < 1. \tag{12}$$

The after-alert factor prevents later anomalies from receiving the same prior as evidence that precedes the observed impact.

Topology relevance is

$$S_{topo}(e_i, a) = \{exp\left(-\frac{d(v_i, a)}{\lambda}\right),\ \pi(v_i, a) \ne \emptyset,\ 0,\ \pi(v_i, a) = \emptyset. \tag{13}$$

where $\pi(v_i, a)$ is the shortest directed propagation path from v_i to a, when one exists, and $d(v_i, a)$ is its edge length. This term is a structural plausibility prior, not proof of causation.

Temporal-order consistency

For an ordered sequence of covered path entities P_E with one selected evidence timestamp t_j per entity, temporal-order support is

$$S_{order}\left(P_E\right) = \{ \frac{1}{|P_E| - 1} \sum_{j=0}^{|P_E|-2} 1\left[t_j \leq t_{j+1} + \epsilon\right],\ |P_E| \geq 2,\ 1,\ |P_E| = 1,\ 0,\ |P_E| = 0. \tag{14}$$

A single covered entity is temporally consistent because there is no pairwise ordering constraint to violate. The tolerance ϵ accounts for clock skew and timestamp jitter; the controlled evaluation uses $\epsilon = 4\ s$.

Candidate Construction and LLM Selection

For each pair $(h, v) \in H \times V$, let $E_{h,v}^{(q)}$ be the q highest-scoring observations attached to v. Root support is

$$S_{root}(h, v) = \frac{1}{q} \sum_{e \in E_{h,v}^{(q)}} R(e, h, a). \tag{15}$$

Missing observations contribute zero through the fixed denominator q. Cross-modal support is

$$S_{xmod}(h, v) = min\left(1, \frac{|M_{h,v}|}{m_0}\right), \tag{16}$$

where $M_{h,v}$ is the set of represented modalities above the relevance threshold. The controlled implementation uses a threshold of 0.30.

For each connected candidate pair, let $P^{pre} = \pi(v, a)$ be its shortest propagation path. At each path entity, the method selects the highest-relevance admissible observation from the available telemetry. In the controlled implementation, root evidence is admissible when $S_{sem} \geq 0.18$ and anomaly magnitude is at least 0.42; downstream path evidence must contain a propagation-related term and have anomaly magnitude at least 0.38. Let P_E^{pre} denote the path entities for which such evidence is available, and let $\tilde{t}_j$ be the timestamp of the selected observation at the jth covered entity. Preliminary temporal-order support uses Eq. (14) with $\tilde{t}_j$. The pre-LLM path-consistency score is

$$S_{chain}^{pre}(h, v) = \rho_r S_{root}(h, v) + \rho_c \frac{|P_E^{pre}|}{|P^{pre}|} + \rho_o S_{order}\left(P_E^{pre}\right). \tag{17}$$

Candidate pairs then receive the bounded preliminary score

$$J_0(h, v) = \left(1 - \omega_x - \omega_p\right) S_{root}(h, v) + \omega_x S_{xmod}(h, v) + \omega_p S_{chain}^{pre}(h, v), \tag{18}$$

where ω_p controls pre-LLM path consistency. No ground-truth incident onset or ground-truth path is used in this calculation. Disconnected pairs are excluded, the top r pairs are retained, and evidence is selected separately from each candidate path so that a single retrieval error does not remove all alternatives. For evidence-context selection only, the controlled implementation adds 0.12 to observations on the candidate root and 0.05 to observations elsewhere on the candidate path; these bonuses do not change Eq. (5).

The LLM receives the alert, the top candidates, their bounded evidence sets, and a local topology view. It must return a typed structure containing root entity, fault type, ordered propagation path, and evidence IDs. The typed output limits the verifier's input surface and makes each claim addressable.

Consistency and Quantitative Verification

Let the LLM propose $P = (v_0, \dots, v_s)$ with $v_0 = \hat{v}$ and $v_s = a$. The graph is used to check structural consistency, not to establish causation by itself [18]. Let P_E be the ordered subset of path entities with cited evidence. The verifier computes temporal-order support with Eq. (14); if several cited observations belong to the same path entity, t_j is the timestamp of the highest-relevance cited observation for that entity.

Path consistency combines root evidence, entity coverage, and temporal order:

$$S_{chain}(h, v, P) = \rho_r S_{root}(h, v) + \rho_c \frac{|P_E|}{|P|} + \rho_o S_{order}(P_E), \tag{19}$$

where $\rho_r + \rho_c + \rho_o = 1$. For a connected and schema-valid explanation, final confidence is

$$C(h, v, P) = (1 - \omega_x - \omega_v) S_{root}(h, v)$$
$$+\omega_x S_{xmod}(h, v) + \omega_v S_{chain}(h, v, P), \tag{20}$$

with $0 \le \omega_x + \omega_v \le 1$. All terms lie in $[0, 1]$, so $C \in [0, 1]$. The final-confidence weight ω_v and abstention thresholds are deployment parameters. They are not instantiated in the reported controlled evaluation or the frozen LLM pilot.

The verifier also checks cited IDs and structured values. A statement such as "CPU exceeded 90 %" is accepted only if the cited metric contains a matching value in the relevant interval. A trace claim must cite a span on the proposed path. An event claim must refer to the canonical entity associated with the hypothesis. If the best explanation fails validation, has confidence below $c_{\min}$, or has margin below $\Delta_{\min}$, the system requests more evidence or abstains. When two validated candidates are available, the confidence margin is the difference between the highest and second-highest values of C. The two thresholds are deployment parameters.

Algorithm 1. Evidence-guided LLM root cause analysis

Require: Alert a, window W, evidence E, graph G, faults H, budget k

1: Normalize entities, timestamps, units, and modality schemas

2: **for all** $(h, v) \in H \times V$ **do**

3: Score observations with Eq. (5)

4: Compute $J_0(h, v)$ and discard disconnected candidates

end for
Retain the top r candidate pairs and allocate evidence budget k
Ask the LLM for a typed diagnosis, path, and evidence IDs
Validate schema, citations, topology, time order, and numeric claims
if validation succeeds **then**
Compute $C(h, v, P)$ with Eq. (20)
end if
if validation fails or confidence margin is insufficient **then**
request additional evidence or abstain
else
return $\left(\hat{v}, \hat{h}, \widehat{P}, \widehat{E}, c\right)$
end if

Complexity

A direct implementation evaluates $|H||V|$ pairs over N observations, giving $O(|H||V|N)$ scoring time. Entity indexing limits each pair to local observations and nearby path entities. Cached shortest paths make structural validation $O(|P| + k)$. A production implementation may replace the controlled token-overlap semantic score with an embedding index; time and topology terms remain deterministic filters.

Evaluation

The evaluation has three parts. First, the controlled mechanism evaluation isolates ranking, topology, time, cross-modal support, and pre-LLM path consistency from model-specific behavior. Second, a frozen LLM-in-the-loop pilot evaluates whether a model can consume candidate-aware context and reports basic post-LLM schema, evidence-ID, root-support, and propagation-path checks. Third, the production-cluster evaluation tests the deterministic ranking pipeline on a 500-entity topology with high-volume telemetry. Numeric-claim validation, confidence thresholds, and iterative evidence requests are not evaluated in the pilot.

Data status

The large-scale results report measured production-cluster telemetry collected under the evaluation protocol. The controlled results remain reproducible reference values and provide a fixed comparison point.

Controlled Incident Corpus

The controlled environment contains 15 entities: ten application services, PostgreSQL, Redis, and three worker nodes. The caller-to-dependency graph and the cause-to-effect propagation graph are stored separately. Worker-node edges point from a node to its hosted services; dependency failures propagate from a dependency to its callers. The topology is informed by AIOps2025, RCA100, Online Boutique, and OpenTelemetry Demo [2, 7, 17]. The fault taxonomy follows common Kubernetes chaos-engineering practice [3].

Using seed 29, the controlled reference corpus contains 480 incidents. Twelve fault types are grouped into resource, network, runtime, middleware/database, application logic, and orchestration/platform categories. Each incident contains a root entity, alert entity, propagation path, and multimodal key evidence. The reference workload also contains three distractor classes: a topology-valid competing fault, a downstream symptom cluster, and observations on unrelated entities. Distractors may share vocabulary and anomaly magnitude with the true root.

The corpus contains 31,855 observations, or 66.4 per incident. Each incident has 4.19 key observations on average. Propagation paths contain one to three entities: 56 local failures, 277 two-entity paths, and 147 three-entity paths. Category counts are 108 resource, 87 network, and 74 runtime incidents. The remaining categories contain 61 middleware/database, 75 application-logic, and 75 orchestration/platform incidents.

Table 2.

Controlled evaluation corpus

Property	Value
Incidents	480
Entities	15
Fault types / categories	12 / 6
Telemetry modalities	5 + topology
Total observations	31,855
Mean observations per incident	66.4
Mean key evidence per incident	4.19
Path distribution	56 / 277 / 147
Main-evaluation seed	29

Compared Strategies

The controlled evaluation compares five deterministic strategies:

Keyword aggregation. Lexical support and anomaly magnitude are aggregated over all observations.

Greedy traversal. Starting at the alert entity, the method repeatedly inspects the most anomalous direct dependency or hosting node. This is a deterministic traversal baseline, not an implementation of ReAct.

Flat retrieval. Semantic, anomaly, modality, topology, and alert similarity are used without temporal scoring or explicit path verification.

Temporal retrieval. Flat retrieval is extended with temporal plausibility and a weak topology prior, but no explicit consistency verifier.

EG-RCA. The complete candidate ranking and pre-LLM path-consistency mechanism is used. The highest-scoring candidate is selected deterministically so that the controlled results do not depend on a particular language model. The frozen LLM pilot evaluates only the basic post-LLM checks listed above, not numeric-claim or confidence validation.

The main configuration uses

$$(\alpha, \beta, \gamma, \delta, \eta, \zeta) = (0.34, 0.18, 0.16, 0.16, 0.10, 0.06), \tag{21}$$

with pre-LLM path-consistency weight $\omega_p = 0.40$, cross-modal weight $\omega_x = 0.08$, and evidence budget $k = 12$. The same parameter configuration is used for all incidents and is not tuned by fault category.

Metrics

Following the Localization–Identification–Reason structure of Cai et al. [2], localization accuracy (LA) is exact root-entity match and type accuracy (TA) is exact fault-type match. Evidence precision and recall compare cited evidence IDs with the key-evidence set; their harmonic mean is Evidence F1. Chain F1 represents each ordered path as position–entity pairs $\{(j, v_j)\}$ and computes set F1 over those pairs, so both entity identity and path position must match. All evidence and path metrics are computed per incident and then macro-averaged. Consequently, the reported mean Evidence F1 is not generally the harmonic mean of the reported mean precision and mean recall. The process score is

$$PS = 0.55\, EvidenceF1 + 0.45\, ChainF1. \tag{22}$$

The overall score is

$$Overall = 0.40\, LA + 0.30\, TA + 0.30\, PS. \tag{23}$$

The 0.40/0.30/0.30 weighting follows the entity/fault/process emphasis of RCA100, but the process score in Eq. (22) is a controlled surrogate defined for this controlled reference corpus and is not the native RCA100 process metric. Overall scores are computed from unrounded incident-level values and then averaged; displayed component values are rounded independently.

Approximate prompt tokens are estimated from all word occurrences in the incident description and selected evidence, using a fixed 90-word instruction/output-schema allowance, an 8-word metadata allowance per selected observation, and 0.75 words per token. Percentile-bootstrap 95 % intervals use 5,000 resamples of the 480 incident-level scores.

LLM-in-the-Loop Pilot

The LLM pilot uses a separate seed (260824) and 12 held-out incidents, one per fault type. It is a closed-set diagnosis task: each prompt includes the allowed entities, allowed fault types, and the propagation graph. Each condition uses an evidence budget of at most 12 observations. The flat condition uses the Flat Retrieval context. The guided condition contains two ranked fault/entity candidates. With two candidates, the context builder first allocates up to five observations per candidate path, ensures that one alert observation is present, and fills any remaining positions with the highest-scoring evidence from those paths. GPT-5.6 Sol Pro produces one JSON response per frozen prompt with no retries or external tools. Responses contain root entity, fault type, evidence IDs, and path. The prompts, outputs, labels, and scoring script are included in the artifact.

The pilot prompt set is kept frozen so the recorded responses remain auditable. It predates the revised single-entity temporal-order convention used in the controlled evaluation, so the pilot numbers are reported only as an interface demonstration and are not used to validate the revised controlled candidate-ranking scores. The basic pilot verifier checks output schema, cited-ID validity, root-support evidence, and propagation-path consistency. It does not evaluate numeric claims, final confidence, abstention thresholds, or iterative evidence requests. The two prompt conditions also differ in candidate hints and evidence allocation, so the pilot does not isolate a verifier effect. With only 12 incidents, it cannot establish model-level generality.

Large-Scale Production-Cluster Evaluation

The production-cluster evaluation uses a separate topology with 500 logical entities: 360 application services, 40 data or messaging entities, 80 worker nodes, and 20 platform entities. A deployment map associates 2,066 runtime instances with these logical entities. The evaluation contains 24 incident windows covering the 12-type evaluation taxonomy. Each window contains between 50,013 and 146,879 telemetry observations, with a mean of 95,674 and 2,296,172 observations in total. Telemetry records are associated with runtime instances and normalized to their logical entities before scoring. The measured windows include the background and incident-related telemetry present during collection, including concurrent and unrelated activity observable in the same time interval.

All three large-scale methods use the same evidence budget $k = 12$. A shared feature index is built once per incident window and computes the text, anomaly, temporal, modality, and topology features required by the compared methods. Because this common pass performs full-window ingestion and feature extraction, its timing is treated only as an implementation baseline rather than as an algorithmic speedup. The shared feature-index pass takes 6.03 s on average in the evaluated environment.

Results and discussion

Results

Controlled Diagnostic Quality

Table 3 reports the controlled results with a matched 12-observation evidence budget for every method. EG-RCA obtains 65.4 % localization accuracy, 81.2 % type accuracy, a process score of 56.6, and an overall score of 67.5. Temporal Retrieval is the strongest retrieval baseline at 60.9 overall. The paired improvement over Temporal Retrieval is 6.58 percentage points, with a bootstrap 95 % interval of 3.49–9.91 points. The improvement over Flat Retrieval is 11.59 points, with an interval of 8.00–15.12 points.

Relative to Temporal Retrieval, localization improves by 12.9 points and Chain F1 by 11.9 points, while type accuracy differs by only 0.6 points in favor of EG-RCA. The main controlled gain therefore comes from root attribution and path quality rather than fault-family classification.

Table 3.

Controlled results over 480 incidents with a matched k=12 evidence budget for all methods. Scores are macro-averaged percentages; higher is better except prompt tokens. Overall intervals are percentile-bootstrap 95 % confidence intervals. Displayed components are rounded independently

Metric	Keyword aggregation	Greedy traversal	Flat retrieval	Temporal retrieval	EG-RCA
Loc.	40.2	12.3	46.2	<u>52.5</u>	**65.4**

Metric	Keyword aggregation	Greedy traversal	Flat retrieval	Temporal retrieval	EG-RCA
Type	78.1	79.2	79.6	80.6	**81.2**
Ev. P	22.0	28.9	24.9	**30.1**	29.5
Ev. R	63.1	81.8	71.9	**86.5**	75.9
Ev. F1	32.5	42.5	36.8	**44.5**	42.2
Chain	49.0	15.5	55.4	62.3	**74.2**
Process	40.0	30.3	45.2	52.5	**56.6**
Overall (95% CI)	51.5 [48.4, 54.5]	37.8 [35.7, 39.8]	55.9 [52.9, 58.8]	60.9 [58.0, 64.0]	**67.5 [64.3, 70.6]**
Tokens	427	417	427	428	**403**

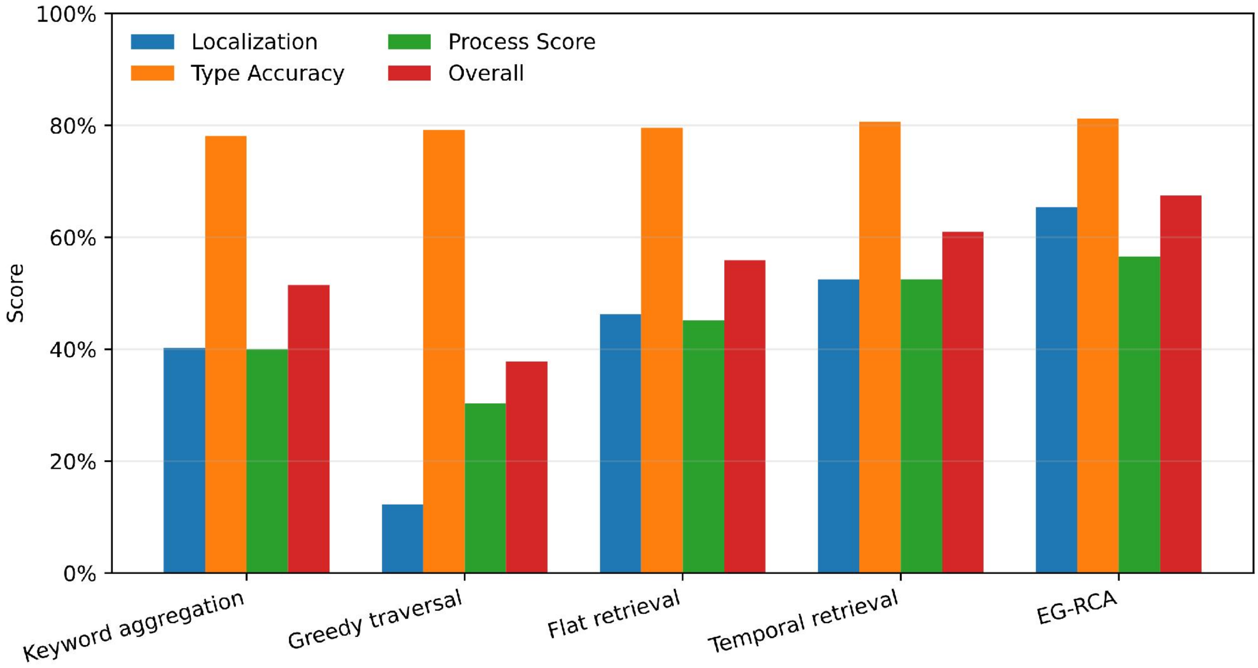

Figure 2. Localization, type accuracy, process quality, and overall score in the controlled evaluation

Performance by Fault Category

Figure 3 shows category-level results. EG-RCA has the highest overall score in five of six categories: runtime (70.9), resource (70.2), network (69.7), orchestration/platform (69.4), and application logic (62.7). Keyword Aggregation leads middleware/database incidents at 73.1, compared with 59.1 for EG-RCA. Explicit SQL and Redis vocabulary makes that category favorable to lexical matching in the controlled reference corpus. Application-logic failures remain difficult because a downstream HTTP error and an initiating application defect can produce nearly identical text.

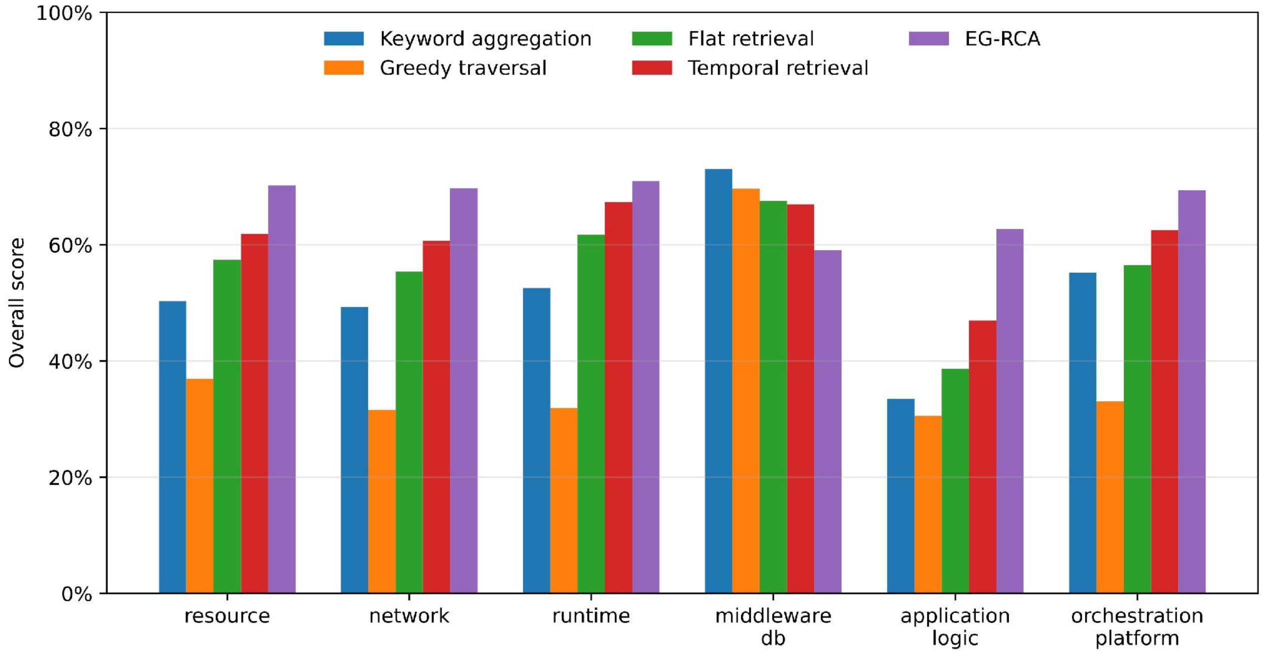

Figure 3. Overall score by operational fault category. EG-RCA leads five categories; Keyword Aggregation is strongest for middleware/database faults

Evidence Budget and Context Efficiency

Figure 4 evaluates evidence budget k on the same 480 incidents as the main experiment. Overall score is 69.3 at $k = 5$, 68.8 at $k = 8$, and 67.5 at $k = 12$. The decline after $k = 5$ is modest, while evidence recall rises from 59.2 % at $k = 5$ to 75.9 % at $k = 12$. The main configuration uses $k = 12$ to preserve a broader auditable evidence set while keeping the prompt context bounded and slightly smaller than the retrieval baselines.

With the matched $k = 12$ budget, all three ranking methods use comparable estimated prompt contexts of approximately 400–430 tokens per incident: 403 for EG-RCA, 428 for Temporal Retrieval, and 427 for Flat Retrieval. The diagnostic gain of EG-RCA therefore does not result from a larger evidence budget or a larger estimated prompt context. Figure 5 shows the corresponding quality–context trade-off.

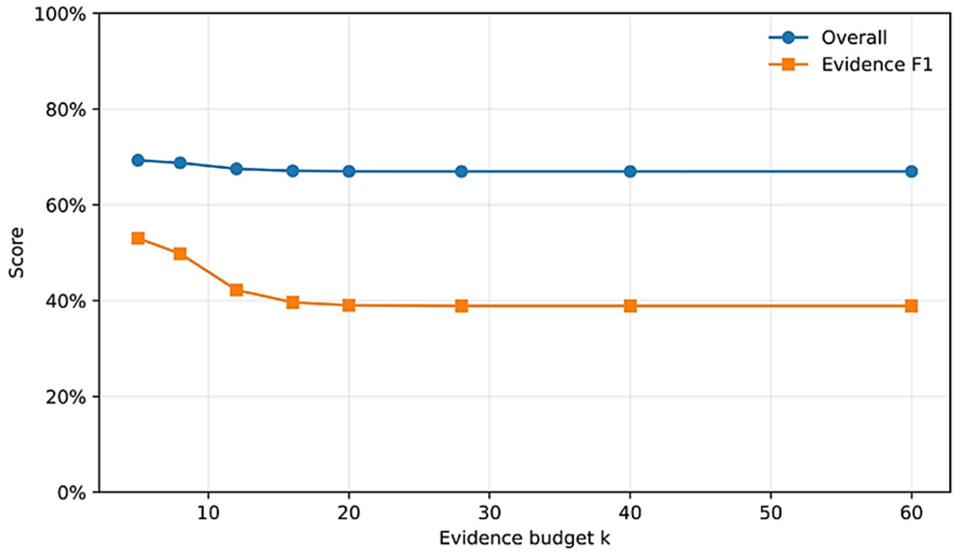

Figure 4. Sensitivity to evidence budget k. Smaller contexts produce the highest overall score, while larger budgets improve key-evidence recall

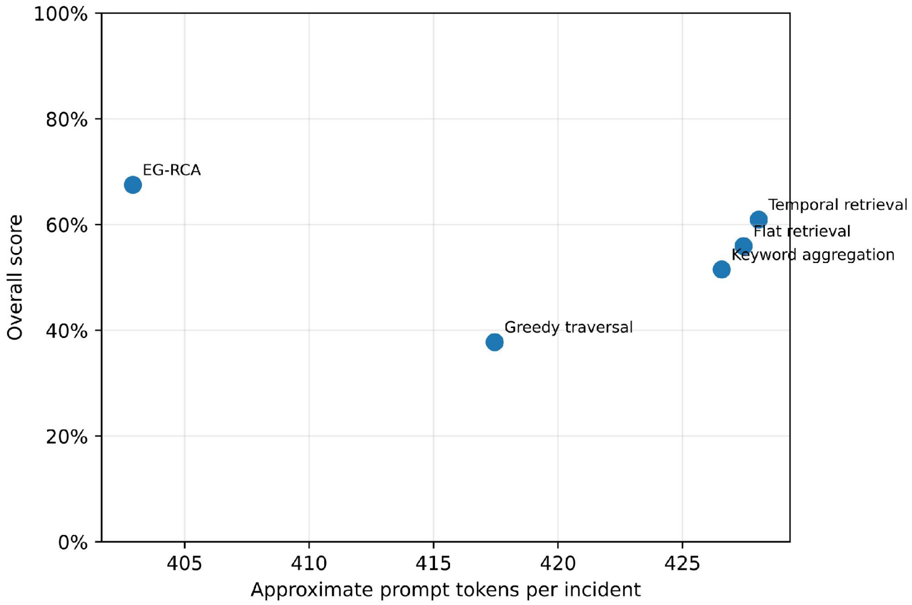

Figure 5. Controlled diagnostic quality versus approximate prompt context. Higher and farther left is better

LLM-in-the-Loop Pilot

Table 4 reports the recorded LLM outputs. Both contexts produce 75.0 % localization accuracy, 83.3 % type accuracy, and 75.0 % joint root/type accuracy. The candidate-aware context changes the quality of the supporting explanation: Evidence F1 increases from 57.0 to 75.4, and overall score rises from 74.5 to 77.6. The improvement comes from more complete and less extraneous citations rather than from additional correct root diagnoses.

The basic post-LLM checks accept 11 of 12 Flat Retrieval outputs and all 12 guided outputs. Acceptance does not imply correctness: structurally valid but incorrect hypotheses remain possible. Three guided diagnoses are wrong despite passing these checks. Acceptance therefore indicates consistency with the checked schema, evidence IDs, root support, and propagation path, not that the inferred cause is necessarily correct.

Table 4.

Recorded GPT-5.6 Sol Pro pilot on 12 held-out incidents. Scores are percentages; "Accepted" means that the output passed the basic schema, evidence-ID, root-support, and propagation-path checks

Context	Loc.	Type	Joint	Ev. F1	Overall	Accepted
Flat retrieval	75.0	83.3	75.0	57.0	74.5	91.7
Frozen guided	75.0	83.3	75.0	75.4	77.6	100.0

Large-Scale Production-Cluster Results

Table 5 reports the 24-window production-cluster evaluation. With the same $k = 12$ evidence budget, EG-RCA reaches 89.5 overall, compared with 65.2 for Temporal Retrieval and 54.0 for Flat Retrieval. For EG-RCA, localization and fault-type accuracy are 95.8 % and 95.8 %, respectively. Because the production-cluster evaluation and the controlled corpus use different topologies and data distributions, their absolute scores are not directly comparable. The 24 measured windows characterize this evaluation campaign rather than production environments in general.

All methods ingest the full incident window through the same shared feature-index pass. The shared feature-index pass takes 6.03 s on average in the evaluated environment. With the matched $k = 12$ budget, estimated prompt contexts are 418, 430, and 436 tokens for EG-RCA, Temporal Retrieval, and Flat Retrieval, respectively. The large-scale evaluation therefore examines diagnostic behavior at higher topology and telemetry volume rather than claiming an implementation-specific post-index speedup.

Table 5.

Production-cluster evaluation over 24 incident windows, 500 logical entities, 2,066 runtime instances, and 2,296,172 observations. All methods use. Ablation Study

Method	Loc.	Type	Ev. F1	Chain	Overall	Tokens
Flat retrieval	58.3	62.5	23.6	59.2	54.0	436
Temporal retrieval	70.8	70.8	35.3	72.7	65.2	430
EG-RCA	95.8	95.8	57.6	95.8	89.5	418

$$k = 12$$

The ablation study uses the first 240 incidents from the seed-29 controlled corpus. Removing structured ranking causes the largest decrease, from 66.5 to 45.5 overall. Removing cross-modal support reduces the score by 7.1 points, removing pre-LLM path consistency by 6.3 points, removing temporal scoring by 4.2 points, and removing topology by 2.0 points. This ablation concerns candidate scoring before the LLM; it does not measure the post-LLM checks.

The no-ranking condition has higher Evidence F1 than the full method because anomaly-only selection retrieves many key IDs together with unrelated observations. Evidence overlap does not identify which entity initiated the incident or whether the observations form the correct path. Localization and Chain F1 capture that failure. Reporting evidence precision, recall, and path quality separately prevents Evidence F1 from being interpreted as a complete diagnosis metric.

Table 6.

Ablation results on 240 incidents. Scores are percentages

Metric	Full EG-RCA	– cross-modal support	– temporal scoring	– topology	– path consistency	– structured ranking
Loc.	**64.2**	53.8	58.8	59.6	54.6	29.2
Type	**80.8**	77.1	77.9	80.8	78.8	67.5
Ev. P	28.5	25.9	27.2	29.6	26.0	**34.3**
Ev. R	74.5	67.8	70.6	84.8	69.3	**98.0**
Ev. F1	41.0	37.3	39.1	43.7	37.6	**50.7**
Chain	**72.6**	64.1	66.7	67.8	63.1	39.0
Process	**55.2**	49.3	51.5	54.6	49.1	45.4
Overall	**66.5**	59.4	62.3	64.4	60.2	45.5

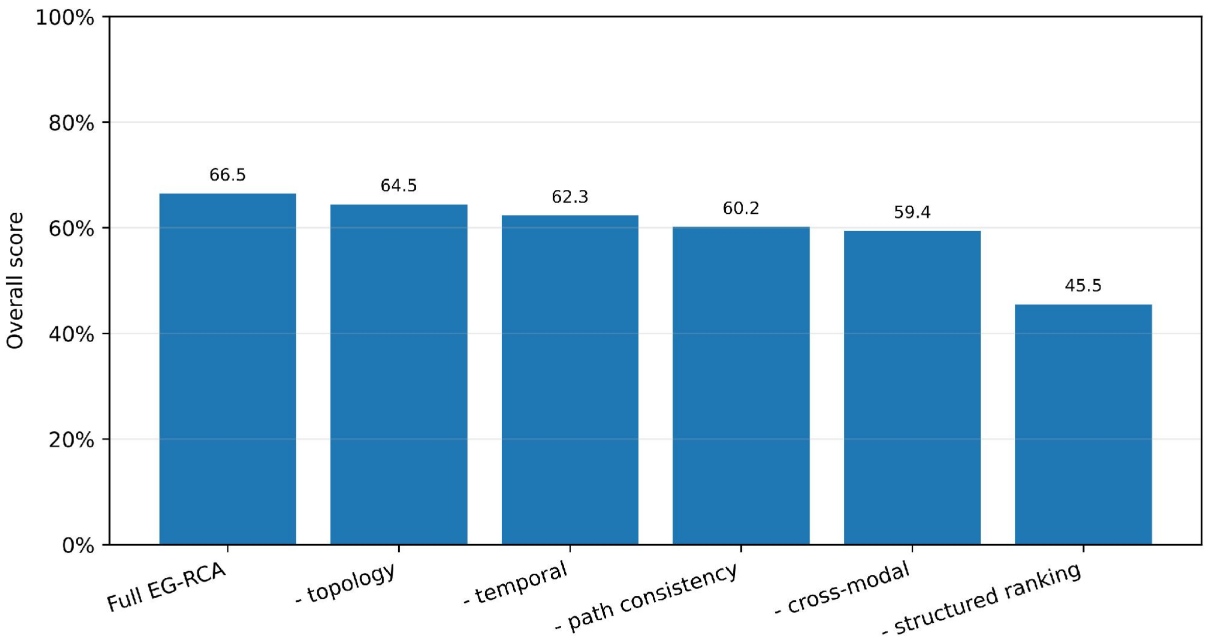

Figure 6. Ablation of EG-RCA components. Structured ranking provides the largest contribution; cross-modal support and pre-LLM path consistency provide the largest additional gains

Case Study

Figure 7 illustrates a controlled reference slow-SQL incident. PostgreSQL emits metric, log, and trace evidence near $t = 0$ indicating a lock-related query delay. At $t = 37.6\,s$, the product-catalog service shows dependency retries and increased span duration. At $t = 76.6\,s$, the frontend produces user-visible latency and error evidence; the incident alert time is $t_a = 88.4\,s$. The resulting propagation path consists of postgres, productcatalog, and frontend, in that order.

A CPU-saturation cluster at an unrelated service can receive high semantic and anomaly scores. Flat retrieval may rank it highly. The EG-RCA path-consistency stage excludes that candidate because no directed propagation path connects the candidate entity to the frontend alert. The retained output cites evidence IDs, allowing an operator to inspect each supporting observation.

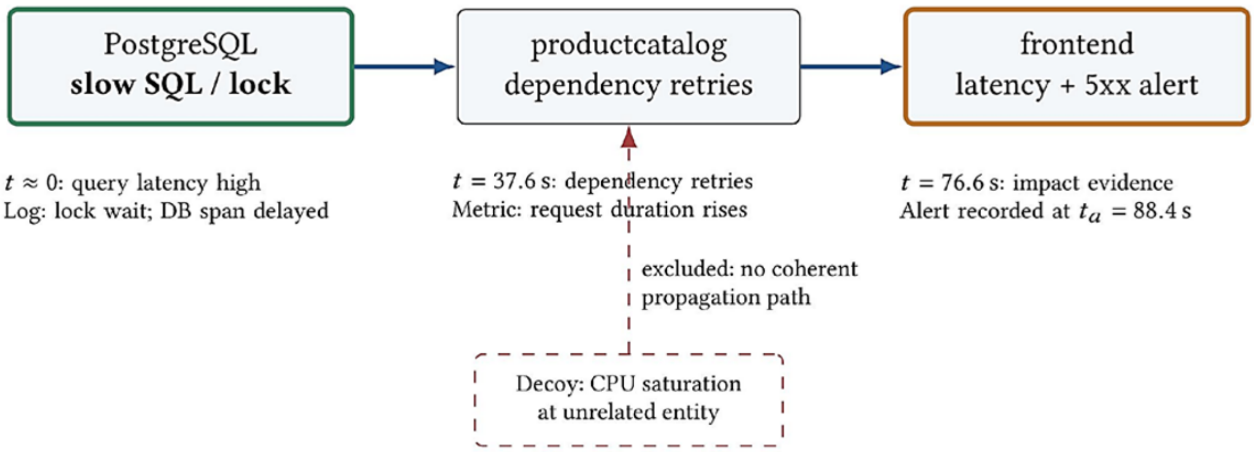

Figure 7. Case study of an evidence-grounded slow-SQL diagnosis. The retained candidate follows the propagation path and temporal order; a semantically elevant but structurally disconnected decoy is excluded during candidate ranking

Discussion

Effect of Verification

Temporal Retrieval and EG-RCA have similar type accuracy, but EG-RCA improves root localization and path quality. Semantic evidence is often sufficient to recognize a fault family such as packet loss, full GC, or slow SQL. The larger error source is assigning that fault to the initiating entity and connecting it to the observed alert.

The post-LLM verifier specifies explicit checks for the selected explanation. A database explanation must connect to a caller, a lifecycle explanation must cite an event, root evidence should precede downstream impact, and numeric claims must match structured values. The pilot evaluates only the basic schema, citation-ID, root-support, and propagation-path subset of these checks. These checks can reject explanations that violate the encoded consistency conditions, but the current experiment does not isolate their causal effect on diagnostic accuracy.

Relationship to AgenticOpsEval

Cai et al. provide key-evidence labels in AIOps2025 and causal-chain checkpoints in RCA100 [2]. Equation (5) supports key-evidence ranking, while Eq. (19) scores path consistency. The next external validation should run EG-RCA on both datasets and report their native scoring protocols alongside Eq. (23); the controlled process score used here should not be interpreted as a native AgenticOpsEval or RCA100 score.

The architecture also fits interactive environments. AIOpsLab provides live fault injection and tool execution [4]; ITBench and SREGym can test whether evidence-grounded diagnosis improves subsequent remediation [6, 11]. In such deployments, the verifier should remain between diagnosis and action, and proposed commands should pass separate policy, blast-radius, and rollback controls.

From RCA to Semi-Autonomous Remediation

The same interface can be extended from diagnosis to remediation. A remediation proposal should identify its target entity, preconditions, expected effect, validation query, and rollback operation. Prior incident-mitigation research also supports grounding corrective actions in operational evidence and executable guidance [12]. The workflow can be separated into the following stages:

$$(1)\ diagnosis, (2)\ proposal,\ (3)\ verification,$$

$$(4)\ execution,\ (5)\ observation, (6)\ rollback\ when\ necessary.$$

A correct diagnosis does not guarantee a safe change. Increasing a connection pool, for example, may reduce local queueing while increasing database saturation. Remediation therefore requires controls beyond diagnostic evidence verification.

Threats to Validity

Production measurement scope

The controlled corpus provides exact labels and repeatability, while the large-scale evaluation uses measured telemetry from one production cluster. Results may differ across clusters, traffic regimes, instrumentation quality, and concurrent-fault conditions.

Evaluation design and selection effects

The controlled corpus is structured to test temporal, topological, and evidence-grounding mechanisms. The production-cluster evaluation reduces dependence on that construction, but incident selection, labeling, topology snapshots, and evidence annotation can still affect results. Independent benchmark and multi-cluster validation remain necessary.

Deterministic controlled strategies

The 480-incident comparison uses deterministic strategy implementations. This isolates candidate ranking and evidence orchestration from model behavior but does not measure hallucination, tool-selection error, or prompt sensitivity. The separate frozen LLM pilot covers candidate-aware prompting and basic post-LLM checks on only 12 incidents; it does not evaluate the full verifier or establish model-level generality.

Topology and fault-catalog coverage

The evaluation uses a 15-entity controlled topology and a measured production topology with 500 logical entities; both use the same 12-type evaluation taxonomy. The production-cluster measurement adds runtime replicas, shared infrastructure, longer paths, and high-volume telemetry, but one cluster cannot establish coverage of regional failover, asynchronous business workflows, or continuously changing service discovery. A deployed topology may also be incomplete or stale.

Metric design

Equation (23) assigns 40 % to localization, 30 % to fault type, and 30 % to process quality. The 0.40/0.30/0.30 weights mirror the entity/fault/process emphasis of RCA100, but the process term is our controlled surrogate rather than the native RCA100 process metric. The weights remain a design choice. Evidence F1 also ignores attribution unless it is paired with localization and path metrics.

Approximate context cost

Prompt tokens are estimated from lexical word counts and fixed metadata allowances rather than a model-specific tokenizer. Repeated words are counted separately. Controlled runtime excludes LLM inference, network latency, telemetry queries, and tool execution.

Operational Safety and Ethics

An RCA agent may access logs containing secrets, personal data, customer identifiers, or proprietary source information. Evidence normalization should enforce access control, field-level redaction, retention limits, and auditable data lineage. The LLM should receive only the observations needed for the incident.

The system described here does not autonomously modify infrastructure. Diagnosis and remediation permissions should remain separate. High-impact actions require policy evaluation and, where appropriate, human approval. Passing the evidence checks does not imply that a remediation action is safe.

Conclusion

This paper presented EG-RCA, an architecture for improving LLM root cause analysis with bounded evidence selection and deterministic verification. The controlled candidate-ranking evaluation showed higher localization, path quality, and overall score than lexical, traversal, and retrieval baselines under a matched evidence budget. A separate 500-entity production-cluster evaluation measured the same ranking design at roughly 50,013–146,879 observations per incident window. The recorded LLM pilot had higher Evidence F1 with candidate-aware context while root/type accuracy was unchanged on the 12-case sample.

The results support separating hypothesis generation from evidence validation in LLM-based AIOps systems. They do not establish production readiness. Evaluation on expert-labeled benchmarks, multiple model families, incomplete topologies, and interactive remediation tasks remains necessary.

Appendix A. Controlled Fault Taxonomy

Table 7.

Fault types used in the controlled evaluation

Category	Fault types
Resource	CPU saturation; memory pressure
Network	network latency; packet loss

Category	Fault types
Runtime	full GC; thread exhaustion
Middleware/DB	slow SQL; Redis unavailable
Application logic	HTTP 5xx; rate limiting
Orchestration/platform	pod CrashLoop; DNS failure

Appendix B. Scoring Configuration

Table 8.

Main EG-RCA parameters

Parameter	Value
Semantic weight α	0.34
Temporal weight β	0.18
Anomaly weight γ	0.16
Topology weight δ	0.16
Modality weight η	0.10
Alert relevance ζ	0.06
Temporal scale τ	36 s
Base delay μ_0	52 s
Per-hop delay μ_1	18 s
After-alert factor χ	0.25
Temporal-order tolerance ϵ	4 s
Topology scale λ	3.1
Root evidence count q	4
Desired modalities m_0	3
Cross-modal relevance threshold	0.30
Candidate count r	2
Root weight ρ_r	0.34
Coverage weight ρ_c	0.38
Order weight ρ_o	0.28
Pre-LLM path-consistency ω_p	0.40
Cross-modal weight ω_x	0.08
Evidence budget k	12

Controlled implementation details

The pre-LLM path scorer uses root admissibility thresholds $S_{sem} \geq 0.18$ and anomaly magnitude ≥ 0.42; downstream path evidence requires a propagation-related term and anomaly magnitude ≥ 0.38. Evidence-context ranking adds a root bonus of 0.12 and a path bonus of 0.05. The prompt-cost estimator uses a 90-word fixed instruction/schema budget, 8 metadata words per selected observation, and 0.75 words per approximate token. These constants are also recorded in the accompanying code.

Appendix C. Baseline Configuration

Flat and Temporal Retrieval share a two-stage fault-family and entity-localization procedure but use different evidence weights. Table 9 records their controlled configuration. Keyword Aggregation and Greedy Traversal do not use this weight vector. All five methods in Table 3 use the same final evidence budget $k = 12$.

Table 9.

Controlled Flat and Temporal Retrieval parameters

Parameter	Flat	Temporal
Semantic	0.56	0.40
Temporal	0.00	0.20
Anomaly	0.18	0.15
Topology	0.10	0.12
Modality	0.10	0.08
Alert relevance	0.06	0.05
Cross-modal weight	0.04	0.05
Cross-modal threshold	0.28	0.28
Disconnected penalty	-0.14	-0.14
Global fault evidence count	8	8
Entity evidence count	5	5
Evidence budget k	12	12

Appendix D. Temporal-Prior Sensitivity

To check that the controlled result is not tied to the exact temporal-prior constants, we reevaluate the first 240 seed-29 incidents under four parameter settings. The scorer never receives the reference incident onset. Table 10 shows that EG-RCA remains ahead of Temporal Retrieval for each tested prior, although absolute performance varies with the assumed propagation delay.

Table 10.

Sensitivity to the temporal prior on 240 fixed incidents. Scores are percentages

Setting	μ_0	μ_1	τ	Temporal	EG-RCA
Main	52	18	36	60.2	66.5
Shorter delay	40	12	48	55.2	59.8
Longer delay	68	24	48	62.0	66.9
Weak temporal prior	52	18	72	59.2	65.0

Appendix E. Topology-Valid Distractor Stress Test

The stress test uses 90 fixed reference incidents constructed with seed 1029 and increases the number of semantically plausible observations attached to entities that retain a propagation path to the alert. At 20 distractor observations, EG-RCA reaches 65.1 overall, compared with 44.4 for Flat

Retrieval and 50.1 for Temporal Retrieval. At 120 distractors, EG-RCA reaches 28.6, Temporal Retrieval 7.2, and Flat Retrieval 3.3. EG-RCA remains ahead at all tested noise levels, but its score falls substantially as topology-valid contamination increases.

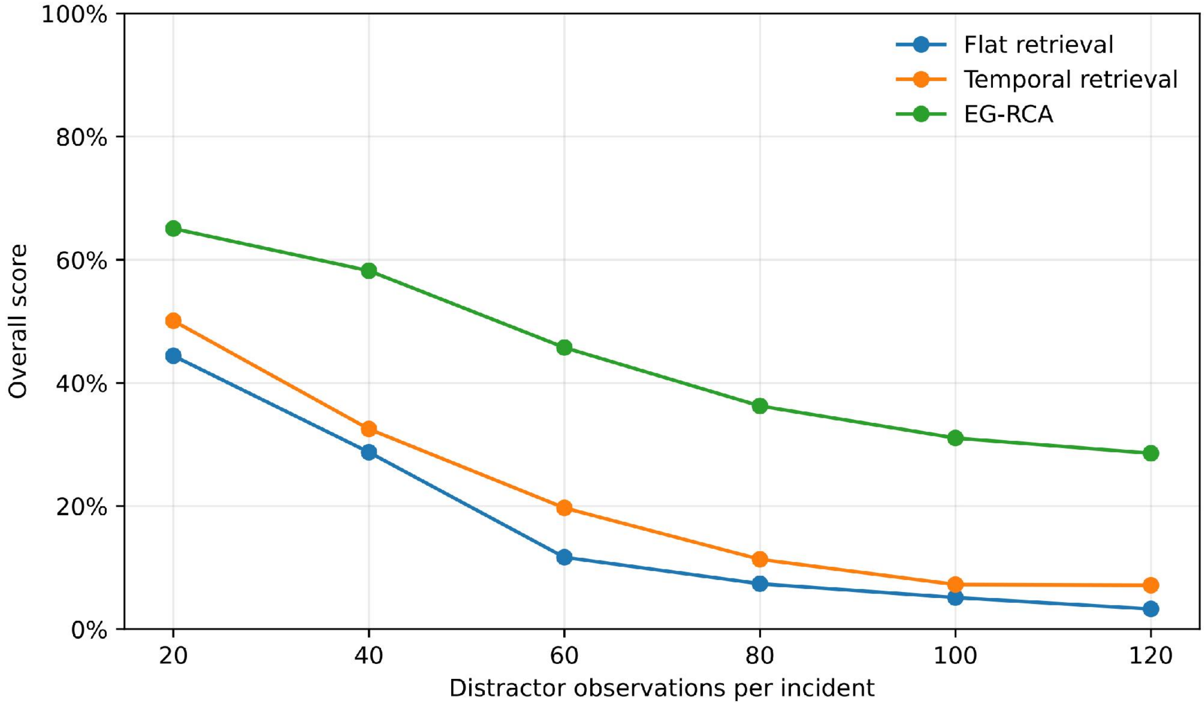

Figure 8. Overall score as topology-valid distractor observations are added

Appendix F. Reproducibility

The accompanying artifact includes reference-workload construction scripts, parameters, per-incident outputs, bootstrap intervals, tables, figures, frozen LLM prompts and responses, and scoring code. The controlled evaluation uses seed 29; ablation and temporal-prior studies use its first 240 incidents, and the topology-valid noise test uses seed 1029. The large-scale production-cluster evaluation is reproduced from exported incident windows, the topology snapshot, runtime-instance mapping, labels, and evaluation outputs included in the accompanying artifact. The LLM pilot uses seed 260824 and can be rescored exactly from its recorded outputs; its frozen prompt set predates the revised single-entity temporal-order convention and is not part of the revised controlled-score comparison.

References

1. Beyer B. Site Reliability Engineering: How Google Runs Production Systems / B. Beyer [et al.]. – Sebastopol: O'Reilly Media, 2016. – 524 P.
2. Cai Y. A Multi-Dataset Benchmark for Evaluating LLM Agents in Microservice Failure Diagnosis / Y. Cai [et al.]. // arXiv preprint arXiv:2606.29193. – 2026. (дата обращения: 14.09.2026).
3. Chaos Mesh: A Powerful Chaos Engineering Platform on Kubernetes – URL:. – 2020. (дата обращения: 14.09.2026).

4. Chen Y. AIOpsLab: A Holistic Framework to Evaluate AI Agents for Enabling Autonomous Clouds / Y. Chen [et al.] // Proceedings of Machine Learning and Systems. – 2025. – Vol. 7.
5. Chen Y. Automatic Root Cause Analysis via Large Language Models for Cloud Incidents / Y. Chen [et al.] // Proceedings of the Nineteenth European Conference on Computer Systems. – 2024. – P. 674–688. – DOI: 10.1145/3627703.3629553.
6. Clark J. SREGym: A Live Benchmark for AI SRE Agents with High-Fidelity Failure Scenarios / J. Clark [et al.]. // arXiv preprint arXiv:2605.07161. – 2026. (дата обращения: 14.09.2026).
7. Online Boutique (HipsterShop): A Cloud-First Microservices Demo Application / Google Cloud Platform. – URL:. – 2018. (дата обращения: 14.09.2026).
8. Huang D. TiDB: A Raft-Based HTAP Database / D. Huang [et al.] // Proceedings of the VLDB Endowment. – 2020. – Vol. 13, No. 12. – P. 3072–3084. – DOI: 10.14778/3415478.3415535.
9. Ikram A. Root Cause Analysis of Failures in Microservices through Causal Discovery / A. Ikram [et al.] // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – P. 31158–31170.
10. Jaeger: Open Source, End-to-End Distributed Tracing – URL:. – 2017. (дата обращения: 14.09.2026).
11. Jha S. ITBench: Evaluating AI Agents across Diverse Real-World IT Automation Tasks / S. Jha [et al.] // Proceedings of Machine Learning Research. – 2025. – Vol. 267. – P. 27134–27197.
12. Jiang J. How to Mitigate the Incident? An Effective Troubleshooting Guide Recommendation Technique for Online Service Systems / J. Jiang [et al.] // Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering. – 2020. – P. 1410–1420. – DOI: 10.1145/3368089.3417054.
13. Karpukhin V. Dense Passage Retrieval for Open-Domain Question Answering / V. Karpukhin [et al.] // Proceedings of the. – 2020. – P. 6769–6781. – DOI: 10.18653/v1/2020.emnlp-main. 550.
14. Lewis P. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks / P. Lewis [et al.] // Advances in Neural Information Processing Systems. – 2020. – Vol. 33. – P. 9459–9474.
15. Lin J. MicroScope: Pinpoint Performance Issues with Causal Graphs in Micro-Service Environments / J. Lin, P. Chen, Z. Zheng // International Conference on Service-Oriented Computing. – 2018. – P. 3–20. – DOI: 10.1007/978-3-030-03596-9_1.
16. Liu Y. OpsEval: A Comprehensive IT Operations Benchmark Suite for Large Language Models / Y. Liu [et al.]. // arXiv preprint arXiv:2310.07637. – 2023. (дата обращения: 14.09.2026).
17. OpenTelemetry Demo: A Microservice-Based Distributed Application – URL:. – 2022. (дата обращения: 14.09.2026).
18. Pearl J. Causality: Models, Reasoning, and Inference. – 2nd ed. – Cambridge: Cambridge University Press, 2009. – 484 P.
19. Pei C. UModel: An Agent-Ready Observability Data Modeling Method at Scale / C. Pei [et al.]. // arXiv preprint arXiv:2606.04799. – 2026. (дата обращения: 14.09.2026).

20. Pei C. Flow-of-Action: SOP Enhanced LLM-Based Multi-Agent System for Root Cause Analysis / C. Pei [et al.] // Companion Proceedings of the ACM on Web Conference. – 2025. – P. 422–431. – DOI: 10.1145/3701716.3715225.
21. Pham L. BARO: Robust Root Cause Analysis for Microservices via Multivariate Bayesian Online Change Point Detection / L. Pham, H. Ha, H. Zhang // Proceedings of the ACM on Software Engineering. – 2024. – Vol. 1. – P. 2214–2237. – DOI: 10.1145/3660805.
22. Pham L. RCAEval: A Benchmark for Root Cause Analysis of Microservice Systems with Telemetry Data / L. Pham [et al.] // Companion Proceedings of the ACM on Web Conference. – 2025. – P. 777–780. – DOI: 10.1145/3701716.3715290.
23. Prometheus: Monitoring System and Time Series Database – URL:. – 2012. (дата обращения: 14.09.2026).
24. Qin Y. ToolLLM: Facilitating Large Language Models to Master 16000+ Real-World APIs / Y. Qin [et al.] // International Conference on Learning Representations. – 2024.
25. Reimers N. Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks / N. Reimers, I. Gurevych // Proceedings of the. – 2019. – P. 3982–3992. – DOI: 10.18653/v1/D19-1410.
26. Sun Y. TrioXpert: An Automated Incident Management Framework for Microservice Systems / Y. Sun [et al.] // IEEE/ACM 40th International Conference on Automated Software Engineering. – 2025. – P. 3239–3250.
27. Wei J. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models / J. Wei [et al.] // Advances in Neural Information Processing Systems. – 2022. – Vol. 35. – P. 24824–24837.
28. Wu L. MicroRCA: Root Cause Localization of Performance Issues in Microservices / L. Wu [et al.] // IEEE/IFIP Network Operations and Management Symposium. – 2020. – P. 1–9. – DOI: 10.1109/NOMS47738.2020.9110353.
29. Xu J. OpenRCA: Can Large Language Models Locate the Root Cause of Software Failures? / J. Xu [et al.] // The Thirteenth International Conference on Learning Representations. – 2025.
30. Yao S. ReAct: Synergizing Reasoning and Acting in Language Models / S. Yao [et al.] // International Conference on Learning Representations. – 2023.
31. Zhang S. Robust Failure Diagnosis of Microservice Systems through Multimodal Data / S. Zhang [et al.] // IEEE Transactions on Services Computing. – 2023. – Vol. 16, No. 6. – P. 3851–3864.
32. Zhang S. Failure Diagnosis in Microservice Systems: A Comprehensive Survey and Analysis / S. Zhang [et al.] // ACM Transactions on Software Engineering and Methodology. – 2025. – Vol. 35, No. 1. – P. 1–55. – DOI: 10.1145/3715005.
33. Zhang W. mABC: Multi-Agent Blockchain-Inspired Collaboration for Root Cause Analysis in Microservices Architecture / W. Zhang [et al.] // Findings of the Association for Computational Linguistics: EMNLP. – 2024. – P. 4017–4033. – DOI: 10.18653/v1/2024.findings-emnlp.232.

34. Zhou X. D-Bot: Database Diagnosis System Using Large Language Models / X. Zhou [et al.] // Proceedings of the VLDB Endowment. – 2024. – Vol. 17, No. 10. – P. 2514–2527. – DOI: 10.14778/3675034.3675043.

INDUSTRY AND MANUFACTURING

CHEMICAL TECHNOLOGIES AND CHEMICAL INDUSTRY

Статья поступила в редакцию: 10.08.2026
Статья принята к публикации: 16.08.2026
Статья опубликована: 28.09.2026

УДК 547.314.2:54 – 44

DOI: 10.32743/UniTech.2026.150.9.23304

Catalytic synthesis of acetic aldehyde from acetylene

Omanov Behruzjon Shuhrat o‘g‘li
doctor of Philosophy of Technical Sciences, Professor, Navoi State University
Uzbekistan, Navoi

Dosimbetova Sevara Gulmirzayevna
Doctoral student of the Department of Chemistry
Navoi State University
Uzbekistan, Navoi

Каталитический синтез уксусного альдегида из ацетилена

Оманов Бехрузжон Шухрат угли
д-р техн. наук, проф.
Навоийский государственный университет
Республика Узбекистан, г. Навои

Досимбетова Севара Гулмирзаевна
аспирант кафедры химии
Навоийский государственный университет
Республика Узбекистан, г. Навои

Abstract

The objective of this work is to develop and evaluate active, selective, stable, and long-lasting catalysts for acetaldehyde production processes and to determine the main factors influencing their catalytic performance. The effects of catalyst composition, preparation method, active components, support materials, and operating conditions on activity, selectivity, stability, and service life were

---

Библиографическое описание: Оманов Б.Ш.У., Досимбетова С.Г. Каталитический синтез уксусного альдегида из ацетилена // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23304

For citation: Omanov B.S.O., Dosimbetova S.G. Catalytic Synthesis of Acetic Aldehyde from Acetylene // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23304

analyzed. Particular attention was paid to metal oxide and modified multicomponent catalytic systems because of their high thermal stability, catalytic activity, and resistance to deactivation. The results indicate that highly active catalysts can improve raw material conversion and increase acetaldehyde yield, whereas highly selective systems reduce the formation of undesirable by-products. Stable and durable catalysts can also decrease energy consumption, production costs, and the frequency of catalyst replacement. In addition, the possibility of catalyst regeneration is an important factor in increasing the economic efficiency of industrial processes. Environmental aspects of catalyst application were also considered. The use of low-toxic, environmentally friendly, and regenerable catalysts can reduce hazardous waste and minimize adverse environmental impacts. Thus, the development of efficient, selective, durable, and recyclable catalytic systems is a promising approach for improving the technological efficiency, environmental safety, and sustainability of acetaldehyde production.

Аннотация

Целью данной работы является разработка и оценка активных, селективных, стабильных и долговечных катализаторов для процессов производства ацетальдегида, а также определение основных факторов, влияющих на их каталитические характеристики. Было проанализировано влияние состава катализатора, метода его приготовления, активных компонентов, материалов-носителей и условий проведения процесса на активность, селективность, стабильность и срок службы катализаторов. Особое внимание уделено металлоксидным и модифицированным многокомпонентным каталитическим системам благодаря их высокой термической стабильности, каталитической активности и устойчивости к дезактивации. Результаты показывают, что высокоактивные катализаторы способны повышать степень конверсии исходного сырья и увеличивать выход ацетальдегида, тогда как высокоселективные системы способствуют снижению образования нежелательных побочных продуктов. Стабильные и долговечные катализаторы также позволяют снизить энергопотребление, производственные затраты и частоту их замены. Кроме того, возможность регенерации катализаторов является важным фактором повышения экономической эффективности промышленных процессов. Также были рассмотрены экологические аспекты применения катализаторов. Использование малотоксичных, экологически безопасных и регенерируемых катализаторов способствует сокращению объёма опасных отходов и минимизации негативного воздействия на окружающую среду. Таким образом, разработка эффективных, селективных, долговечных и пригодных к повторному использованию каталитических систем является перспективным направлением повышения технологической эффективности, экологической безопасности и устойчивости производства ацетальдегида.

Keywords: acetaldehyde, catalyst, catalytic activity, selectivity, metal oxides, chemical industry, stability.

Ключевые слова: ацетальдегид, катализатор, каталитическая активность, селективность, оксиды металлов, химическая промышленность, стабильность.

Introduction

The Kucherov reaction is a method for synthesizing acetaldehyde by the hydration of acetylene with water in the presence of a mercury (II) oxide catalyst dissolved in sulfuric acid. This process is one of the most important reactions in the organic synthesis industry and is based on the electrophilic addition reaction of acetylene. The reaction initially produces vinyl alcohol, which is converted to acetaldehyde by tautomerization. The process is usually carried out in an acidic environment and at a certain temperature [1-2].

The mechanism of this reaction is still unclear. It is believed that the following reactions occur in the presence of mercuric sulfate. Based on the following mechanism, an unsaturated alcohol is regrouped into an oxo compound.

The process of acetylene hydration is mainly carried out in industry on a continuous technological scheme. In this case, acetylene gas is passed through a catalyst solution – that is, through a solution of mercury (II) sulfate in sulfuric acid (to ensure good mixing of gas and liquid and efficient reaction) is carried in the form of tiny gas bubblesDuring the process, acetylene reacts with water to initially form vinyl alcohol, which quickly tautomerizes to acetaldehyde. This method is one of the important technological processes for obtaining acetaldehyde in the organic synthesis industry [3-4].

The concentration of sulfuric acid in the catalyst solution in various technological devices is usually in the range of 6–35 %. With an increase in the acid concentration, the hydration reaction of acetylene accelerates and the volume of the catalyst solution decreases. At the same time, if the acidity of the medium is too high, a number of additional reactions occur. In particular, the formation of crotonaldehyde, tarring of organic products, and the formation of resinous substances are observed. These conditions reduce the productivity of the process and lead to contamination of the equipment [5-6].

The content of mercury (II) oxide in the catalyst usually does not exceed 1 % and is often used in a lower concentration. The hydration process is mainly carried out in the temperature range of 75–100 °C. With increasing temperature, the reaction rate increases, but at the same time the degree of tarring of acetylene and the resulting products also increases. Therefore, the optimal temperature is selected depending on the catalyst composition, the design of the apparatus and the volumetric velocity of acetylene [7-8].

Experimental and Discussion Part

In the preparation of the catalyst, $Zn(NO_3)_2{\cdot}6H_2O$, $Mg(NO_3)_2{\cdot}6H_2O$, $Fe(NO_3)_3{\cdot}6H_2O$, $Mn(NO_3)_2{\cdot}6H_2O$ and $ZrO(NO_3)_2$ salts were used as sources of active components. The mass fractions of these salts in the initial catalytic composition were taken in the ratio of 5:3:5:3:1, respectively. These ratios represent the mass percentages of the initial salts, not the oxides [9-10].

Accordingly, to prepare 100 g of the initial catalytic composition, 5.0 g of $Zn(NO_3)_2{\cdot}6H_2O$, 3.0 g of $Mg(NO_3)_2{\cdot}6H_2O$, 5.0 g of $Fe(NO_3)_3{\cdot}6H_2O$, 3.0 g of $Mn(NO_3)_2{\cdot}6H_2O$, and 1.0 g of $ZrO(NO_3)_2$ were taken. Since the total amount of salts was 17.0 g, the remaining 83.0 g of the composition corresponds to the contribution of activated bentonite (YKS) used as a carrier.

Table 1.

Amount of components per 100 g of starting composition

No.	Component	Mass fraction, %	Quantity, g
1	$Zn(NO_3)_2 \cdot 6H_2O$	5	5.0
2	$Mg(NO_3)_2 \cdot 6H_2O$	3	3.0
3	$Fe(NO_3)_3 \cdot 6H_2O$	5	5.0
4	$Mn(NO_3)_2 \cdot 6H_2O$	3	3.0
5	$ZrO(NO_3)_2$	1	1.0
6	Activated bentonite (HSZ)	83	83.0

Initially, 83.0 g of activated bentonite (HSZ) was prepared. Metal salts, which are precursors of the active components, were weighed on an analytical balance at specified masses and dissolved in distilled water separately or under compatible dissolution conditions. The resulting salt solutions were gradually introduced into the UCS-based system and continuously mixed using a magnetic stirrer until a homogeneous suspension was formed [9-10].

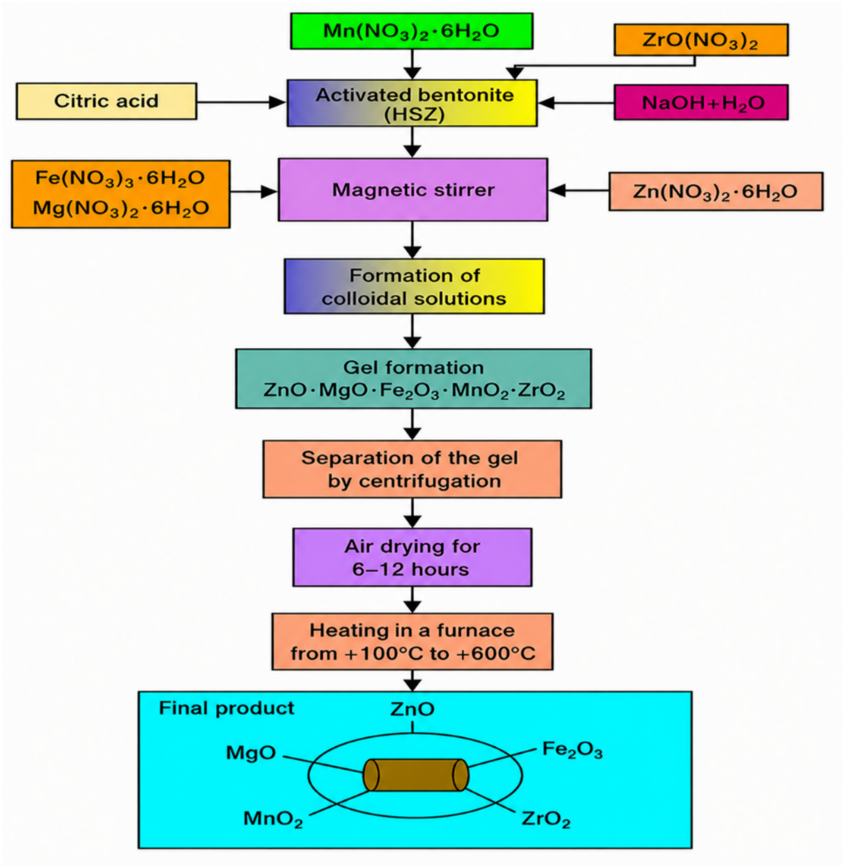

***Figure 1.** $ZnO:MgO:Fe_2O_3:MnO_2:ZrO_2$/ HSZ nanoparticle synthesis scheme*

During the mixing process, a uniform distribution of metal ions on the carrier surface was ensured and a colloidal system was formed. In the next step, the system was converted to a gel state, forming a homogeneous gel containing Zn, Mg, Fe, Mn and Zr components. The resulting gel was separated from the liquid phase by centrifugation. The separated gel was air-dried under atmospheric conditions for 6–12 h. The dried sample was initially kept at low temperature, then the temperature was gradually increased and thermally treated up to 600 °C. During the thermal treatment, the dehydration and decomposition of the precursor salts resulted in the formation of catalytically active oxide phases, which ensured the formation of a multicomponent catalytic system consisting of $ZnO·MgO·Fe_2O_3·MnO_2·ZrO_2$ on the surface of the HSZ.

Thus, the mass composition of the initial components is $Zn(NO_3)_2·6H_2O:Mg(NO_3)_2·6H_2O:Fe(NO_3)_3·6H_2O:Mn(NO_3)_2·6H_2O:ZrO(NO_3)_2$

A catalytic composition with the ratio of :HSZ = 5:3:5:3:1:83 was prepared.

Results and Discussion

The effect of various factors (temperature, volume flow rate, acetylene:water ratio, catalyst composition, etc.) on the product yield and process selectivity in the catalytic hydration reaction of acetylene was studied. Table 2 presents the effect of the catalyst active components on the reaction products [10].

Table 2.

Effect of starting materials on catalyst activity in the catalytic hydration reaction of acetylene

No.	Catalyst composition	C_2H_2 of conversion, %		Selectivity %
		General	To acetaldehyde	
1	$ZnO:Fe_2O_3:MnO_2$/ HSZ	62.4	50.8	81.4
2	$ZnO:MgO:MnO_2$/ HSZ	63.2	51.9	82.1
3	$ZnO:MgO:Fe_2O_3$/ HSZ	70.2	60.3	85.8
4	$ZnO:Fe_2O_3:MnO_2:ZrO_2$/ HSZ	75.8	66.5	87.7
5	$ZnO:MgO:MnO_2:ZrO_2$/ HSZ	85.7	78.9	92
6	$ZnO:MgO:Fe_2O_3:MnO_2$/ HSZ	92.8	87.6	94.3
7	**$ZnO·MgO·Fe_2O_3·MnO_2·ZrO_2$/** HSZ	**97.7**	**93.5**	**95.7**
8	$ZnO:Co_2O_3:Cr_2O_3:MnO_2:ZrO_2$/ HSZ	90.2	84.2	93.3
9	$ZnO:Co_2O_3:Cr_2O_3:Mn_2O_3:ZrO_2$/ HSZ	84.3	76.2	90.3
10	$ZnO:Fe_2O_3:Cr_2O_3:Mn_2O_3:V_2O_5$/ HSZ	78.3	65.4	83.5

As can be seen from the data presented in the table, the activity of pure $ZnO:Fe_2O_3$: MnO_2/ HSZ is much higher. With its participation, the total conversion of acetylene is 62.4 %, and the selectivity of acetaldehyde is 81.4 %. The introduction of Co_2O_3 and Mn_2O_3 into this catalyst leads to a decrease in the yield of acetaldehyde. The introduction of MgO and ZrO_2 into the catalyst composition prepared from ZnO:Fe2O3/YKS has a different effect on the yield of acetaldehyde.

Studies have shown that the highest acetaldehyde yield is achieved when ZrO_2 is added to the ZnO:MgO: Fe_2O_3: MnO_2/ HSZ catalyst. The addition of ZrO_2 to the catalyst composition had a positive effect on the activity of the catalyst. With the addition of ZrO_2, the acetylene conversion increased from 92.8 % to 97.7 %, and the selectivity increased from 94.3 % to 95.7 %.

Selectivity indicates how much of the starting material (i.e., the converted material, in our Table 1, the converted acetylene) is converted into the target product (acetaldehyde).

$$S\% = \frac{93.5}{97.7} \cdot 100\% = 95.7\%$$

Based on the above, we selected a catalyst with high activity, selectivity, and productivity, with high selectivity and thermal stability, using a suspension method based on local raw materials.

When studying the effect of temperature on the yield of acetaldehyde formation in the presence of the newly developed catalyst No. 7, which has high activity and productivity, it was found that the optimal temperature for the acetylene hydration reaction is 330 – 340℃ (Table 3).

Table 3.

Effect of temperature on the yield of acetaldehyde and the selectivity of the process for the conversion of acetylene (cat. No. 7)

No.	**Temperature, ^{oh}C**	**Total acetylene c onversion, %**	**AcetaldehydeProduction, %**	**Selectivity S % to Acet-aldehyde**
1	10-250	66.2	47.6	71.9
2	255-275	79.4	62.2	78.3
3	280-300	87.2	73.2	83.9
4	310-325	89.9	77.7	86.4
5	**330-340**	**97.7**	**93.5**	**95.7**
6	345-375	90.1	82.4	91.4
7	380-400	82.4	70.5	85.5

As can be seen from the table, when the temperature reaches 330 – 340℃, the yield of acetaldehyde formation is 93.5 %, and the selectivity of the process towards acetaldehyde is 95.7 %.

The metal reactor (7) is 40 mm in diameter and 380 mm high, with a mesh substrate for the catalyst installed in the lower part. The upper flange cover is equipped with two fittings: a valve for supplying the steam-gas mixture (acetylene saturated with water vapor) and a thermowell (8). The lower flange-bottom part is equipped with fittings for removing the reaction products.

TEN-06 heating elements with a total power of 3.6 kV are evenly distributed around the reactor vessel (7), which are connected to a 220 V network through a control transformer (9). To prevent overheating, it is not allowed to supply a voltage of more than 220 V.

The reactor surface is insulated and wrapped to retain heat and ensure a stable temperature. The temperature in the reactor was measured by a THC thermocouple with signals output to the temperature control unit (11), the readings were recorded on a KSP potentiometer or indicating device.

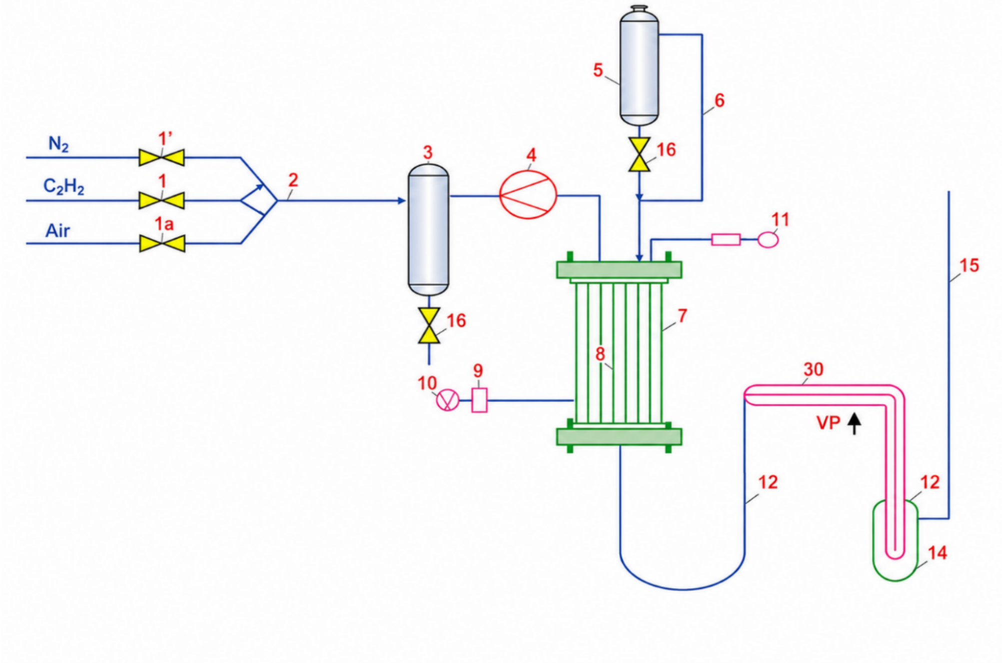

Figure 2. Scheme of acetaldehyde in the laboratory

1-needle valve; 2-flexible removable part; 3-drip eliminator;

4-rheometer; 5-water distribution funnel; 6-bypass line; 7-reactor; 8-thermowell; 9-transformer; 10-voltmeter; 11-block for temperature control in the reactor; 12-air cooler; 13- water-cooled refrigerator (Lebix); 14-reaction gas scrubber; 15 – exhaust gas discharge line; 16-tap

To carry out the synthesis, a certain amount of catalyst was loaded into the reactor. The catalyst was heated in a nitrogen stream. When the temperature reached 330 – 340°C, the valve on the mixed gas line of workshop No. 007 was opened (or acetylene concentrate from a cylinder), and the valve on the nitrogen line was closed. Water (5) with a temperature of ≈70 – 80°C was passed through the bypass line (6) to saturate it with water vapor. Then the gas saturated with vapor entered the reactor, where it reacted with the catalyst at a temperature of 330 – 340°C. During the reaction, the parameters of the technological regime, temperature conditions, and flow rates of the supplied components were measured.

The gaseous reaction products were cooled first in an air cooler (12) and then in a water-cooled refrigerator (Liebig) (13). The condensate and acetaldehyde vapors were absorbed in absorbers with xylene (14) placed in an ice bath (3 – 8°C). The reaction products obtained with the collected xylene were chromatographically analyzed for their composition. The inlet and outlet gases were periodically analyzed by gas chromatography.

After 120 hours, the supply of acetylene to the system is stopped. The system is purged with nitrogen and the catalyst is regenerated. Regeneration is carried out at a temperature of 330 – 340℃ for 12 hours. After the acetylene is squeezed out of the system, a nitrogen-air mixture is supplied. Regeneration is considered complete if the carbon dioxide content in the gases does not exceed 0.4 %.

Based on the results of the experiment and the qualitative and quantitative composition of the reaction products, the following mechanism of acetone formation in the presence of the given catalyst was proposed:

The formation of acetaldehyde can be expressed as the following summation equation:

$2C_2H_2 + 3H_2O \rightarrow CH_3COCH_3 + 2H_2 + CO_2$

We explain the formation of acetylene based on the following plausible mechanism: first, acetylene is hydrated and acetaldehyde is formed:

$C_2H_2 + H_2O \rightarrow CH_3CHO$

Conclusion

The development of active and long-lasting catalysts for the production of acetaldehyde is important for increasing the efficiency of the process. Highly active catalysts increase product yield, reduce energy consumption and production costs. Metal oxides and modified catalytic systems provide high stability and selectivity. Also, the use of environmentally friendly and regenerable catalysts is relevant for the modern chemical industry.

References

1. Wang, Q., Dai, B., Zhang, J. Catalytic Acetylene Hydration over the Zn/Zr-MCM Catalyst // Catalysis Communications. – 2022. – Vol. 165. – Art. 106455.
2. Chen, Z., Zhou, L., Wang, Q., Dai, B. High Catalytic Performance of Zn-Bi/MCM-41 in Acetylene Hydration to Acetaldehyde // Catalysis Communications. – 2022. – Vol. 172. – Art. 106539. – DOI: 10.1016/j.catcom.2022.106539.
3. Zhang, S., Wang, Q., Chen, Z., Dai, B. A Highly Effective Zinc-Methanesulfonic Acid Catalyst for Acetylene Hydration // Energy Advances. – 2022. – Vol. 1, No. 12. – P. 1021–1027. – DOI: 10.1039/D2YA00272H.
4. Liu, X., Long, Z., Wang, L., Dai, B. Effect of N, P Co-Doped Activated Carbon Supported Cu-Based Catalyst for Acetylene Hydration // Molecular Catalysis. – 2022. – Vol. 522. – Art. 112223. – DOI: 10.1016/j.mcat.2022.112223.
5. Zhang, Q., Wang, X., Dai, B. Supported Ionic Liquid-Copper Catalyst for Synergistically Catalytic and Stable Acetylene Hydration // International Journal of Hydrogen Energy. – 2022. – Vol. 47.
6. Najafian, A., Cundari T.R. Computational Study of Acetylene Hydration by Bio-Inspired Group Six Catalyst Models // Polyhedron. – 2018. – Vol. 154. – P. 114–122.
7. Chen, X., Wang, B. Ligand-Regulated Cu Catalysts for Stable Acetylene Hydration // Molecular Catalysis. – 2024.
8. Zhou, L., Dai, B. Vacancy-Engineered Oxide Supports for Acetylene Hydration Catalysts // Chemical Engineering Journal. – 2025.
9. Икрамов А.И., Кулдошев А.К., Зиядуллаев А.Э. Сравнительные активности гетерогенных катализаторов на основе соединений металлов d-группы при гидратации ацетилена // Universum: технические науки. – 2022. – No. 12 (105).
10. Omanov B.Sh., Turobjonov S.M., Fayzullayev N.I., Safarova L.B. Catalyst development for acetylene catalytic hydrogenation synthesis using “wet” and suspension methods // Science and Innovation International Scientific Journal. Toshkent. – 2025. – Vol. 4, No. 3. – P. 4–9. – URL: https://www.scientists.uz/view?id=8757. (accessed 27.09.2026)

Статья поступила в редакцию: 04.09.2026
Статья принята к публикации: 15.09.2026
Статья опубликована: 28.09.2026

УДК 661.632.14:631.893

DOI: 10.32743/UniTech.2026.150.9.23408

Investigation of the process for producing nitrogen-phosphorus fertilizers by processing Shorsu phosphorites with wet-process phosphoric acid activated with ammonium nitrate

Rustamov Ilyos
Doctoral student
Namangan state technical university
Republic of Uzbekistan, Namangan
E-mail: ilyosbekrustamov3773@gmail.com

Shamshidinov Israiljon
doctor of technical sciences, professor
Namangan state technical university
Republic of Uzbekistan, Namangan
E-mail: israiljon2010@mail.ru

Kodirova Gulnoza
PhD, associate professor
Namangan state technical university
Republic of Uzbekistan, Namangan
E-mail: qgulnoza1911@gmail.com

Najmiddinov Rikhsitilla Yunusali ugli
PhD, associate professor
Namangan state technical university
Republic of Uzbekistan, Namangan
E-mail: najmiddinov9191@gmail.com

Sayfiddinov Osimkhon
Master's student Namangan state technical university
Republic of Uzbekistan, Namangan
E-mail: sosimxon04@gmail.com

Библиографическое описание: Исследование процесса получения азотно-фосфорных удобрений путем переработки фосфоритов Шорсу экстракционной фосфорной кислотой, активированной аммиачной селитрой // Universum: технические науки : электрон. научн. журн. Рустамов И.Т. [и др.]. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23408

For citation: Investigation of the process for producing nitrogen-phosphorus fertilizers by processing shorsu phosphorites with wet-process phosphoric acid activated with ammonium nitrate // Universum: Technical Sciences : electronic scientific journal. Rustamov I. et al. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23408

Исследование процесса получения азотно-фосфорных удобрений путем переработки фосфоритов Шорсу экстракционной фосфорной кислотой, активированной аммиачной селитрой

Рустамов Илёс Тухтаназарович
докторант
Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Шамшидинов Исраилжон Тургунович
д-р техн. наук, проф.
Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Кодирова Гулноза Кодиржоновна
PhD, доц.
Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Нажмиддинов Рихситилла Юнусали угли
PhD, доц.
Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Сайфиддинов Осимхон Обиддин-угли
магистрант
Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Abstract

This study investigates the production of complex nitrogen-phosphorus fertilizers containing calcium and magnesium from low-grade phosphate rock from the Shorsu deposit (Fergana Region, Uzbekistan), containing 6,87 % P_2O_5. To enhance the efficiency of raw material decomposition, wet-process phosphoric acid (WPA) containing 19,50 % P_2O_5, produced from washed calcined phosphate concentrate of the Central Kyzylkum deposits and activated with ammonium nitrate (NH_4NO_3), was used. The effects of the phosphorite-to-WPA mass ratios (100:80 and 100:100), neutralization of the slurry with gaseous ammonia to pH 3,1 – 3,5, and drying at 100 – 105°C on the product composition and the degree of conversion of phosphates into plant-available forms [K(avail.)] were investigated. The optimal process conditions were determined as follows: phosphorite-to-WPA (100 % P_2O_5 basis) ratio of 100:100, temperature of 80°C, and process duration of 90 – 120 min. Under these conditions, the resulting concentrated complex fertilizer contained 43,63 % total P_2O_5, 42,15 % available P_2O_5, 2,85 % N, 6,46 % CaO, and 4,56 % MgO, while the relative content of available phosphates with respect to total phosphates [K(avail.)] reached 96,61 %. The results demonstrate the feasibility of efficiently utilizing low-grade Shorsu phosphorites for the production of multicomponent fertilizers containing nitrogen, phosphorus, calcium, and magnesium.

Аннотация

В данной работе исследован процесс получения сложных азотно-фосфорных удобрений, содержащих кальций и магний, из низкосортного фосфоритного сырья месторождения Шорсу (Ферганская область, Узбекистан) с содержанием P_2O_5 6,87 %. Для повышения эффективности разложения сырья использовалась экстракционная фосфорная кислота (19,50 % P_2O_5), полученная на основе мытого обожженного фосфоконцентрата Центральных Кызылкумов и активированная нитратом аммония (NH_4NO_3). В исследовании изучено влияние массовых соотношений фосфорит: ЭФК (100:80 и 100:100), нейтрализации пульпы газообразным аммиаком до pH=3,1 – 3,5 и сушки при температуре 100 – 105°C на состав продукта и степень перехода фосфатов в усвояемую форму [K(усв.)]. Определены оптимальные условия процесса: соотношение фосфорит:ЭФК (100 % P_2O_5) — 100:100, температура – 80°C, продолжительность – 90 – 120 минут. При этом в составе полученного концентрированного сложного удобрения содержание P_2O_5общ. составило 43,63 %, P_2O_5усв. – 42,15 %, N – 2,85 %, CaO – 6,46 % и MgO – 4,56 %, а относительное содержание усвояемой формы фосфатов к их общей форме [K(усв.)] достигло 96,61 %. Результаты демонстрируют возможность эффективного использования бедных фосфоритов Шорсу для производства многокомпонентных удобрений, содержащих азот, фосфор, кальций и магний.

Keywords: phosphorite, low-grade phosphate rock, extraction phosphoric acid, ammonium nitrate, activation, ammoniation, retrogradation, nitrogen-phosphorus fertilizer, calcium- and magnesium-containing fertilizer.

Ключевые слова: фосфорит, низкосортное фосфатное сырье, экстракционная фосфорная кислота, нитрат аммония, активация, аммонизация, ретроградация, азотно-фосфорное удобрение, кальций- и магнийсодержащее удобрение.

Introduction

Worldwide, including in the Republic of Uzbekistan, the demand for complex fertilizers containing not only the primary macronutrients (N, P, and K) but also secondary nutrients such as calcium and magnesium is steadily increasing. Such multicomponent fertilizers are of particular importance for increasing crop yields and improving soil fertility [10; 11; 17]. At the same time, due to the annual depletion of conventional high-grade phosphate resources, there is growing interest in utilizing relatively poor, compositionally complex, and low-grade phosphate ores for the production of phosphorus-containing fertilizers [12; 14-16].

The industrial production of wet-process phosphoric acid (WPA) from low-grade phosphate rocks containing approximately 20 % P_2O_5 is economically inefficient due to the high consumption of reagents, the large amount of impurities, and the need for additional processing costs. Consequently, such phosphate raw materials are used only to a limited extent, mainly for the production of single superphosphate. However, conventional single superphosphate produced from low-grade phosphate rocks is characterized by a relatively low nutrient content (typically about 9 – 11 % P_2O_5 and 1 – 2 % N). This reduces the efficiency of fertilizer production and application and also increases the relative costs associated with its transportation, packaging, and storage [18].

The partial neutralization of wet-process phosphoric acid obtained from phosphate rocks with a high content of magnesium and other foreign impurities negatively affects product quality. When secondary phosphate is processed with such acid, its decomposition rate does not exceed 50%, resulting in a fertilizer that fails to meet standard requirements. To increase this indicator, it is necessary to introduce 2–3 times more acid into the system compared to the stoichiometric norm, or to employ methods for its pre-activation. However, an increase in the acid norm raises the amount of free acid in the finished product, deteriorating its physical and mechanical properties. To eliminate this problem and improve product quality, a technology is proposed that involves neutralizing the acidic slurry formed during the decomposition stage with gaseous ammonia, followed by granulation and drying [17; 18].

To increase the reactivity of wet-process phosphoric acid used in the processing of natural phosphates, various chemical additives are introduced into its composition. Strong mineral acids, such as nitric and sulfuric acids, or salts, such as ammonium nitrate and ammonium sulfate, can be used as such activating components. Research has investigated in detail the processes of decomposing low-grade phosphorites with a low phosphorus pentoxide (P_2O_5) content (13 % or more) using WPA activated through these methods, as well as obtaining high-quality phosphorus fertilizers from them [10; 11].

Phosphate rocks with an extremely poor chemical composition (less than 13 % P_2O) also occur in nature. Due to such a low content of phosphorus pentoxide and the high amount of foreign impurities in the raw material, the processes of their direct chemical processing into mineral fertilizers have not been sufficiently studied to date. This determines the necessity of the complex utilization of the local raw material base and the development of new technological solutions for activating such low-grade ores.

Significant reserves of extremely low-grade phosphate ore are concentrated in the Shorsu region of the Uzbekistan District, Fergana Region, Uzbekistan. The P_2O_5 content of these ores is only 2 – 8 %, which substantially limits their direct use in conventional phosphorus fertilizer production technologies. The phosphate rock investigated in the present study contains 6,87 % P_2O_5, along with significant amounts of calcium and magnesium compounds. Therefore, the development of an efficient processing method capable of converting this low-grade local phosphate raw material into a concentrated multicomponent fertilizer is of both scientific and practical importance.

The present study investigates a technological approach based on the decomposition of Shorsu phosphorite using wet-process phosphoric acid produced from washed calcined phosphate concentrate (WCPC) of the Central Kyzylkum deposits and pre-activated with ammonium nitrate (NH_4NO_3).

The processes of WPA activation in the presence of ammonium nitrate have been studied in detail previously. The WPA used in the present study contained 19,50 % P_2O_5, and ammonium nitrate was added at an amount corresponding to 1,5 % of the acid mass for its activation. WPA activated with ammonium nitrate is used to intensify the decomposition of relatively low-grade phosphate raw materials. Subsequent ammoniation and drying of the acidic phosphate slurry obtained by decomposing the phosphate raw material with the activated acid are employed to optimize the chemical composition of the product and increase its nitrogen content while maintaining a high proportion of plant-available phosphates.

Accordingly, the aim of this study is to determine the technological conditions for processing extremely low-grade Shorsu phosphorite using WPA activated with ammonium nitrate and to evaluate the effects of the phosphorite-to-WPA ratio, ammoniation, and phosphate slurry drying on the degree of conversion of phosphates into plant-available forms and on the chemical composition of the resulting complex fertilizer containing calcium and magnesium.

Materials and Methods

Raw materials and their preparation. Low-grade phosphate ore from the Shorsu deposit, located in the Uzbekistan District of the Fergana Region, Uzbekistan, was used as a secondary phosphate raw material. Its chemical composition was as follows (wt.%): P_2O_5 = 6,87; CaO = 19,60; MgO = 4,20; CO_2 = 13,63; K_2O = 1,97; Fe_2O_3 = 0,48; FeO = 2,24; Al_2O_3 = 0,37; SO_3 = 0,52, and other components. To increase the specific contact surface area of the raw material during acid decomposition, the phosphorite was ground and sieved to obtain a fraction with a particle size of less than 0,25 mm.

Wet-process phosphoric acid (WPA), produced industrially by the dihydrate process from washed calcined phosphate concentrate (WCPC) of the Central Kyzylkum deposits, was used for the decomposition of the secondary phosphate raw material. The WPA had the following chemical composition (wt.%): P_2O_5 = 19,50; CaO = 0,28; MgO = 0,85; Fe_2O_3 = 0,57; Al_2O_3 = 0,63; F = 1,18; SO_3 = 2,67. Prior to phosphorite decomposition, the WPA was chemically activated. A commercial-grade sample of dry ammonium nitrate (NH_4NO_3) was used as the activating additive.

Experimental procedure. The production of a complex nitrogen-phosphorus fertilizer containing the macronutrients calcium and magnesium was carried out under laboratory conditions using a method analogous to the continuous process for double superphosphate production. The experimental procedure comprised the stages of acid activation, phosphorite decomposition, ammoniation of the phosphate slurry, evaporation of the ammoniated slurry, followed by its drying and granulation by mixing with recycle material.

Activation of WPA. A predetermined amount of NH_4NO_3 was added to the initial WPA, after which the mixture was continuously stirred until a homogeneous solution was obtained.

Acid decomposition of phosphorite. A three-necked flask was placed in a water bath maintained at 80°C. A stirrer equipped with a liquid seal was inserted through the central neck of the flask and driven by an electric motor. A water-cooled reflux condenser was fitted into the second neck. A predetermined amount of WPA was introduced through the third neck using a dropping funnel [at a phosphorite:WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:80 – 100, corresponding to 100 – 120 % of the stoichiometric amount required for the formation of monocalcium phosphate from phosphorite]. A predetermined amount of ammonium nitrate was then added through the funnel, and the mixture was stirred until the salt was completely dissolved. Subsequently, a pre-weighed amount of the secondary phosphate raw material – Shorsu phosphorite – was introduced through the same neck. After the addition of phosphorite was completed, a thermometer was inserted into the same neck of the flask. The phosphate slurry was stirred for 90 – 120 min, with the reaction time measured from the beginning of phosphorite addition to the activated WPA.

Ammoniation of the phosphate slurry. To enrich the product with nitrogen, the acidic superphosphate slurry obtained after decomposition of the secondary phosphate raw material with activated WPA was neutralized with gaseous ammonia (NH_3) to a pH of 3,1-3,5.

Evaporation, drying, and granulation of the ammoniated slurry. The resulting ammoniated slurry was evaporated until the residual moisture content reached 20 – 25 %. One portion of the evaporated slurry was dried at 100 – 105°C for use as recycle material. It was then mixed with the remaining portion of the slurry until the moisture content of the resulting mass reached 8 – 10 %. The resulting mass was granulated and dried in a drying oven at 100 – 105°C. The granules of the final product were sieved, and the commercial-size fraction of 2 – 4 mm was collected. The oversized fraction (>4 mm) was crushed and mixed with the powder fraction (<1 mm) generated during sieving; this fine material was returned to the process as recycle material.

Chemical analysis and calculation of the degree of conversion of phosphates into plant-available [K(avail.)] and water-soluble [K(w.s.)] forms. The components present in the products and slurries obtained by decomposition of natural phosphorite with activated WPA (wet-process phosphoric acid) and subsequent ammoniation, namely total, available, and water-soluble P_2O_5, free P_2O_5 [1, 7, 9], CaO and MgO [3, 9], SO_3 [5, 9, 13], N [6, 8, 9], Fe_2O_3, Al_2O_3, F and K_2O [9], moisture (H_2O) [2, 9], as well as the pH of the slurries [4; 9], were determined using analytical methods based on the applicable standards.

The efficiency of phosphate conversion into plant-available and water-soluble forms was evaluated in terms of the degrees of conversion into the available and water-soluble forms [K(avail.) and K(w.s.), respectively], calculated using the following equations:

$$K_{avail.} = \frac{P_2O_{5avail.}}{P_2O_{5total}} \times 100$$

$$K_{w.s.} = \frac{P_2O_{5w.s.}}{P_2O_{5total}} \times 100$$

where: P_2O_5avail. and P_2O_5w.s. are the contents of plant-available P_2O_5 and water-soluble P_2O_5 in the product, respectively; P_2O_5total is the total P_2O_5 content in the fertilizer.

Results and discussion. The technological parameters of the process for producing calcium- and magnesium-containing nitrogen-phosphorus fertilizers using low-grade phosphorite (6,87 % P_2O_5), as well as the chemical composition of the resulting products, are presented in Tables 1 and 2.

Initially, a comparative analysis was carried out to evaluate the chemical characteristics of the initial acidic slurry obtained after the phosphoric acid decomposition of Shorsu phosphorite and the intermediate ammoniated slurry formed during its subsequent neutralization with gaseous ammonia.

As shown in Table 1, when activated WPA containing 19,50 % P_2O_5 was used at a phosphorite-to-WPA mass ratio of 100:80, after decomposition for 1,5 – 2,0 h (90 – 120 min), the degree of conversion of phosphates into the plant-available form [K(avail.)] reached 95,90 %.

Table 1.

Technological parameters of Shorsu phosphorite decomposition with activated wet-process phosphoric acid and the chemical composition of the resulting phosphate slurry

Parameters and components	Chemical composition of the slurry components (wt.%) and technological process parameters			
	Acidic slurry		Ammoniated slurry	
Mass ratio of phosphorite to WPA (calculated on a 100% P_2O_5 basis)	100:80	100:100	100:80	100:100
WPA rate relative to the stoichiometric requirement, %	100	125	100	125
Slurry pH	3,2	2,4	3,5	3,1
P_2O_5 (total)	19,26	19,81	19,98	20,86
P_2O_5 (available)	18,47	19,33	18,46	20,12
P_2O_5 (water-soluble)	14,44	16,75	13,77	17,09
K_2O (total)	0,17	0,19	0,18	0,20
CaO (total)	3,82	3,17	3,76	3,09
MgO (total)	2,31	2,28	2,26	2,19
N (total)	0,43	0,45	0,57	1,35
H_2O	54,46	56,11	52,76	53,68
(P_2O_5avail./P_2O_5total)·100 [K(avail.)], %	**95,90**	**97,57**	**93,89**	**96,45**
(P_2O_5w.s./P_2O_5total)·100 [K(w.s.)], %	**74,97**	**84,38**	**68,92**	**81,93**

During the neutralization with gaseous ammonia of the slurry obtained at a phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:80, reverse chemical transformations associated with phosphorus retrogradation (conversion of phosphates into plant-unavailable forms) occurred to a certain extent. As the slurry medium changed with increasing pH, the readily plant-available mono- and dicalcium phosphates were partially converted into sparingly soluble tricalcium phosphate, resulting in a decrease in the proportion of plant-available phosphates. During ammoniation of the acidic slurry, the pH increased from 3,2 to 3,5, while the P_2O_5 content increased from 19,26 to 19,98 %. As a result of the retrogradation of 2,01 % of the plant-available phosphates, K(avail.) decreased from 95,90 to 93,89 %. At the same time, the proportion of water-soluble phosphates [K(w.s.)] also decreased markedly from 74,97 to 68,92 %.

Increasing the WPA rate had a positive effect on the degree of phosphorite decomposition. When the acid rate was increased to 1,25 times the stoichiometric requirement, i.e., when the phosphorite-to-WPA mass ratio was increased to 100:100, K(avail.) in the acidic slurry increased to 97,57 %, while K(w.s.) increased to 84,38 %. During neutralization of the acidic slurry with gaseous ammonia from an initial pH of 2,4 to a final pH of 3,1, despite the observed increases in total P_2O_5 from 19,81 to 20,86 %, plant-available P_2O_5 from 19,33 to 20,12 %, and water-soluble P_2O_5 from 16,75 to 17,09 %, the ratio of plant-available phosphates to total phosphates [K(avail.)] decreased from 97,57 to 96,45 %, while the proportion of water-soluble phosphates [K(w.s.)] decreased from 84,38 to 81,93 %.

The slurries obtained during the decomposition of natural phosphorite with activated WPA contained approximately 52 – 56 % water. Subsequently, the slurry was evaporated to reduce its water content to 20 – 25 %. After the addition of dry fertilizer as recycle material, the moisture content was further reduced to 8 – 10 %, after which the resulting mass was granulated and dried at 100 – 105°C.

During the drying of the acidic and ammoniated slurries obtained by decomposition of Shorsu phosphorites with activated WPA, the chemical equilibrium in the system underwent further changes. The experimental results obtained for the dried products are presented in detail in Table 2.

Table 2.

Chemical composition of acidic and ammoniated products obtained from Shorsu phosphorite and activated WPA and the technological process parameters

Parameters and components	Chemical composition of the product components (wt.%) and technological process parameters			
	Acidic product		Ammoniated product	
Mass ratio of phosphorite to WPA (calculated on a 100% P_2O_5 basis)	100:80	100:100	100:80	100:100
WPA rate relative to the stoichiometric requirement, %	100	125	100	125
Slurry pH	3,2	2,4	3,5	3,1
P_2O_5 (total)	41,40	44,50	41,46	43,63
P_2O_5 (available)	39,76	43,84	38,94	42,15
P_2O_5 (water-soluble)	28,15	35,81	25,70	34,24
P_2O_5 (free)	1,76	8,42	0,41	3,57
K_2O (total)	0,36	0,43	0,38	0,41
CaO (total)	8,24	7,12	7,80	6,46
MgO (total)	5,02	5,10	4,69	4,56
N (total)	0,92	1,03	1,18	2,85
H_2O	2,21	1,37	1,95	3,22
(P_2O_5avail./P_2O_5total)·100 [K(avail.)], %	**96,04**	**98,52**	**93,92**	**96,61**
(P_2O_5w.s./P_2O_5total)·100 [K(w.s.)], %	**68,00**	**80,47**	**61,99**	**78,48**

During the drying of phosphate slurries, further decomposition of the residual phosphorite present in the slurry may continue. As a result, depending on the amount of free acid remaining in the slurry, the proportion of plant-available P_2O_5 may increase slightly. When the acidic and ammoniated slurries obtained at a phosphorite-to-WPA ratio of 100:80 were dried at 100 – 105°C, the K(avail.) values of the resulting products reached 96,04 % and 93,92 %, respectively. Compared with the corresponding values for the initial slurries, these increases were only 0,14 % and 0,03 %, respectively. The contents of free P_2O_5 in the resulting acidic and ammoniated products were 1,76 % and 0,41 %, respectively.

Although the degree of conversion of phosphates into the plant-available form [K(avail.)] reached 93,92 % after ammoniation and subsequent drying of the slurry obtained by decomposition of natural phosphorite with activated WPA at a phosphorite-to-WPA mass ratio (calculated on a

100 % P_2O_5 basis) of 100:80, this value was below the established quality requirements. Therefore, the production of an ammoniated fertilizer under these conditions was considered technologically impractical.

In contrast, when the slurry obtained at the same phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:80 was dried without prior ammoniation, the K(avail.) value of the resulting product reached 96,04 %. In addition, the free P_2O_5 content of the product (1,76 %) met the established technological requirements. Therefore, under these conditions, direct drying of the acidic slurry provided a higher proportion of plant-available phosphates than neutralization followed by drying.

Thus, the acidic product obtained by decomposing Shorsu phosphorite with WPA activated with ammonium nitrate at a phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:80, followed by slurry evaporation, granulation, and drying, had the following chemical composition (wt.%): total P_2O_5 = 41,40; plant-available P_2O_5 = 39,76; water-soluble P_2O_5 = 28,15; free P_2O_5 = 1,76; K_2O = 0,36; CaO = 8,24; MgO = 5,02; N = 0,92; H_2O = 2,21, along with other components, and met the established technological requirements. The proportion of plant-available phosphates in the product, (P_2O_5avail./P_2O_5total) × 100, was 96,04 %, while the proportion of water-soluble phosphates, (P_2O_5w.s./P_2O_5total) × 100, was 68,0 %.

When the amount of activated WPA used for the decomposition of natural phosphorite was increased to 1,25 times the stoichiometric requirement [at a phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:100], the K(avail.) value of the resulting product increased by 2,48 percentage points, reaching 98,52 %. Nevertheless, the free P_2O_5 content of the final product was 8,42 %, which exceeded the established limit and therefore restricted the applicability of this product.

In this case, i.e., at a phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:100, the most favorable results were achieved by neutralizing the resulting acidic slurry with gaseous ammonia. After ammoniation, the K(avail.) value of the slurry was 96,45 %, and during subsequent drying it increased slightly to 96,61 %. At the same time, the free P_2O_5 content of the final product decreased to 3,57 %. The calcium- and magnesium-containing nitrogen-phosphorus fertilizer obtained under these optimal conditions demonstrated the most favorable combination of high phosphate availability and reduced residual free acidity.

Conclusion

The feasibility of producing a high-quality concentrated complex fertilizer using low-grade Shorsu phosphorite containing only 6,87 % P_2O_5 and WPA containing 19,50 % P_2O_5, activated with ammonium nitrate, was demonstrated.

It was established that at a phosphorite-to-activated WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:80, the ammoniated product was characterized by a relatively low degree of phosphate conversion into the plant-available form [K(avail.) = 93,92 %], indicating that the production of an ammoniated fertilizer under these conditions is technologically impractical. However, drying the acidic slurry without prior ammoniation resulted in a product with a K(avail.) value of 96,04 %.

The following optimal conditions were determined for producing a calcium- and magnesium-containing nitrogen-phosphorus fertilizer by decomposing Shorsu phosphorite with pre-activated WPA, followed by neutralization with gaseous ammonia, evaporation, drying, and granulation of the slurry: initial activated WPA concentration of approximately 19 % P_2O_5; WPA rate for phosphorite decomposition of 125 % of the stoichiometric requirement [corresponding to a phosphorite-to-WPA mass ratio (calculated on a 100 % P_2O_5 basis) of 100:100]; decomposition temperature of 80°C; process duration of 90 – 120 min; neutralization of the slurry with gaseous ammonia to pH 3,1 – 3,5; and drying and granulation of the slurry at 100 – 105°C.

The fertilizer obtained under these conditions had the following chemical composition (wt. %): total P_2O_5 = 43,63; plant-available P_2O_5 = 42,15; water-soluble P_2O_5 = 34,24; free P_2O_5 = 3,57; K_2O = 0,41; CaO = 6,46; MgO = 4,56; N = 2,85; and H_2O = 3,22. The proportion of plant-available phosphates in the product, (P_2O_5avail./P_2O_5total) × 100, reached 96,61 %, while the proportion of water-soluble phosphates, (P_2O_5w.s./P_2O_5total) × 100, was 78,48 %.

Thus, the proposed processing method provides an effective approach for converting extremely low-grade Shorsu phosphorite into a concentrated multicomponent fertilizer containing phosphorus, nitrogen, calcium, and magnesium.

References

1. ГОСТ 20851.2–75 (ИСО 5316–77, ИСО 6598–85, ИСО 7497–84). Удобрения минеральные: Методы определения фосфора. – М.: ИПК Изд-во стандартов, 1997. – 37 с.
2. ГОСТ 20851.4–75 Удобрения минеральные. Методы определения воды. – М.: ИПК Изд-во стандартов, 2000. – 5 с.
3. ГОСТ 24596.4–2015. Фосфаты кормовые. Метод определения кальция. – Москва: Стандартинформ, 2020. – 14 с.
4. ГОСТ 24596.5–2015. Фосфаты кормовые. Метод определения показателя активности водородных ионов. – Москва: Стандартинформ, 2020. – 12 с.
5. ГОСТ EN 15749-2013 (EN 15749:2009, IDT). Удобрения. Определение содержания сульфатов тремя разными методами. – Москва: Стандартинформ, 2013.
6. ГОСТ EN 15750-2016. Удобрения. Определение общего азота в удобрениях, содержащих азот только в нитратной, аммиачной и карбамидной формах, двумя различными методами. – М.: Стандартинформ, 2016. – 22 с.
7. ГОСТ EN 15959-2014. Удобрения. Определение экстрагируемого фосфора. – М.: Стандартинформ, 2014. – 16 с.
8. ГОСТ Р ИСО 15604-2019. Удобрения. Определение различных форм азота в одной и той же пробе, содержащей азот в виде нитратного, аммиачного азота, мочевины и цианамида. – М.: Стандартинформ, 2019. – 23 с.
9. Крашенинников С.А. Технический анализ и контроль в производстве неорганических веществ. – М.: Высшая школа, 1986. – 280 с.

10. Шамшидинов И.Т. Переработка фосфоритов Каратау и Центральных Кызылкумов на экстракционную фосфорную кислоту и концентрированные фосфорсодержащие удобрения. – М.: Издательство Lambert Academic Publishing, 2021. – 225 с.
11. Шамшидинов И.Т. Разработка усовершенствованной технологии производства экстракционной фосфорной кислоты и получения концентрированных фосфорсодержащих удобрений из фосфоритов каратау и Центральных Кызылкумов. Диссертация доктора технических наук. (DSc), 2017. – 193 с.
12. Allamuratova A.Zh., Erkaev A.U., Reimov A.M., Kazakhbaev S., Kurbiyazov D. Improving technology for producing complex fertilizer from low-grade phosphorites of the Central Kyzylkum // Chemical Technology, Control and Management. – 2024.
13. ISO 22887:2020. Fertilizers – Determination of total sulfur by high-temperature combustion. – ISO, 2020.
14. Kodirova G., Shamshidinov I., Sultonov B., Najmiddinov R., Mamurov B. Investigation of the Process of Purification of Wet-Process Phosphoric Acid and Production of Concentrated Phosphoric Fertilizers Based on It // Chemical Science International Journal. – 2021.
15. Reymov A., Namazov Sh. Nitrogen-phosphorous fertilizers on the base of concentrated ammonium nitrate solution and Central Kyzylkum phosphate raw material // Polish Journal of Chemical Technology. – 2014.
16. Shamshidinov I.T., Gafurov K.G., Ikramov M.M. Investigation on the phosphoric acid production from low grade phosphorites with high content of magnesium // Journal of Chemical Technology and Metallurgy. – 2016.
17. Shamshidinov I.T., Mirzakulov X.Ch., Melikulova G.E., Isroiljonova S.Q. Mineral o‘g‘itlar va tuzlar kimyoviy texnologiyasi. Darslik. 436 b. – T.: “Innovatsion rivojlanish nashriyot-matbaa uyi” nashriyoti, 2026.
18. Shamshidinov I.T. Noorganik moddalar va mineral o‘g‘itlar texnologiyasi. 324 b. – T.: Iqtisod-moliya, 2014.

Статья поступила в редакцию: 02.09.2026
Статья принята к публикации: 07.09.2026
Статья опубликована: 28.09.2026

УДК 661.86+553.65+338.2

DOI: 10.32743/UniTech.2026.150.9.23381

Processing of calcium nitrate solutions generated during nitric acid treatment of Central Kyzylkum phosphorites

Sultonov Bokhodir Elbekovich
doctor of Technical Sciences, Associate Professor
Namangan State University
Republic of Uzbekistan, Namangan
E-mail: bse-chemist-68@mail.ru

Giyasidinov Abduaziz Lutfidinovich
PhD, Associate Professor
Department of Chemical Engineering Namangan State Technical University
Republic of Uzbekistan, Namangan

Переработка растворов нитрата кальция, образующихся при азотнокислотной обработке фосфоритов Центральных Кызылкумов

Султонов Боходир Элбекович
д-р техн. наук, проф.
Наманганский государственный университет
Республика Узбекистан, г. Наманган

Гиясидинов Абдуазиз Лутфидинович
PhD, доц.
кафедры Химическая инженерия, Наманганский государственный технический университет
Республика Узбекистан, г. Наманган

Abstract

This study aimed to evaluate the processing of calcium nitrate solutions generated during nitric acid treatment of Central Kyzylkum phosphorites by cooling crystallization and evaporative concentration, and to determine technologically rational operating conditions for calcium nitrate recovery and concentration. A representative process solution containing 24.40 wt.% $Ca(NO_3)_2$ was used. The methodology comprised material-balance calculations based on 1000 g of the initial solution, stepwise cooling from +20 to -20 °C at 5 °C intervals, and evaporative concentration at

Библиографическое описание: Султонов Б.Э., Гиясидинов А.Л. Переработка растворов нитрата кальция, образующихся при азотнокислотной обработке фосфоритов Центральных Кызылкумов // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23381

For citation: Sultonov B.E., Giyasidinov A.L. Processing of calcium nitrate solutions generated during nitric acid treatment of central kyzylkum phosphorites // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23381

atmospheric pressure. During cooling, the calculated recovery of calcium nitrate into the crystalline phase increased as temperature decreased. Between 0 and -15 °C, recovery rose from 59.74 to 94.21 %. Further cooling to -20 °C increased recovery only to 95.04 %, corresponding to an additional 2.03 g of $Ca(NO_3)_2$ equivalent per 1000 g of initial solution; therefore, cooling below -15 °C was considered economically inefficient under the investigated conditions. During evaporation, removal of 531.58 g of water increased the $Ca(NO_3)_2$ concentration from 24.40 to 52.09 wt.%, while approximately 37 wt.% was identified as a technologically relevant intermediate concentration for subsequent conversion processes. The results indicate that cooling crystallization and evaporative concentration are complementary processing routes for obtaining calcium nitrate-containing crystalline products or concentrated solutions. These approaches promote more complete utilization of calcium and nitrate nitrogen during nitric acid processing of Central Kyzylkum phosphate raw materials.

Аннотация

Целью настоящего исследования являлась оценка переработки растворов нитрата кальция, образующихся при азотнокислотной обработке фосфоритов Центральных Кызылкумов, методами охлаждающей кристаллизации и выпарного концентрирования, а также определение технологически рациональных условий извлечения и концентрирования нитрата кальция. В качестве исходного технологического раствора использовали репрезентативный раствор, содержащий 24.40 мас.% $Ca(NO_3)_2$. Методология включала материально-балансовые расчеты на основе 1000 г исходного раствора, ступенчатое охлаждение от +20 до -20 °C с интервалом 5 °C и выпарное концентрирование при атмосферном давлении. В процессе охлаждения расчетная степень извлечения нитрата кальция в кристаллическую фазу возрастала по мере снижения температуры. В интервале от 0 до -15 °C степень извлечения увеличилась с 59.74 до 94.21 %. Дальнейшее охлаждение до -20 °C повысило степень извлечения лишь до 95.04 %, что соответствовало дополнительному извлечению 2,03 г $Ca(NO_3)_2$ в пересчете на 1000 г исходного раствора; поэтому охлаждение ниже -15 °C в исследованных условиях было признано экономически неэффективным. В процессе выпаривания удаление 531.58 г воды позволило повысить концентрацию $Ca(NO_3)_2$ с 24.40 до 52.09 мас.%, при этом концентрация около 37 мас.% была определена как технологически значимый промежуточный уровень для последующих процессов конверсии. Полученные результаты показывают, что охлаждающая кристаллизация и выпарное концентрирование являются взаимодополняющими технологическими направлениями для получения кальций-нитратсодержащих кристаллических продуктов или концентрированных растворов. Применение данных подходов способствует более полному использованию кальция и нитратного азота при азотнокислотной переработке фосфатного сырья Центральных Кызылкумов.

Keywords: Central Kyzylkum phosphorites, nitric acid processing, calcium nitrate, cooling crystallization, evaporative concentration, material balance, nitrogen–calcium fertilizer, resource utilization.

Ключевые слова: фосфориты Центральных Кызылкумов, азотнокислотная переработка, нитрат кальция, охлаждающая кристаллизация, выпарное концентрирование, материальный баланс, азотно-кальциевое удобрение, комплексное использование ресурсов.

Introduction

Phosphate rock is the principal mineral raw material used for the production of phosphoric acid and phosphate fertilizers. Its technological behavior during chemical processing is determined not only by the phosphorus content but also by the mineralogical and chemical composition of the accompanying components. Calcium is generally the dominant non-phosphorus constituent, while magnesium, iron, aluminum, fluorine, silica, carbonates, and other components may also occur in significant amounts. These constituents affect acid consumption, dissolution behavior, phase separation, filtration, and ultimately the composition and quality of the resulting products [1]. Consequently, the efficient processing of phosphate raw materials increasingly requires an integrated approach aimed not only at phosphorus recovery but also at the rational utilization of other valuable components of the feedstock. This consideration is particularly relevant to the phosphate resources of the Central Kyzylkum region. Along with conventional phosphate raw materials, mining, beneficiation, and thermal processing operations generate mineralized mass and phosphorite slurry characterized by relatively low phosphorus contents and high proportions of calcium- and carbonate-containing components. Recent studies have demonstrated that such Central Kyzylkum phosphate materials can be involved in nitric-acid-based processing and converted into phosphorus-containing fertilizer products [2,3]. These findings indicate that low-grade and off-specification phosphate materials should not necessarily be regarded solely as waste streams, but may represent secondary mineral resources suitable for further chemical processing. The selection of the acid used for phosphate-rock decomposition strongly influences the subsequent distribution of calcium within the technological system. In the conventional sulfuric acid route, calcium originating from the phosphate mineral matrix is predominantly separated as calcium sulfate, producing phosphogypsum as the major solid by-product of wet-process phosphoric acid production [4]. Although various approaches for phosphogypsum utilization have been proposed, its large-scale generation and the presence of residual phosphorus, fluorine, trace metals, and other impurities continue to complicate its comprehensive utilization [4]. Nitric acid processing provides a fundamentally different pathway because calcium released during phosphate-rock decomposition remains largely in the liquid phase in the form of soluble calcium nitrate rather than being precipitated as calcium sulfate [5,6]. This difference creates an opportunity to recover calcium and nitrogen as useful products within an integrated phosphate-processing scheme. Previous investigations of Central Kyzylkum phosphorites have demonstrated the feasibility of nitric acid treatment for producing phosphorus-containing fertilizers and fertilizer precipitates. In particular, mineralized mass from Central Kyzylkum has been processed with nitric acid to obtain phosphorus fertilizer products, and physicochemical characterization has confirmed substantial changes in the mineral and elemental composition of the material after treatment [2]. More recent work has also shown that mineralized mass and slurry phosphorite generated during thermal processing can be treated with nitric acid to produce activated nitrogen–phosphate fertilizer materials [3]. These studies provide an important technological basis for the nitric-acid processing of Central Kyzylkum phosphate resources. However, their principal emphasis has been placed on phosphate conversion and the properties of the resulting solid fertilizer products rather than on the subsequent recovery and utilization of the calcium nitrate-containing liquid phase.

The calcium nitrate solution formed during nitric acid processing represents a potentially valuable intermediate stream because it contains both calcium and nitrate nitrogen. The management of calcium nitrate is also an important technological issue in other nitric-acid-based phosphate-processing schemes, where its presence and separation can significantly influence downstream product quality and process configuration [6]. From a physicochemical standpoint, aqueous calcium nitrate systems exhibit strongly temperature- and concentration-dependent solid–liquid and vapor–liquid equilibria [7]. Therefore, water removal and temperature control can be used to modify the concentration state of the solution and promote the formation of calcium nitrate-containing crystalline phases. The feasibility of recovering calcium nitrate crystals from concentrated aqueous solutions has also been demonstrated using crystallization-based separation approaches [8]. These findings suggest that calcium nitrate-containing process liquors may be treated not only through chemical conversion but also through physical concentration and crystallization routes.

Despite the established potential of nitric acid for processing Central Kyzylkum phosphorites, comparatively little attention has been given to the direct treatment of the resulting calcium nitrate solutions by cooling crystallization and evaporative concentration. Evaluation of these routes is important for determining whether the calcium nitrate-containing liquid phase can be transformed into a concentrated or crystalline product suitable for further processing or fertilizer-related applications while simultaneously improving the overall utilization of calcium and nitrogen within the technological system.

Accordingly, the aim of the present study was to evaluate the processing of calcium nitrate solutions generated during nitric acid treatment of Central Kyzylkum phosphorites using two technological approaches: cooling crystallization and evaporative concentration. The study examined the effect of temperature on calcium nitrate recovery from solution and evaluated the material balance of progressive water removal during evaporation. On this basis, technological options for obtaining calcium nitrate-containing crystalline products and concentrated solutions intended for further processing or potential use as nitrogen–calcium fertilizer materials were assessed.

2. Materials and Methods

Calcium nitrate-containing process solutions

The calcium nitrate-containing solutions used in this study were obtained as the liquid phase generated during nitric acid processing of Central Kyzylkum phosphorite materials. In the preceding technological stage, the phosphate raw material was treated with nitric acid, after which the resulting nitrocalcium phosphate slurry was neutralized and separated into solid and liquid phases. The solid phase represented the phosphorus-containing fertilizer product, whereas the filtrate contained calcium predominantly in the form of soluble calcium nitrate.

Four representative filtrates obtained under different nitric acid consumption conditions and solid-phase separation modes were considered. The nitric acid dosage corresponded to 100 or 105 % of the stoichiometric requirement calculated with respect to dicalcium phosphate formation. The principal characteristics of the resulting calcium nitrate-containing solutions are summarized in Table 1.

Table 1.

Composition of calcium nitrate-containing solutions generated during nitric acid processing of Central Kyzylkum phosphorites

Processing condition	Nitric acid dosage, % of stoichiometric requirement	Water-soluble CaO, wt.%	N, wt.%	Equivalent $Ca(NO_3)_2$ concentration, wt.%
Without preliminary removal of insoluble residue	100	8.33	6.21	24.40
Without preliminary removal of insoluble residue	105	8.49	6.28	24.86
After removal of insoluble residue	100	8.39	6.24	24.57
After removal of insoluble residue	105	8.52	6.30	24.95

The solutions contained approximately 24.40-24.95 wt.% $Ca(NO_3)_2$ equivalent, indicating relatively small differences among the investigated process conditions. The detailed cooling and evaporation calculations presented in this study were based on the solution containing 24.40 wt.% $Ca(NO_3)_2$ equivalent, corresponding to a water-soluble CaO content of 8.33 wt.%.

Cooling crystallization

The calcium nitrate solution was cooled stepwise to evaluate the effect of temperature on calcium nitrate recovery from the liquid phase. The solution was cooled from +20 °C to -20 °C in 5 °C increments. The temperatures investigated were +20, +15, +10, +5, 0, -5, -10, -15, and -20 °C.

Material-balance calculations were based on 1000 g of the initial solution containing 24.40 wt.% $Ca(NO_3)_2$, equivalent to 244.0 g of $Ca(NO_3)_2$. At each temperature, we determined the amounts of calcium nitrate transferred to the crystalline phase and remaining in the mother liquor from the material balance.

Quantities transferred to the solid phase are reported as $Ca(NO_3)_2$ equivalent on an anhydrous-salt basis. No specific hydrate composition was assigned because independent phase characterization of the crystalline product was not available.

Evaporative concentration

The calcium nitrate solution was also concentrated by evaporation at atmospheric pressure. The initial solution contained 24.40 wt.% $Ca(NO_3)_2$ and 75.60 wt.% water.

For the material-balance calculations, we considered 1000 g of the initial solution, corresponding to 244.0 g of $Ca(NO_3)_2$ and 756.0 g of water. During progressive evaporation, calcium nitrate was treated as the nonvolatile dissolved component, and the decrease in solution mass was attributed to water removal.

At each concentration stage, we calculated the residual solution mass, the amount of evaporated water, the calcium nitrate concentration, and the amount of water remaining in the solution. The evaporation process was evaluated up to a $Ca(NO_3)_2$ concentration of approximately 52 wt.%.

Material-balance evaluation

Material-balance calculations evaluated both processing routes. For cooling crystallization, the initial calcium nitrate content was distributed between the crystalline phase and the mother liquor. For evaporative concentration, we evaluated the progressive removal of water and the resulting increase in calcium nitrate concentration.

All concentrations and mass-flow values were reported on a mass basis. Because the original dataset did not report replicate measurements or statistical variability, the results were treated as technological and material-balance data, with no inferential statistical analysis.

3. Results and Discussion

Effect of cooling temperature on calcium nitrate recovery

The effect of temperature on the distribution of calcium nitrate between the crystalline phase and the residual mother liquor was evaluated over the range from +20 to -20 °C. The results, calculated based on 1000 g of the initial solution containing 24.40 wt.% $Ca(NO_3)_2$ are presented in Table 2.

Table 2.

Effect of cooling temperature on the material balance of calcium nitrate recovery

Temperature, °C	$Ca(NO_3)_2$ Transferred to the crystalline phase, % of initial amount	$Ca(NO_3)_2$ in the crystalline phase, g	$Ca(NO_3)_2$ remaining in the mother liquor, % of initial amount	$Ca(NO_3)_2$ remaining in the mother liquor, g
+20	25.27	61.66	74.73	182.34
+15	31.02	75.69	68.98	168.31
+10	36.77	89.72	63.23	154.28
+5	48.26	117.75	51.74	126.25
0	59.74	145.77	40.26	98.23
-5	71.23	173.80	28.77	70.20
-10	82.72	201,84	17.28	42.16
-15	94.21	229.87	5.79	14.13
-20	95.04	231.90	4.96	12.10

As shown in Table 2, lowering the temperature progressively increased the calculated recovery of calcium nitrate into the crystalline phase. At +20 °C, 25.27 % of the initial $Ca(NO_3)_2$ equivalent was in the crystalline fraction, whereas cooling to 0 °C increased this value to 59.74 %. Further cooling produced a more pronounced recovery: 71.23 % at -5 °C, 82.72 % at -10 °C, and 94.21 % at -15 °C.

For a 1000 g initial solution, the starting amount of $Ca(NO_3)_2$ was 244.0 g. At -15 °C, 229.87 g of $Ca(NO_3)_2$ equivalent was transferred to the crystalline phase, leaving only 14.13 g in the mother liquor. When the temperature was further decreased to -20 °C, the calculated amount transferred to the crystalline phase increased only to 231.90 g, leaving 12.10 g in the liquid phase.

An important feature of the temperature dependence is the diminishing incremental recovery observed below -15 °C. Cooling from -10 to -15 °C increased calcium nitrate recovery by 11.49 percentage points, whereas a further decrease from -15 to -20 °C increased recovery by only 0.83 percentage points, corresponding to an additional 2.03 g of $Ca(NO_3)_2$ equivalent per 1000 g of initial solution. Thus, within the investigated temperature range, most of the achievable recovery had already been attained at -15 °C.

The observed temperature dependence is consistent with the strong influence of temperature on solid–liquid equilibria in the $Ca(NO_3)_2$-H_2O system [7]. However, because calcium nitrate can form hydrated crystalline phases in aqueous systems, the solid-phase quantities in the present study are expressed as $Ca(NO_3)_2$ equivalents on an anhydrous-salt basis rather than assigned to a specific calcium nitrate hydrate. Identifying the exact crystalline phase would require independent structural or thermal characterization.

From a technological and economic standpoint, cooling the solution below -15 °C was considered inefficient under the investigated conditions. Lowering the temperature from -15 to -20 °C increased calcium nitrate recovery by only 0.83 percentage points, equivalent to an additional 2.03 g of $Ca(NO_3)_2$ per 1000 g of the initial solution. This limited increase in product recovery does not justify the additional cooling requirement. Therefore, the temperature range from 0 to -15 °C is a rational operating range for calcium nitrate crystallization, providing a recovery of 59.74-94.21 %.

Evaporative concentration of calcium nitrate solutions

A second processing route involved concentrating the calcium nitrate solution by evaporation at atmospheric pressure. The initial solution mass was 1000.0 g and contained 24.40 wt.% $Ca(NO_3)_2$, corresponding to 244.0 g of calcium nitrate and 756.0 g of water. During evaporation, water was progressively removed while calcium nitrate remained in the liquid phase, causing a continuous decrease in total solution mass and a corresponding increase in $Ca(NO_3)_2$ concentration.

Table 3.

Material balance of evaporative concentration of the calcium nitrate solution

Stage	Solution mass, g	Cumulative evaporated water, g	$Ca(NO_3)_2$ concentration, wt.%	Water content, wt.%	Water remaining, g
Initial	1000.0	-	24.40	75.60	756.0
1	781.8	218.2	31.21	68.79	537.8
2	712.0	288.0	34.27	65.73	468.0
3	659.1	340.9	37.02	62.98	415.1
4	572.23	427.77	42.64	57.36	328.23
5	551.29	448.71	44.26	55.74	307.29
6	528.14	471.86	46.20	53.80	284.14
7	506.33	493.67	48.19	51.81	262.33
8	487.90	512.10	50.01	49.99	243.90

Stage	Solution mass, g	Cumulative evaporated water, g	$Ca(NO_3)_2$ concentration, wt.%	Water content, wt.%	Water remaining, g
9	468.42	531.58	52.09	47.91	224.42

As shown in Table 3, progressive evaporation systematically increased the calcium nitrate concentration. After removing 218. 20 g of water, the solution mass decreased to 781.80 80 g, and the $Ca(NO_3)_2$ concentration rose from 24. 24.40 to 31. 21 wt. %. Further evaporation reduced the solution mass to 659.10 10 g, at which point the calcium nitrate concentration reached 37. 02 wt. %.

Continued water removal produced progressively more concentrated calcium nitrate solutions. At a residual solution mass of 506.33 g, 33 g, the $Ca(NO_3)_2$ concentration reached 48. 19 wt. %, whereas at 487.90 90 g, it reached approximately 50 wt. %. At the final investigated stage, the solution mass decreased to 468.42 42 g, and the calcium nitrate concentration increased to 52. 09 wt. %.

Overall, 531.58 g 58 g of water was removed from the initial 756.00 g, corresponding to approximately 70.33% of the initial water content. The remaining concentrated solution contained 224.42 g of water together with the initial 244.00 g of calcium nitrate. The complete mass balance was therefore preserved throughout the calculation.

The results show that evaporation directly increases calcium nitrate concentration without introducing additional chemical reagents. In contrast to cooling crystallization, which transfers calcium nitrate into a solid phase, evaporative concentration retains calcium nitrate in the liquid phase while reducing water content. Consequently, the two approaches serve different technological purposes: cooling crystallization may be used when recovery of a crystalline calcium nitrate-containing product is required, whereas evaporation may be preferred when a concentrated liquid calcium nitrate stream is desired for subsequent processing or direct fertilizer-related applications.

A particularly relevant concentration level was reached at approximately 37 wt.% $Ca(NO_3)_2$, corresponding to a residual solution mass of 659.10 g after removal of 340.90 g of water. Previous investigations of calcium nitrate solutions generated during nitric acid processing of Central Kyzylkum phosphorites identified a $Ca(NO_3)_2$ concentration of approximately 37 wt.% as a suitable condition for subsequent conversion with ammonium carbonate to ammonium nitrate and calcium carbonate [9-10]. Therefore, concentrating the initial 24.40 wt.% solution to at least approximately 37 wt.% may provide a technologically suitable intermediate stream for subsequent conversion. Further evaporation in the present study increased the $Ca(NO_3)_2$ concentration to 52.09 wt.%, indicating that the final concentration can be adjusted according to the intended downstream processing route.

Comparative technological assessment of cooling crystallization and evaporative concentration

The results for the two investigated processing routes indicate that cooling crystallization and evaporative concentration can serve different technological purposes in treating calcium nitrate-containing solutions generated during nitric acid processing of Central Kyzylkum phosphorites.

Cooling crystallization provides a direct route for transferring calcium nitrate from the liquid phase into a solid crystalline fraction. Within the investigated temperature range, the calculated recovery increased continuously as the temperature decreased. Cooling from 0 to -15 °C increased recovery from 59.74 to 94.21 %, while further cooling to -20 °C yielded only a small additional increase to 95.04 %. Thus, the interval of 0 to -15 °C can be regarded as the rational operating range for calcium nitrate recovery under the investigated conditions. The limited increase of only 0.83 percentage points between -15 and -20 °C indicates that deeper cooling provides little additional product recovery while requiring more intensive refrigeration.

Evaporative concentration operates on a different technological principle. Rather than transferring calcium nitrate to a solid phase, evaporation retains the dissolved salt in the liquid phase while progressively removing water. Starting from a solution containing 24.40 wt.% $Ca(NO_3)_2$, the concentration increased to 37.02 wt.% after removing 340.90 g of water and reached 52.09 wt.% at the final investigated stage. Overall, 531.58 g of water, corresponding to approximately 70.3 % of the initial water content, was removed from 1000 g of the starting solution.

A concentration of approximately 37 wt.% $Ca(NO_3)_2$ is technologically relevant because previous studies of calcium nitrate solutions derived from Central Kyzylkum phosphorites identified this range as suitable for subsequent conversion to ammonium nitrate and calcium carbonate [9]. Related work has also demonstrated the feasibility of converting calcium nitrate solutions obtained from Kyzylkum phosphate raw materials with ammonium carbonate [10]. Therefore, evaporative concentration to approximately 37 wt.% may be a practical intermediate step when subsequent chemical conversion is intended. From a process-selection perspective, cooling crystallization may be preferred when the objective is to recover calcium nitrate as a solid, whereas evaporative concentration is more suitable when a concentrated liquid stream is required for subsequent conversion or potential fertilizer application. The two routes should therefore not be regarded as mutually exclusive alternatives. Instead, they represent complementary processing options that can be selected based on the desired final product and downstream technological requirements. Overall, both approaches lead to more complete utilization of the calcium and nitrogen in the liquid phase formed during nitric acid processing of Central Kyzylkum phosphorites. Their application reduces the need to treat the calcium nitrate-containing filtrate as a residual process stream and supports more integrated utilization of the mineral raw material.

Conclusions

The study demonstrated the technological feasibility of processing calcium nitrate solutions generated during nitric acid treatment of Central Kyzylkum phosphorites by cooling crystallization and evaporative concentration.

Cooling of the initial 24.40 wt.% Ca(NO3)2 solution from +20 to -15 °C progressively increased the calculated calcium nitrate recovery into the crystalline phase from 25.27 to 94.21 %. Further cooling to -20 °C increased the recovery only slightly, to 95.04 %, corresponding to an additional 2.03 g of $Ca(NO_3)_2$ equivalent per 1000 g of the initial solution. Therefore, the temperature range from 0 to -15 °C was identified as the rational operating range for calcium nitrate crystallization, while further cooling to -20 °C was considered economically inefficient because of the limited additional recovery. Evaporative concentration provided an alternative route for retaining calcium nitrate in the liquid phase while removing water. Starting from 1000 g of solution containing 24.40 wt.% Ca(NO3)2, removal of 531.58 g of water increased the calcium nitrate concentration to 52.09 wt.%. This process removed approximately 70.3 % of the initial water content. A Ca(NO3)2 concentration of approximately 37 wt.% was reached after removal of 340.90 g of water and represents a technologically relevant intermediate level for subsequent conversion processes reported for calcium nitrate solutions derived from Central Kyzylkum phosphorites. Further evaporation to concentrations above 50 wt.% provides additional flexibility depending on the intended downstream use. Overall, cooling crystallization and evaporative concentration represent complementary processing routes. The first is suitable for recovering calcium nitrate as a crystalline product, whereas the second provides concentrated calcium nitrate solutions for subsequent chemical conversion or potential nitrogen–calcium fertilizer applications. The proposed approaches improve the utilization of calcium and nitrate nitrogen contained in the liquid phase and contribute to a more integrated processing scheme for Central Kyzylkum phosphate raw materials.

References

1. Ryszko U., Rusek P. Kołodyńska D. Quality of phosphate rocks from various deposits used in wet phosphoric acid and P-fertilizer production // Materials. – 2023. – Vol. 16, No. 2. – Art. 793. – DOI: 10.3390/ma16020793.
2. Giyasidinov A.L., Sultonov B.E., Dormeshkin O. Study on the composition of phosphorus fertilizers obtained on the basis of Kizilkum phosphorites and nitric acid // Journal of Chemical Technology and Metallurgy. – 2023. – Vol. 58, No. 2. – P. 347–352.
3. Kholmatov D.S., Rasulov A.A. Treatment of phosphate waste generated during thermal processing of phosphorites of the Central Kyzylkum // Obogashchenie Rud. – 2024. – No. 4. – P. 1–8. – DOI: 10.17580/or.2024.04.06.
4. Akfas F., Elghali A., Aboulaich A., Munoz M., Benzaazoua M., Bodinier J.L. Exploring the potential reuse of phosphogypsum: a waste or a resource? // Science of the Total Environment. – 2024. – Vol. 908. – Art. 168196. – DOI: 10.1016/j.scitotenv.2023.168196.
5. Wang B., Zhou Z., Xu D., Wu J., Yang X., Zhang Z., Yan Z. A new enrichment method of medium–low grade phosphate ore with high silicon content // Minerals Engineering. – 2022. – Vol. 181. – Art. 107548. – DOI: 10.1016/j.mineng.2022.107548.
6. Xu C., Zhang M., Su J., Peng X., Kong X., Luo Y., Sun G. Removal of calcium from phosphoric acid produced by the nitric acid process using solvent extraction with TCHDGA and stripping with water // Hydrometallurgy. – 2024. – Vol. 229. – Art. 106379. – DOI: 10.1016/j.hydromet. 2024.106379.
7. Novikov A.A., Dzuban A.V., Kovalenko N.A., Uspenskaya I.A. Thermodynamic properties and phase equilibria in the H_2O–$Ca(NO_3)_2$ system // Journal of Chemical & Engineering Data. – 2021. – Vol. 66, No. 5. – P. 1839–1855. – DOI: 10.1021/acs.jced.1c00102.
8. Nasiraee M., Mousavi S.M., Saljoughi E., Kiani S., Razmgar K. Production of calcium nitrate crystals via membrane distillation crystallization using polyvinylidene fluoride/sorbitan trioleate membranes // Advanced Powder Technology. – 2021. – Vol. 32, No. 5. – P. 1463–1471. – DOI: 10.1016/j.apt.2021.02.043.

9. Dekhkanov Z.K., Seytnazarov A.R., Namazov Sh.S., Beglov B.M. Conversion of calcium nitrate as a by-product of chemical beneficiation of Central Kyzylkum phosphorites // Khimicheskaya Promyshlennost. – 2013. – Vol. 90, No. 2. – P. 87–92.
10. Allamuratova A.J., Erkaev A.U., Reymov A.M. Conversion of calcium nitrate solution obtained from Kyzylkum phosphorite with ammonium carbonate // Chemical Science International Journal. – 2016. – Vol. 16, No. 4. – P. 1–6. – DOI: 10.9734/ACSJ/2016/28146.

Статья поступила в редакцию: 04.08.2026
Статья принята к публикации: 05.09.2026
Статья опубликована: 28.09.2026

УДК 622.7+553.6

DOI: 10.32743/UniTech.2026.150.9.23382

Research on the flotation enrichment process of “Changi” glauconite using alkylbenzenesulfonic acid

Tursunov Azamjon
Doctoral student of Andijan State Technical Institute
teacher, Fergana State Technical University
Republic of Uzbekistan, Fergana
E-mail: turcunov-farpi@mail.ru

Turdialiyev Umid
professor of Andijan State Technical Institute
Republic of Uzbekistan, Andijan

Исследование процесса флотационного обогащения глауконита «Чанги» с использованием алкилбензолсульфоновой кислоты

Турсунов Азамджон
аспирант
Андижанский государственный технический институт, преподаватель Ферганский государственный технический университет
Республика Узбекистан, г. Фергана

Турдиалиев Умид
проф.
Андижанский государственный технический институт
Республика Узбекистан, Андижан

Abstract

The technological parameters of the use of alkylbenzenesulfonic acid as a reagent in the enrichment of the “Changi” glauconite mineral using a flotation machine were studied in laboratory conditions. A glauconite sample passed through a 0.075 mm sieve was used for the study. Flotation experiments were conducted in a laboratory flotation machine in an environment with pH = 5 – 7. The pulp composition was prepared from 332 g of mineral and 2 l of water (solid: liquid = 1: 6). Flotation was carried out in three stages. According to the final results, the total mass of the

---

Библиографическое описание: Турсунов А., Турдиалиев У. Исследование процесса флотационного обогащения глауконита «Чанги» с использованием алкилбензолсульфоновой кислоты // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23382

For citation: Tursunov A., Turdialiyev U. Research on the flotation enrichment process of “Changi” glauconite using alkylbenzenesulfonic acid // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23382

concentrate was 145 g, the sediment was 184 g, and the loss was 3 g. The yield of the concentrate in terms of percentage concentration was 43.67 %, the sediment yield was 55.42 %, and the loss was 0.90 %. The results obtained indicate that alkylbenzenesulfonic acid has the potential to increase flotation efficiency by selectively hydrophobizing glauconite particles.

Аннотация

Изучены технологические параметры использования алкилбензолсульфоновой кислоты в качестве реагента при обогащении минерала глауконита «Чанги» с помощью флотационной машины в лабораторных условиях. Флотация — один из наиболее эффективных физико-химических методов разделения при обогащении полезных ископаемых. Этот метод основан на изменении гидрофобных или гидрофильных свойств поверхности минеральных частиц. Эффективность реагента зависит от крупности частиц, плотности пульпы, расхода реагента, продолжительности флотации и значения pH среды. Для исследования использовался образец глауконита, прошедший через сито размером 0,075 мм. Флотационные эксперименты проводились в лабораторной флотационной машине в среде с pH = 5–7. Состав пульпы готовился из 332 г минерала и 2 л воды (твердое: жидкое = 1: 6). Флотация проводилась в три стадии. По окончательным результатам общая масса концентрата составила 145 г, осадка — 184 г, потерь — 3 г. Выход концентрата в процентном отношении концентрации составил 43,67 %, выход осадка — 55,42 %, потери — 0,90 %. Полученные результаты показывают, что алкилбензолсульфоновая кислота имеет потенциал повышения эффективности флотации путем селективной гидрофобизации частиц глауконита.

Keywords: Glauconite, flotation, alkylbenzenesulfonic acid, collector, enrichment, concentrate, mineral raw materials, proportions, crushing.

Ключевые слова: Глауконит, флотация, алкилбензолсульфоновая кислота, собиратель, обогащение, концентрат, минеральное сырье, пропорции, дробление.

Introduction

Glauconite is an iron-potassium hydrosilicate with a layered structure, which, due to its high content of K_2O (up to 6–8 %), Fe_2O_3 (up to 20–25 %), as well as aluminium and silicon, is a promising multifunctional raw material for agriculture, the chemical industry, and the production of sorbents and mineral pigments. Glauconite-bearing rocks are widespread in the region, and among them, the “Changi” deposit is of particular interest. It was previously established that glauconite sandstones from eastern Uzbekistan are a promising raw material for the production of mineral pigments [1]. The general characteristics of glauconite-containing sands in the republic and the directions for their practical use are substantiated in work [2].

The issues of beneficiation of glauconite-containing rocks from the “Changi” deposit have been considered in a number of publications. Thus, D.K. Adylov and his co-authors investigated the possibility of beneficiation of glauconite-containing sandstones from this deposit [3]. Makhsudova Z.I. and her co-authors studied the enrichment of pigment-containing sandstones from the same deposit [4]. Bauatdinov S.N. and Aimbetov I.K. developed recommendations for the use of enriched glauconite sand as a fertilizer in agriculture [5].

Flotation is one of the most effective physicochemical separation methods used in the enrichment of minerals. It is based on changing the hydrophobic or hydrophilic properties of the surface of mineral particles due to the adsorption of collectors. Among the anion-active reagents, alkylbenzenesulfonic acid (ABSC) stands out, which is characterized by its ability to adsorb on the surface of silicate minerals and stabilize the foam layer [9]. Previously, the effectiveness of ABSC application was shown in the flotation enrichment of graphite ores from the Taskazgan deposit [6, 7, 8].

At the same time, there are no systematic studies of flotation enrichment of glauconite from the Changi deposit using alkylbenzenesulfonic acid as a collector in the available literature. The influence of the reagent consumption, the duration, and the number of stages in the process on the concentrate yield remains unstudied. Chemical and mineralogical studies have confirmed the suitability of "Changi" glauconites for the production of ceramic pigments [10], which further substantiates the relevance of developing an effective technology for their enrichment.

The aim of this work is to study the technological parameters of flotation enrichment of glauconite from the "Changi" deposit using alkylbenzene sulfonic acid under laboratory conditions and to determine the optimal process conditions.

Materials and methods

The enrichment of "Changi" glauconite with alkylbenzenesulfonic acid in a flotation machine under laboratory conditions, the effect of reagent consumption and flotation duration on concentrate yield was determined.

Object of the study. The object was a sample of glauconite ore from the Changi deposit, collected using the standard method for sampling solid mineral deposits.

Material preparation. The initial sample weighing 322 g was dried at 105 °C until it reached a constant mass, then crushed and ground, after which it was sifted through a laboratory sieve with a mesh size of 0.075 mm. The choice of this particle size class is due to the need to increase the specific surface area of the particles to intensify the adsorption of the collector and increase the likelihood of adhesion to air bubbles.

Equipment. Flotation experiments were conducted in a laboratory flotation machine with mechanical mixing and air supply. The pulp pH was monitored using a pH meter and maintained within the range of 5–7 using buffer solutions.

Pulp preparation. The pulp was prepared by mixing 332 g of the prepared material with 2 L of tap water, with a solid-to-liquid ratio of 1:1.6. Alkylbenzene sulfonic acid (ABS) was used as a collector – an anion-active reagent that provides hydrophobization of the surface of silicate minerals.

Flotation scheme. The flotation was carried out in three stages. At each stage, the reagent consumption (in drops, converted to grams), the duration of the process, and the mass of the obtained concentrate were recorded. Stage 1: 28 min at a consumption of 0.390 g (15 drops); Stage 2: 20 min at a consumption of 0.208 g (8 drops); Stage 3: 18 min at a consumption of 0.130 g (5 drops). After each stage, the foam product (concentrate) was separated, dried, and weighed. The chamber product (sediment) was sent to the next stage.

Assessment of results. The enrichment efficiency was assessed based on the concentrate yield (the ratio of the concentrate mass to the mass of the original sample, %). Losses were determined as the difference between the mass of the original sample and the sum of the masses of the concentrate and the sediment.

Research results

During the three-stage flotation of glauconite ore from the Changi deposit using alkylbenzene sulfonic acid, the following results were obtained.

At the first stage (reagent consumption: 0.390 g, 28 min), 69 g of concentrate was separated, which amounted to 42.76 % of the total yield. The froth layer was thick and stable, which indicates active release of hydrophobic particles.

At the second stage (0.208 g, 20 min), 39 g of concentrate (26.90 %) was obtained. The froth layer became less intense, but flotation continued steadily.

At the third stage (0.130 g, 18 min), an additional 44 g of concentrate (30.34 %) was isolated. A fine-bubbled froth layer was observed, which ensured the additional extraction of the remaining particles.

The total mass of the concentrate after three stages was 152 g (69 + 39 + 44), the mass of the sludge was 184 g, and the losses were 3 g. The concentrate yield was 43.67 %, the sludge yield was 55.42 %, and the losses were 0.90 % (Tables 1, 2).

The highest concentrate yield was achieved at the first stage, which is explained by the maximum consumption of the collector and the high content of hydrophobic particles in the initial pulp. The decrease in yield at subsequent stages is naturally linked to a reduction in reagent consumption and the separation of the bulk of the flotable particles at previous stages. The three-stage scheme ensures a higher degree of recovery compared to the single-stage one. Losses of less than 1 % confirm the correctness of the methodology and the maintenance of the material balance.

Table 1.

Results of triple flotation enrichment of glauconite ore

Name	Water quantity (l)	Glauconite ore (gr)	Amount of ABSK flotation agent, (drops, gr)	Flotation time (min)	Finished concentrate (gr), (%)	Residue after flotation (gr), (%)
First enrichment						184 (55,42%)
Sample 1	2,0+0,9=2,9	332.0	15 / 0,390	28	69 (42,76%)	
Second enrichment						
Sample 2	1,0	-	8 / 0,208	20	39 (26,90%)	
Third enrichment						
Sample 3	1,0	-	5 / 0,130	18	44 (30,34%)	

Table 2.

Concentrate formed during flotation enrichment

Name	Weight (g)	Mass fraction (%)
Concentrate	145	43,67

Name	Weight (g)	Mass fraction (%)
Sediment	184	55,42
Loss	3	0,90
Total	**332**	**100,00**

In the second stage, the amount of reagent was reduced from 0.390 g to 0.208 g. Due to the decrease in reagent concentration and the separation of most of the flotation-prone particles in the first stage, the concentrate yield was 39 g. This represents the natural kinetic law observed in multi-stage flotation.

Although the reagent consumption in the third stage was reduced to 0,130 g, 44 g of concentrate was separated. This result is explained by the additional flotation of small and relatively weak hydrophobic particles remaining in the pulp after the second stage. This indicates that the three-stage flotation scheme provides a higher overall separation efficiency compared to single-stage flotation.

In total, 145 g of concentrate was recovered, representing a concentrate yield of 43.67 %. The tailing (sediment) fraction was 55.42 %, while the technical loss was only 0.90 %. A loss of less than 1 % indicates that the experimental methodology was correctly selected and that the mass balance was almost fully preserved under laboratory conditions.

The following photographs show the gradual formation of flotation foam. In the first stage, the foam layer was thick and stable, indicating active separation of hydrophobic particles. In the second stage, the foam layer became relatively thin, but flotation continued. In the third stage, a fine-bubble foam was observed, which provided additional separation of the remaining particles [7, 8, 9, 10, 11].

Figure 1. General view of the foam layer formed in the first flo tation stage (0.390 g reagent, 28 min)

Figure 2. Foam stabilization in the seco nd flotation stage (0.208 g of reagent, 20 min)

Figure 3. Structure of the fine-bub ble foam formed in the third flota-tion stage (0.130 g reagent, 18 min)

The flotation process is a complex process that depends on the physicochemical properties of mineral particles, reagent type, pulp composition, flotation duration, and particle size. In this study, alkylbenzenesulfonic acid was used as a collector for “Changi” glauconite, and a three-stage flotation scheme was tested.

Conclusion

The results obtained showed that the “Changi” glauconite of alkylbenzenesulfonic acid can be used as an effective collector in flotation enrichment. The three-stage flotation scheme allows obtaining a higher concentrate than the single-stage process. These results can serve as initial laboratory data for the development of industrial-scale glauconite enrichment technology.

The flotation time in the first stage was 28 minutes, during which time sufficient conditions were created for adsorption of the collector to the mineral surface and the formation of stable aggregates with air bubbles. Although the flotation time was reduced to 20 and 18 minutes in the subsequent stages, the additional concentrate obtained demonstrates the importance of choosing a balance between the process duration and the consumption of reagents.

References

1. Makhsudova Z.I., Mirzaev A.Zh., Kadyrova Z.R., Chernichenko N.I. Chemical-mineralogical investigation of glauconites of the «Changi» deposit for ceramic pigments // Universum: технические науки. – 2022. – No. 5 (98) – URL: https://7universum.com/ru/tech/archive/item/13794 (дата обращения: 21.09.2026).
2. Yakubov S.I., Mirzaev A.Z., Yunusova F.R., Beknazarova Z. Granulometric and microscopic analysis of glauconite-containing sandstones from the «Changi» deposit, Parkent district // AIP Conference Proceedings. – 2025. – Vol. 3286, No. 1. – DOI: 10.1063/5.0280703.
3. Makhsudova Z.I., Kadyrova Z.R., Mirzaev A.Zh. Glauconite sandstones of the Changi deposit – promising raw material for obtaining green and brown pigments // Glass and Ceramics. – 2023. – Vol. 80, No. 1. – DOI: 10.1007/s10717-023-00558-w.
4. Cordeiro L.B.C., Oliveira M.S. de Oliveira K.D., Barrozo M.A.S., de Ávila Neto C.N. Feasibility of concentrating potassium minerals from Verdete ore by cell flotation // The Journal of Engineering and Exact Sciences. – 2021. – Vol. 7, No. 2. – DOI: 10.18540/jcecvl7iss2pp12338-01-12e.
5. Farahat M.M., El-Midany A.A. Removal of major phosphate impurities by flotation using DTAB collector // Tenside Surfactants Detergents. – 2021. – Vol. 58. – DOI: 10.1515/tsd-2020-2319.
6. Encyclopedia of Separation Science / Ed. M. Cooke. – Academic Press, 2000. – Vol. 4. – P. 1863–1872.
7. Kenzhaliyev B.K., Koizhanova A., Fischer D., Magomedov D., Yerdenova M., Smailov K., Abdyldayev N. A Study on the efficiency of organic activator application to process refractory hard-to-beneficiate raw materials // Jurnal Pendidikan Teknologi Kejuruan. – 2024. – Vol. 7, No. 3. – DOI: 10.24036/jptk.v7i3.38323.
8. Tursunov A.S. et al. Studies of the process of flotation enrichment of graphite ore from the Taskazgan deposit of the Republic of Uzbekistan // Universum: технические науки. – 2019. – No. 10.
9. Tursunov A., Turdialiyev U. The Use of Graphite ore Concentrate from the Taskazgan Deposit as a Lubricant // E3S Web of Conferences. – 2023. – Vol. 449. – DOI: 10.1051/e3sconf/202344906012.

10. Tursunov A., Turdialiyev U. Study of the flotation enrichment process of graphite ore from the Taskazgan deposit // Interpretation and Researches. – 2023. – Vol. 1, No. 32.

METALLURGY, MATERIALS SCIENCE AND MINING

Статья поступила в редакцию: 11.09.2026
Статья принята к публикации: 20.09.2026
Статья опубликована: 28.09.2026

УДК 622.271

DOI: 10.32743/UniTech.2026.150.9.23432

Improvement of drilling and blasting parameters in open-pit mining of the Karakhtai limestone deposit

Kushnabaev Mirziyod Mirzohid ugli
Mining Engineer, Independent Researcher
graduate of the Branch of NUST MISIS in Almalyk
Republic of Uzbekistan, Almalyk
E-mail: qoshnaboyev@gmail.com

Совершенствование параметров буровзрывных работ при открытой разработке Карахтайского месторождения известняков

Кушнабаев Мирзиёд Мирзохид угли
горный инженер
независимый исследователь, выпускник Филиала Национального исследовательского технологического университета МИСИС в г. Алмалык
Республика Узбекистан, Алмалык

Abstract

The study addresses the improvement of drilling and blasting operations in the open-pit mining of the Karakhtai limestone deposit in the Akhangaran district of the Tashkent Region, Republic of Uzbekistan. The objective is to assess the existing blast-preparation scheme and substantiate a project-based direction for reducing drilling-and-blasting costs while preserving acceptable fragmentation and operational safety. The analysis uses geological and hydrogeological characteristics of the deposit, physical and mechanical properties of limestone, parameters of the mining system, drilling equipment, blasthole layout, initiation systems, and comparative technical and economic calculations contained in the graduation thesis. The studied limestone has a density of about 2.56 t/m³ and a uniaxial compressive strength of 41.3–44.8 MPa in the natural state and 27.4–30.2 MPa when water-saturated. The compared project variant increases blasthole diameter from

---

Библиографическое описание: Кушнабаев М.М.У. Совершенствование параметров буровзрывных работ при открытой разработке Карахтайского месторождения известняков // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23432

For citation: Kushnabaev M.M.U. Improvement of drilling and blasting parameters in open-pit mining of the Karakhtai limestone deposit // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23432

130 to 152 mm and changes the drilling pattern from 4 × 4 m to 4 × 6 m with a staggered arrangement. The calculated annual number of blastholes decreases from 9,800 to 6,402, while the considered drilling-and-blasting costs decrease from 851,019.75 to 698,746.29 conventional units. The results are interpreted as a project hypothesis requiring controlled field verification before industrial implementation.

Аннотация

В статье рассматривается совершенствование буровзрывных работ при открытой разработке Карахтайского месторождения известняков в Ахангаранском районе Ташкентской области Республики Узбекистан. Цель исследования заключается в анализе существующей схемы подготовки массива и обосновании проектного направления снижения затрат на буровзрывные работы при сохранении приемлемого качества дробления и требований безопасности. Исходными материалами являются горно-геологические и гидрогеологические характеристики месторождения, физико-механические свойства известняка, параметры системы разработки, применяемое буровое оборудование, параметры сетки скважин, системы инициирования и сравнительные технико-экономические расчёты выпускной квалификационной работы. Плотность известняка составляет около 2,56 т/м³, предел прочности при одноосном сжатии — 41,3–44,8 МПа в естественном состоянии и 27,4–30,2 МПа в водонасыщенном. В рассматриваемом проектном варианте диаметр скважин увеличивается с 130 до 152 мм, а сетка бурения изменяется с 4 × 4 м до 4 × 6 м при шахматном расположении. Расчётное годовое количество скважин уменьшается с 9800 до 6402, а затраты по рассматриваемым статьям БВР — с 851019,75 до 698746,29 условных единиц. Полученные результаты следует рассматривать как проектную гипотезу, подлежащую опытно-промышленной проверке.

Keywords: Karakhtai deposit, limestone, open-pit mining, drilling and blasting, rock fragmentation, water-bearing horizons, blast optimization.

Ключевые слова: Карахтайское месторождение, известняк, открытая разработка, буровзрывные работы, дробление горных пород, водоносные горизонты, оптимизация БВР.

Introduction

The efficiency of open-pit mining of cement and construction raw materials is determined not only by the quantity of extracted rock but also by the quality of its preparation for excavation, transportation, and subsequent crushing. Recent international studies increasingly treat blasting as one component of an integrated size-reduction chain. Fragmentation affects downstream energy demand, productivity, operating costs, recovery, and environmental performance, which means that blast design should be evaluated against the whole production system rather than against drilling or blasting costs alone [1, 2].

Another important research direction is the use of drilling-monitoring information to characterize rock-mass variability and adapt blast design to actual geological conditions. Navarro et al. demonstrated that measurement-while-drilling data can be used to classify structural conditions and support blastability assessment in open-pit operations [1]. This approach is relevant to heterogeneous deposits where rock strength and fracturing vary spatially.

Full-scale research also confirms that the relationship between explosive energy, powder factor, and initiation delay is nonlinear and must be considered as a combined design problem [3]. Sensitivity studies show that several blast-design parameters jointly affect fragmentation outcomes, while digital monitoring and machine-vision approaches increasingly support quantitative assessment of the post-blast muckpile [4–6]. The influence of initiation conditions has also been demonstrated in full-scale field tests at the Kevitsa open pit mine [8, 9].

The Karakhtai limestone deposit is located in the Akhangaran district of the Tashkent Region of the Republic of Uzbekistan, approximately 75 km southeast of Tashkent and 10 km northeast of Akhangaran. The deposit serves as a raw-material base for cement production. Mined limestone is transported by truck to a crushing complex and then delivered by a conveyor approximately 7 km long to the cement plant raw-material stockpile. The deposit is characterized by heterogeneous geological conditions, variable bench heights, overburden thicknesses, and water-bearing horizons, which complicate the selection of drilling and blasting parameters.

The aim of this study is to analyze the existing drilling and blasting scheme at the Karakhtai deposit and to substantiate, on the basis of the graduation thesis calculations, a direction for improving blasthole parameters and layout in order to reduce the number of drilling operations and the associated costs while maintaining acceptable fragmentation, slope stability, and safety constraints.

Materials and Methods

The study is based on the graduation thesis ‘Preparation of an Open-Pit Mining Project for the Karakhtai Deposit’, including its special section on drilling and blasting in water-bearing horizons. The source materials comprise the geological and hydrogeological description of the deposit, physical and mechanical characteristics of the rocks, parameters of the mining system, drilling and excavation equipment, existing blast-design parameters, safety requirements, and comparative technical and economic calculations [10].

The Karakhtai limestone belongs to the Paleogene carbonate sequence. The limestone density adopted in the project is approximately 2.56 t/m^3, and the loosening coefficient is 1.35. The uniaxial compressive strength is 41.3–44.8 MPa in the natural state and 27.4–30.2 MPa in the water-saturated state. The softening coefficient is 0.61–0.69. These values indicate that the rock mass requires controlled preparation before excavation and that water saturation may materially influence blasting conditions [10].

The deposit is developed by an open-pit transport system with preliminary loosening of the part of the rock mass that cannot be excavated mechanically. Working bench heights vary by section and are generally 10–15 m. The thesis reports that part of the overburden can be excavated mechanically, whereas the remaining portion requires drilling and blasting. The mining equipment includes hydraulic excavators of the Komatsu PC500LC and Hitachi ZX-450-3 class and equivalent equipment [10].

Rotary-percussive drilling rigs of the KAYSHAN-125D type are considered in the project. The initial design uses vertical blastholes with a diameter of 130 mm and a multi-row arrangement. Short-delay initiation is used to control the sequence of rock breakage. For water-bearing horizons, the thesis considers water-resistant emulsion explosive systems, while non-electric initiation systems are used to provide the required sequence of initiation [10].

The comparative analysis focuses on the parameters that directly affect the number of drilling operations and the calculated cost of the blast-preparation stage: blasthole diameter, drilling pattern, arrangement of holes, annual number of holes, quantity of initiation elements, and the resulting cost of the considered drilling-and-blasting items. The proposed project variant increases the hole diameter from 130 to 152 mm and changes the pattern from 4 × 4 m to 4 × 6 m with staggered placement. The thesis also considers adjustment of the calculated specific explosive consumption. These values are reported here as source-derived project parameters and are not presented as a universal operational blasting prescription.

Results and Discussion

The existing project variant uses a 4 × 4 m drilling pattern with 130 mm blastholes. Such a scheme provides a relatively dense distribution of energy in the rock mass, but it requires a comparatively large number of holes. When the annual production requirement is considered, the drilling volume and associated consumption of initiation elements become significant cost components [10].

The proposed variant changes two interrelated geometric parameters: the blasthole diameter is increased to 152 mm, and the drilling pattern is changed to 4 × 6 m with staggered placement. The increase in spacing means that each blasthole is assigned a larger volume of rock. Consequently, fewer holes are required for a given annual production volume. At the same time, the larger spacing introduces a potential risk of poorer fragmentation if the energy distribution is not adequately matched to the rock structure. International research emphasizes that fragmentation depends on the interaction of rock properties, blast geometry, energy distribution, and initiation conditions [2–4].

The calculated comparison from the graduation thesis gives a reduction in annual blastholes from 9,800 to 6,402, corresponding to 3,398 fewer holes or approximately 34.7 %. The quantity of electrical initiation elements decreases from 19,600 to 12,804, while the number of initiating elements reported in the project calculations decreases by the same proportion. The considered drilling-and-blasting costs decrease from 851,019.75 to 698,746.29 conventional units, giving a calculated difference of 152,273.46 conventional units [10].

These results indicate a substantial calculated economic effect at the drilling-and-blasting stage. However, the reduction in direct costs cannot be treated as the final economic outcome. A wider drilling pattern may increase the proportion of oversize fragments, which can require secondary breakage or reduce excavator and crusher productivity. Reviews of the fragmentation chain emphasize that the optimum blast is the one that minimizes the total cost of size reduction from drilling and blasting through crushing and grinding, rather than the cost of blasting alone [2].

The results also support the importance of considering blastability and geological variability. Navarro et al. showed that drilling-monitoring data can be used to distinguish structural conditions of the rock mass and thereby improve the basis for blast design [1, 7]. For Karakhtai, this suggests that future field verification should not rely solely on average rock strength. The spatial variability of fracturing, water saturation, and bench geometry should be incorporated into the assessment of the proposed pattern.

The role of initiation conditions is equally important. Full-scale studies have shown that changes in initiation position and timing can influence fragmentation, while the interaction between powder factor and delay affects the resulting fragment-size distribution [3, 8, 9]. Therefore, the proposed geometric changes should be evaluated together with the existing initiation system rather than as an isolated drilling modification. This is especially important for water-bearing horizons, where the stability of the explosive system and the reliability of initiation are additional constraints [10].

Digital measurement can strengthen this verification stage. Machine-vision and photogrammetric approaches make it possible to estimate fragment-size distributions and compare actual results with the target fragmentation range [6, 9]. A practical monitoring scheme for Karakhtai could compare the proposed and existing patterns using consistent measurements of fragmentation, oversize, excavation productivity, crusher loading, and slope condition. Such a controlled comparison would allow the calculated economic benefit to be tested against the actual production chain.

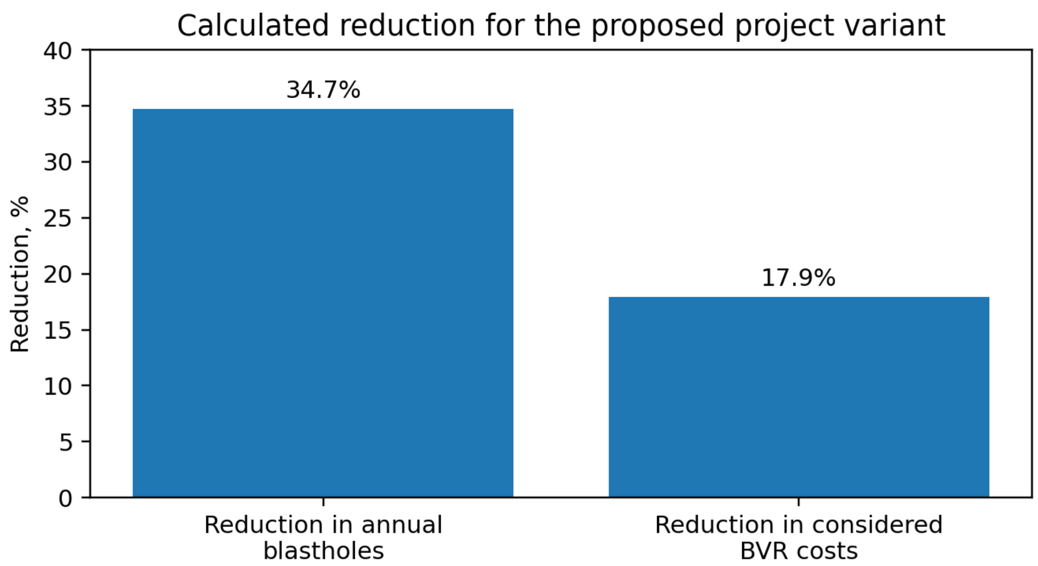

Figure 1. Calculated reduction in annual blastholes and in the considered drilling-and-blasting costs for the proposed project variant

Table 1.

Comparison of the main parameters of the existing and proposed drilling-and-blasting variants

Parameter	Existing variant	Proposed variant
Blasthole diameter, mm	130	152
Drilling pattern, m	4 × 4	4 × 6
Hole arrangement	Row	Staggered
Annual number of blastholes	9,800	6,402
Electrical initiation elements	19,600	12,804
Initiating elements	19,600	12,804

Parameter	Existing variant	Proposed variant
Considered BVR costs, conventional units	851,019.75	698,746.29
Calculated cost difference, conventional units	—	152,273.46

Note — the numerical indicators are taken from the graduation-thesis calculations and should be refined using current prices and controlled field-test results.

The calculated savings are therefore best interpreted as an initial engineering estimate. The proposed pattern should be accepted for industrial use only after controlled trials confirm that the fragmentation distribution, oversize proportion, excavator productivity, crusher performance, slope condition, and safety indicators remain within the project requirements. This limitation is consistent with the broader international trend toward data-supported blast optimization and continuous monitoring [1, 6].

Safety and environmental constraints remain integral to the assessment. The thesis specifies requirements for trained and certified personnel, hazardous-zone control, bench stability, equipment condition, communication, power supply, industrial sanitation, and emergency response [10]. Blast-induced vibration, dust and gas emissions, flyrock, and potential damage to the final pit walls must be considered together with the economic effect. Recent predictive studies likewise treat vibration control and blast-parameter optimization as interconnected tasks [5].

Conclusion

The analysis shows that the efficiency of drilling and blasting at the Karakhtai limestone deposit depends on the combined influence of rock properties, water-bearing conditions, blasthole geometry, drilling pattern, energy distribution, initiation conditions, and downstream excavation and crushing performance.

On the basis of the graduation-thesis calculations, the proposed project variant increases the blasthole diameter from 130 to 152 mm and changes the pattern from 4 × 4 m to 4 × 6 m with staggered placement. The calculated annual number of blastholes decreases from 9,800 to 6,402, while the considered drilling-and-blasting costs decrease from 851,019.75 to 698,746.29 conventional units.

The calculated reduction in drilling requirements and direct BVR costs demonstrates the potential economic value of the proposed approach. At the same time, the result cannot be considered sufficient evidence for immediate industrial implementation because a wider pattern can influence fragmentation quality, oversize generation, excavation productivity, crushing performance, and slope condition.

For water-bearing horizons, the selection of water-resistant explosive systems and reliable non-electric initiation methods is an essential project constraint. The final technical decision should therefore be based on controlled field verification with quantitative monitoring of fragmentation, oversize, equipment productivity, slope stability, environmental effects, and actual operating costs.

The main practical conclusion is that the Karakhtai project has a substantiated basis for further optimization of drilling and blasting parameters, but the proposed values should be treated as project-specific hypotheses requiring experimental confirmation rather than universal blasting recommendations.

References

1. Navarro J., Seidl T., Hartlieb P. et al. Blastability and Ore Grade Assessment from Drill Monitoring for Open Pit Applications // Rock Mechanics and Rock Engineering. – 2021. – Vol. 54. – DOI: 10.1007/s00603-020-02354-2.
2. Zhang Z.-X., Sanchidrián J.A., Ouchterlony F., Luukkanen S. Reduction of Fragment Size from Mining to Mineral Processing: A Review // Rock Mechanics and Rock Engineering. – 2023. – Vol. 56. – DOI: 10.1007/s00603-022-03068-3.
3. Sanchidrián J.A., Segarra P., Ouchterlony F. et al. The Influential Role of Powder Factor vs. Delay in Full-Scale Blasting: A Perspective Through the Fragment Size-Energy Fan // Rock Mechanics and Rock Engineering. – 2022. – Vol. 55. – DOI: 10.1007/s00603-022-02856-1.
4. Hosseini M., Khandelwal M., Lotfi R. et al. Sensitivity Analysis on Blast Design Parameters to Improve Bench Blasting Outcomes Using the Taguchi Method // Geomechanics and Geophysics for Geo-Energy and Geo-Resources. – 2023. – Vol. 9. – DOI: 10.1007/s40948-023-00540-4.
5. Bai R., Zhang P., Zhang Z. et al. Optimization of Blasting Parameters and Prediction of Vibration Effects in Open Pit Mines Based on Deep Neural Networks // Alexandria Engineering Journal. – 2023. – Vol. 70. – DOI: 10.1016/j.aej.2023.02.043.
6. Koteleva N., Khokhlov S., Frenkel I. Digitalization in Open-Pit Mining: A New Approach in Monitoring and Control of Rock Fragmentation // Applied Sciences. – 2021. – Vol. 11, No. 22. – Art. 10848. – DOI: 10.3390/app112210848.
7. Nikkhah A., Vakylabad A.B., Hassanzadeh A., Niedoba T., Surowiak A. An Evaluation on the Impact of Ore Fragmented by Blasting on Mining Performance // Minerals. – 2022. – Vol. 12, No. 2. – Art. 258. – DOI: 10.3390/min12020258.
8. Ylitalo R.M., Zhang Z.-X., Bergström P. Effect of Detonator Position on Rock Fragmentation: Full-Scale Field Tests at Kevitsa Open Pit Mine // International Journal of Rock Mechanics and Mining Sciences. – 2021. – Vol. 147. – Art. 104918. – DOI: 10.1016/j.ijrmms.2021.104918.
9. Yuab Chi L., Ylitalo R.M., Zhang Z.-X. et al. A Case Study of Fragmentation Improvement by Primer Placement in Kevitsa Mine // Mining, Metallurgy & Exploration. – 2024. – Vol. 41. – DOI: 10.1007/s42461-024-00998-0.
10. Kushnabaev M.M. [Preparation of an Open-Pit Mining Project for the Karakhtai Deposit: Graduation Qualification Work]. – Almalyk: Branch of NUST MISIS in Almalyk, 2026. – 108 P. (In Russ.)

MECHANICAL ENGINEERING AND ROBOTICS

Статья поступила в редакцию: 07.07.2026
Статья принята к публикации: 10.09.2026
Статья опубликована: 28.09.2026

УДК 631.3+531.3

DOI: 10.32743/UniTech.2026.150.9.23436

Dynamic problem of a wedge-shaped working tool of soil-tilling implement

Xudjayev Muxiddin Kushshayevich
doctor of Technical Sciences, Associate Professor of the Department of Theoretical Mechanics and Theory of Machines and Mechanisms, Tashkent State Technical University
Republic of Uzbekistan, Tashkent

Botir Khasanov Berdimurot ugli
Senior teacher
Almalyk State Technical Institute
Republic of Uzbekistan, Almalyk

Ibrohimov Shehroz Abdurahmonovich
Senior teacher
Almalyk State Technical Institute
Republic of Uzbekistan, Almalyk

Библиографическое описание: Худжаев М.К., Хасанов Б.Б.Ў., Шехроз И.А. Динамическая проблема клиновидного рабочего инструмента почвообрабатывающего орудия // Universum: технические науки : электрон. научн. журн. 2026. 9(150). URL: https://7universum.com/ru/tech/archive/item/23436

For citation: Xudjayev M.K., Botir K.B.U., Ibrohimov S.A. Dynamic problem of a wedge-shaped working tool of soil-tilling implement // Universum: Technical Sciences : electronic scientific journal. 2026. No. 9(150). Available at: https://7universum.com/ru/tech/archive/item/23436

Динамическая проблема клиновидного рабочего инструмента почвообрабатывающего орудия

Худжаев Мухиддин Кушшаевич
д-р техн. наук, доц.
кафедры теоретической механики и теории машин и механизмов Ташкентский государственный технический университет
Республика Узбекистан, г. Ташкент

Хасанов Ботир Бердимурот ўгли
ст. преп
Алмалыкский государственный технический институт
Республика Узбекистан, г. Алмалык

Шехроз Иброхимов Абдурахмонович
ст. преп
Алмалыкский государственный технический институт
Республика Узбекистан, г. Алмалык

Abstract

This study aims to determine the dynamic reaction forces exerted by the soil against a wedge-shaped soil-tilling implement. Specifically, it focuses on the reaction forces produced by a single-sided wedge. The problem is addressed using theoretical mechanics based on the D’Alembert principle approximation, leading to an expanded solution to existing analytical formulations. A scheme illustrating the active and passive forces acting on the wedge is created by utilizing the principle of release from constraints. Based on the fundamental law of dynamics, equations of motion for the wedge-shaped working element are derived in projections along Cartesian coordinate axes. Using these equations, the dynamic reaction forces for the wedge are determined. Analytical expressions for the reaction force in this dynamic problem are developed under the condition that the wedge is subjected to a constant force, with the wedge's motion following a periodic law. The analysis incorporates an open system, applying Coulomb's formula for dry friction alongside a new method that differs from d'Alembert's principle. Under d'Alembert's principle, the differential is replaced with the object's inertial forces, reducing the differential equation to the ordinary equilibrium equation of statics. In this paper, an alternative approach is used; instead of relying on a differential equation, an ordinary equation is derived by differentiating the given law governing the wedge's motion.

Аннотация

Целью данного исследования является определение динамических реакционных сил, воздействующих со стороны почвы на клинообразное почвообрабатывающее орудие. В частности, исследование сосредоточено на реакционных силах, создаваемых односторонним клином. Задача решена методами теоретической механики на основе принципа Д'Аламбера, что привело к развитию существующих аналитических формулировок. Схема действующих активных и пассивных сил на клин построена с использованием принципа освобождения от связей. На основе фундаментального закона динамики выведены уравнения движения клинообразного рабочего элемента в проекциях на оси декартовой системы координат. С помощью этих уравнений определены динамические реакционные силы клина. Развиты аналитические выражения для реакционной силы в данной динамической задаче при условии, что клин подвергается постоянной силе, а его движение следует периодическому закону. Анализ включает открытую систему с применением формулы Кулона для сухого трения наряду с новым методом, отличающимся от принципа Д'Аламбера. При применении принципа Д'Аламбера дифференциал заменяется силами инерции объекта, что сводит дифференциальное уравнение к обычному уравнению статического равновесия. В данной работе используется альтернативный подход: вместо опоры на дифференциальное уравнение получено обычное уравнение посредством дифференцирования заданного закона, управляющего движением клина.

Keywords: wedge, dynamic problem, equilibrium equation, reaction forces.

Ключевые слова: клин, динамическая задача, уравнение равновесия, реакционные силы.

1. Introduction

Determination of the reaction forces of constraints of composite wedge pairs is the foundation for the study of the dynamic processes of many mechanisms used in mechanical engineering. Agricultural machines are equipped with V-belts and wedge friction gears, wedge valves are used in oil products transportation lines, freight car bogies are equipped with wedge spring-friction sets. Wedge pairs are widely covered in the literature [1, 2].

Using the method of addition of forces between two infinitesimal sections of the rod, in accordance with the d'Alembert principle, a differential equation of longitudinal vibrations of a wedge pair was compiled [3]. The general and particular solutions of the differential equation of longitudinal oscillations of a wedge pair are presented and their propagation under the action of a constant force corresponding to the given initial and boundary conditions is studied.

A generalized model of a spring-friction set of a freight car bogie was created in [4]. Critically discussing the lack of methods of classical mechanics in multiple publications, according to the principle of the release of constraints, an equation for the equilibrium of friction wedges was compiled and an analytical expression for the reaction of a friction wedge was obtained in [5].

This article addresses the dynamic problem of a wedge. The objective of the study is to determine a general solution by applying d'Alembert's principle, as well as a particular solution for a wedge-shaped working element under a periodic law of motion using a new method.

2. Research methods

The research methods are based on the classical methods of theoretical mechanics for compiling the equations of dynamics according to the d'Alembert principle and the analytical method for determining the reaction force of constraints.

3. Material

Let force $\bar{P}$ make an effort to move a homogeneous wedge (Fig. 1). It is necessary to determine the dynamic reaction forces of constraints on the wedge.

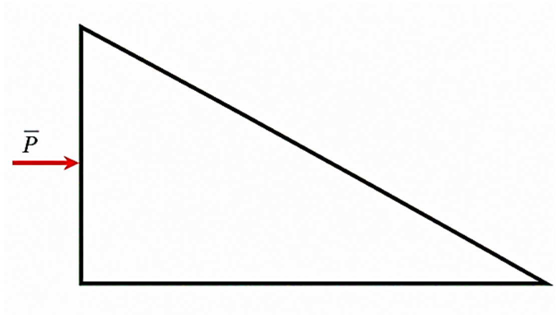

Figure 1. Wedge shape

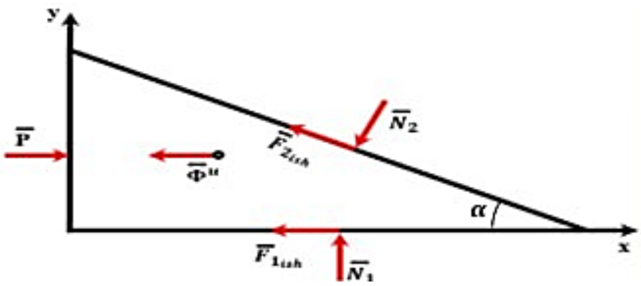

Figure 2. Acting forces

To do this, we use the d'Alembert principle. The equation of dynamics according to the d'Alembert principle has the following form [6, 7, 8]

$$\bar{\Phi}^{и}+\bar{F}^{a}+\bar{R}=0,$$

where $\bar{\Phi}^{и}$ is the force of inertia, expressed as $\bar{\Phi}^{и}=-m\ddot{x}$, $\bar{F}^{a}$ is the active force, expressed by force $\bar{P}$, $\bar{R}$ is the passive force, which consists of the forces of friction $\bar{F}_{ish}$ and the forces of normal reaction $\bar{N}$.

To determine the reaction force, we set a coordinate system and determine the projections of all forces along the coordinate axes of this coordinate system (Fig. 2). Normal reaction forces $\bar{N}$ of the working planes are directed perpendicular to these planes. Friction forces $\bar{F}_{ish}$ of these surfaces lie in the working planes. The force of inertia $\bar{\Phi}^{и}$ is directed in the opposite direction of force $\bar{P}$ direction.

Let us derive an equation for the projections of all wedge forces along the x -axis:

$$\Phi^{и}+P-fN_2\cos\alpha-N_2\sin\alpha-fN_1=0. \tag{1}$$

Let us derive an equation for the projections of all wedge forces along the y -axis:

$$N_1-N_2\cos\alpha+fN_2\sin\alpha=0. \tag{2}$$

Then we determine N_1:

$$N_1=N_2(\cos\alpha-f\sin\alpha). \tag{3}$$

Taking into account relation $F_{ish}=fN$ for the friction force (where f is the coefficient of friction), from the last equation we obtain:

$$\mathrm{P}-\mathrm{fN}_2\cos\alpha-\mathrm{N}_2\sin\alpha-\mathrm{fN}_2\cos\alpha+\mathrm{f}^2\mathrm{N}_2\sin\alpha-\mathrm{m}\ddot{x}=0,$$

Then we determine

$$m\ddot{x} - P = N_2(f^2 \sin\alpha - 2f\cos\alpha - \sin\alpha). \tag{4}$$

By integrating over time t in (4), we obtain an expression for the velocity:

$$\dot{x} = \frac{P}{m}t - \frac{N_2}{m}t(f^2 \sin\alpha - 2f\cos\alpha - \sin\alpha) + C_1. \tag{5}$$

Thus, the forces of normal reaction N_1 and N_2 of the working planes have the following form:

$$N_2 = \frac{m\ddot{x} - P}{f^2 \sin\alpha - 2f\cos\alpha - \sin\alpha}, \tag{6}$$

$$N_1 = \frac{(m\ddot{x} - P)(\cos\alpha - f\sin\alpha)}{f^2 \sin\alpha - 2f\cos\alpha - \sin\alpha}, \tag{7}$$

By substituting the inertial forces of the wedge from equations (6) and (7) into equation (6), we can calculate the dynamic reaction forces of a single-sided wedge, generated on the two working surfaces of the wedge.

4. Results

Thus, the analytical expressions for the dynamic reaction forces of a single-sided wedge are determined. Substituting the inertial forces of the wedge from (6) and (7) into (6), we can determine.

In a particular case, if the law of motion of the wedge is given in the following form:

$$x = \sin kt, \tag{8}$$

then, differentiating this expression twice, we obtain:

$$\ddot{x} = -k^2 \sin kt. \tag{9}$$

Substituting (9) into (6) and from (6), (7) we obtain expressions for the dynamic normal reaction forces

$$N_1 = \frac{(P + mk^2 \sin kt)(\cos\alpha - f\sin\alpha)}{\sin\alpha + 2f\cos\alpha - f^2 \sin\alpha}, \tag{10}$$

$$N_2 = \frac{P + mk^2 \sin kt}{\sin\alpha + 2f\cos\alpha - f^2 \sin\alpha}. \tag{11}$$

5. Conclusion

By applying d'Alembert's principle, an analytical solution to the dynamic problem of a one-sided wedge is obtained.

For a periodic law of wedge motion, the problem is solved using a new method that differs from d'Alembert's principle. According to d'Alembert's principle, the differential is replaced by the object’s inertial forces, which allows us to reduce the differential equations to the ordinary equilibrium equations of statics. In this paper, by differentiating the law of motion of the wedge, we obtain an ordinary equation from which we can derive expressions for the dynamic reaction forces of a single-sided wedge acting on its two surfaces.

References

1. Frolov K.V. Theory of mechanisms and machines / K.V. Frolov, S.A. Popov. – Moscow: Vysshaya shkola, 2005. – 496 P.
2. Artobolevsky, I. Theory of mechanisms and machines. – Moscow: Nauka, 1988. – 640 P.
3. Khudjaev, M. Modeling the dynamics of a wedge pair under the action of a constant force / M. Khudjaev, A. Rizaev, G. Pirnazarov, Sh. Khojikulov // Transportation Research Procedia. – 2022. – Vol. 63. – P. 449–457.
4. Turanov Kh.T. Generalized model of a spring-friction set of a freight car bogie // Transport: Science, Technology and Management. – 2009. – No. 12. – P. 32–36.
5. Turanov Kh.T. My scientific laboratory in mechanics. – Yekaterinburg. – 2013. – 240 P.
6. Szolga V.I. Theoretical mechanics. Part 2. – Berlin. – 2013. – 204 P.
7. Kleppner, D. An introduction to mechanics / D. Kleppner, R. Kolenkow. – Cambridge: Cambridge University Press, 2013. – 542 P.
8. Yablonsky A.A. Course of theoretical mechanics / A.A. Yablonsky, V.M. Nikiforova. – St. Petersburg: Lan, 2002. – 764 P.

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК

ДЛЯ ЗАМЕТОК

UNIVERSUM: ТЕХНИЧЕСКИЕ НАУКИ

Электронный научный журнал

№ 9(150)
Сентябрь 2026 г.
Часть 7

Свидетельство о регистрации СМИ: ЭЛ № ФС 77 – 54434 от 17.06.2013

Мнение авторов может не совпадать с позицией редакции.

Издание содержит научную информацию и в соответствии с ч. 4 ст. 1 Федерального закона от 29.12.2010 № 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию» не подлежит классификации (маркировке) по возрастным категориям

Издательство ООО «Юниверсум»

Объём издания: 96 с.

Минимальные системные требования: процессор с частотой от 1,3 ГГц; ОЗУ от 512 МБ; программа для чтения PDF-файлов; доступ к сети Интернет

Дата размещения в сети Интернет: 28.09.2026

Журнал размещается в:

CYBERLENINKA

Google scholar

BASE

OpenAIRE

НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА ELIBRARY.RU

ULRICHSWEB GLOBAL SERIALS DIRECTORY

СОЦИОНЕТ

EBSCO HOST

Registry of Open Access Repositories (ROAR)