

Novelty Assessment Report

Paper: Synchronizing Probabilities in Model-Driven Lossless Compression

PDF URL: <https://openreview.net/pdf?id=nGzYkW4FSB>

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

It is well-known in the field of lossless data compression that probabilistic next-symbol prediction can be used to compress sequences of symbols. Deep neural networks are able to capture rich dependencies in data, offering a powerful means of estimating these probabilities and hence an avenue towards more effective compression algorithms. However, both compressor and decompressor must have exactly matching predictions; even small non-deterministic differences (which often happen with learned models due to hardware, software, or computation order) can lead to cascading decoding failures. In this paper, we formalize the problem of prediction mismatch in model-driven compression, and introduce Probability Matching Interval Coding (PMATIC), a model-agnostic algorithm that tolerates bounded prediction mismatch with low overhead. PMATIC works with the predicted probabilities, making it compatible as a drop-in replacement for the arithmetic encoder in model-driven compression tools. We show theoretical correctness and performance bounds for PMATIC, and validate these results on text data. These results confirm that, when paired an advanced prediction model, PMATIC is robust to prediction mismatch while achieving compression rates that out-perform standard modern compression tools.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **Lossless Compression with Probabilistic Next-Symbol Prediction under Model Mismatch**

A total of **1 papers** were analyzed and organized into a taxonomy with **2 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Prediction Mismatch Tolerance Mechanisms**
- **Domain-Specific Prediction-Based Compression**

Complete Taxonomy Tree

- Lossless Compression with Probabilistic Next-Symbol Prediction under Model Mismatch Survey Taxonomy
- Prediction Mismatch Tolerance Mechanisms
 - Interval Coding with Probability Matching ★ (1 papers)
 - [0] Synchronizing Probabilities in Model-Driven Lossless Compression (Anon et al., 2026) [View paper](#)
- Domain-Specific Prediction-Based Compression
 - Large Language Model Output Compression (1 papers)
 - [1] Lossless Compression of Large Language Model-Generated Text via Next-Token Prediction (Mao Yu, 2025) [View paper](#)

Narrative

Core task: lossless compression with probabilistic next-symbol prediction under model mismatch. This field addresses the challenge of achieving efficient compression when the probability model used by the encoder and decoder may not perfectly match the true data distribution or each other. The taxonomy organizes work into two main branches. The first, Prediction Mismatch Tolerance Mechanisms, focuses on algorithmic strategies that maintain correctness and efficiency even when predictions are imperfect—such as interval coding schemes that adapt to probability mismatches or methods that synchronize encoder and decoder beliefs. The second branch, Domain-Specific Prediction-Based Compression, explores how prediction-based compression can be tailored to particular data types (text, images, structured data), often leveraging domain knowledge or specialized models to improve compression rates despite inherent model limitations.

Within Prediction Mismatch Tolerance Mechanisms, a particularly active line of work examines how to robustly encode symbols using interval or arithmetic coding when probability estimates are uncertain or evolving. Synchronizing Probabilities[0] sits squarely in this area, specifically within the Interval Coding with Probability Matching subfield, where it addresses how encoder and decoder can maintain consistent probability distributions to avoid decoding errors. Meanwhile, Domain-Specific approaches such as LLM Text Compression[1] illustrate how large language models can serve as powerful predictors for text data, though they must contend with the same mismatch issues when model outputs diverge from actual symbol frequencies. The central tension across these branches is balancing the expressiveness of rich predictive models against the need for reliable, synchronized compression—a trade-off that Synchronizing Probabilities[0] directly engages by proposing mechanisms to keep probabilistic beliefs aligned during the coding process.

Related Works in Same Category

No sibling papers and no sibling subtopics were found under the same parent taxonomy node; the paper appears structurally isolated in the taxonomy.

Contributions Analysis

Overall novelty summary. The paper introduces PMATIC, an algorithm designed to tolerate bounded prediction mismatch in model-driven lossless compression, alongside a formalization of the mismatch problem itself. Within the taxonomy, it occupies the 'Interval Coding with Probability Matching' leaf under 'Prediction Mismatch Tolerance Mechanisms'. Notably, this leaf contains only the original paper—no sibling papers were identified in the taxonomy. This positioning suggests the work addresses a relatively sparse research

direction: while prediction-based compression is well-established, explicit mechanisms for handling encoder-decoder probability discrepancies appear underexplored in the examined literature.

The taxonomy reveals two main branches: mismatch tolerance mechanisms and domain-specific compression methods. The original paper sits in the former, which focuses on algorithmic robustness when predictions diverge. The neighboring 'Large Language Model Output Compression' leaf (one paper) represents domain-specific applications that may encounter similar mismatch issues but do not explicitly address tolerance mechanisms. The taxonomy's scope notes clarify that standard arithmetic coding without mismatch handling belongs outside this branch, emphasizing that PMATIC's contribution lies in extending interval coding to accommodate bounded discrepancies—a boundary that appears sparsely populated in the current taxonomy structure.

Across three identified contributions, the literature search examined 24 candidate papers total, with no refutable pairs found. The formalization of prediction mismatch examined 10 candidates (0 refutable), the PMATIC algorithm examined 4 candidates (0 refutable), and the theoretical bounds examined 10 candidates (0 refutable). These statistics indicate that among the limited set of semantically similar papers reviewed, none provided overlapping prior work on mismatch-tolerant interval coding. The absence of refutations across all contributions, combined with the sparse taxonomy leaf, suggests the work occupies a relatively novel intersection of probabilistic compression and robustness to model discrepancies.

Given the limited search scope (24 candidates from top-K semantic retrieval), this analysis captures nearby work but cannot claim exhaustive coverage of the compression literature. The taxonomy structure and contribution-level statistics consistently point toward a sparse research area, though a broader survey might reveal related work in adjacent fields (e.g., distributed source coding, robust arithmetic coding) not captured by the semantic search. The findings reflect what is visible within the examined candidate set, not a definitive statement on absolute novelty across all compression research.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: Formalization of prediction mismatch problem in model-driven compression

Description: The authors formally define the prediction mismatch problem that arises when encoder and decoder use the same model but obtain slightly different probability predictions due to non-determinism, which can cause cascading decoding failures in arithmetic coding-based compression.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Deep Lossless Compression Algorithm Based on Arithmetic Coding for Power Data

URL: [View paper](#)

Brief Assessment

Arithmetic Coding Power[13] focuses on applying deep learning models (Bi-LSTM and Transformers) with arithmetic coding for power data compression. It does not address or formalize the prediction mismatch problem between encoder and decoder.

2. Lossless Image Compression Using Context-Dependent Linear Prediction Based on Mean Absolute Error Minimization

URL: [View paper](#)

Brief Assessment

Context Linear Prediction[6] focuses on lossless image compression using linear prediction methods and does not address prediction mismatch between encoder and decoder in model-driven compression with arithmetic coding.

3. Digital Image Compression

URL: [View paper](#)

Brief Assessment

Digital Image Compression[7] discusses prediction errors in image compression contexts but does not formalize the prediction mismatch problem arising from non-determinism in neural network-based arithmetic coding systems.

4. Fast Point Cloud Geometry Compression with Context-based Residual Coding and INR-based Refinement

URL: [View paper](#)

Brief Assessment

Point Cloud Geometry[10] focuses on point cloud compression using arithmetic coding with spatial contexts, not on prediction mismatch between encoder/decoder models due to non-determinism.

5. Climate science data can be compressed efficiently by dual-stage extreme compression with a variational auto-encoder transformer

URL: [View paper](#)

Brief Assessment

Climate Dual-Stage[11] focuses on atmospheric data compression using variational auto-encoders and does not address prediction mismatch problems in arithmetic coding-based compression systems.

6. DNACoder: a CNN-LSTM attention-based network for genomic sequence data compression

URL: [View paper](#)

Brief Assessment

DNACoder[8] mentions arithmetic coding's sensitivity to errors in compressed data but does not formalize the prediction mismatch problem arising from non-deterministic probability predictions between encoder and decoder in model-driven compression.

7. Variable-Bitrate Neural Compression via Bayesian Arithmetic Coding

URL: [View paper](#)

Brief Assessment

Variable Bitrate Neural[9] addresses compression of latent representations in variational autoencoders using adaptive discretization, not the prediction mismatch problem between encoder/decoder in arithmetic coding-based compression systems.

8. Lossless Image Coding Using Non-MMSE Algorithms to Calculate Linear Prediction Coefficients

URL: [View paper](#)

Brief Assessment

Non-MMSE Prediction[14] focuses on lossless image compression using linear prediction coefficients and arithmetic coding for prediction errors, not on the prediction mismatch problem between encoder and decoder in model-driven compression systems.

9. Enhanced Color Palette Modeling For Lossless Screen Content Compression

URL: [View paper](#)

Brief Assessment

Color Palette Modeling[12] focuses on lossless screen content image compression using color palette and context modeling stages, not on prediction mismatch problems in model-driven compression with arithmetic coding.

10. Adaptive Context Modeling for Arithmetic Coding Using Perceptrons

URL: [View paper](#)

Brief Assessment

Adaptive Context Modeling[15] focuses on online perceptron-based context modeling for arithmetic coding, not on prediction mismatch between encoder and decoder due to non-determinism in neural networks.

Contribution 2: PMATIC algorithm for mismatch-tolerant compression

Description: The authors propose PMATIC, a drop-in replacement for arithmetic coding that quantizes probabilities into bins and uses helper bits to ensure encoder-decoder agreement on probability distributions despite bounded prediction mismatch.

This contribution was assessed against **4 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Robust 1-bit compressed sensing via hinge loss minimization

URL: [View paper](#)

Brief Assessment

Robust 1-bit Sensing[3] focuses on 1-bit compressed sensing with quantization for signal recovery, not lossless data compression with probability matching for encoder-decoder synchronization.

2. Universally quantized neural compression

URL: [View paper](#)

Brief Assessment

Universally Quantized Neural[2] addresses quantization in lossy image compression using universal quantization to eliminate train-test mismatch, while PMATIC focuses on lossless text compression with bounded prediction mismatch between encoder-decoder LLM predictions. These are fundamentally different compression paradigms (lossy vs. lossless) with different technical challenges.

3. Mixture of Quantized Experts (MoQE): Complementary Effect of Low-bit Quantization and Robustness

URL: [View paper](#)

Brief Assessment

Mixture Quantized Experts[4] focuses on weight quantization for Mixture of Experts models to reduce memory consumption, not on mismatch-tolerant compression algorithms using probability quantization and helper bits.

4. Robust 2-bit Quantization of Weights in Neural Network Modeled by Laplacian Distribution

URL: [View paper](#)

Brief Assessment

Robust 2-bit Quantization[5] focuses on quantizing neural network weights using Laplacian distribution models and logarithmic companding, not on probability matching or mismatch-tolerant compression algorithms for lossless data compression.

Contribution 3: Theoretical correctness and performance bounds for PMATIC

Description: The authors establish formal guarantees showing PMATIC correctly decodes when conditional total variation distance is bounded, and derive theoretical bounds on the compression overhead required to achieve mismatch robustness.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Compressive quantum waveform estimation

URL: [View paper](#)

Brief Assessment

Compressive Quantum Waveform[23] addresses quantum sensor measurements for waveform estimation using compressive sensing mathematics, not probabilistic compression with prediction mismatch. The domains and technical problems are entirely different.

2. Improving prediction-based lossy compression dramatically via ratio-quality modeling

URL: [View paper](#)

Brief Assessment

Ratio-Quality Modeling[19] focuses on analytical models for error-bounded lossy compression in scientific data, not lossless compression with prediction mismatch robustness guarantees.

3. Nonlinear-by-Linear: Guaranteeing Error Bounds in Compressive Autoencoders

URL: [View paper](#)

Brief Assessment

Nonlinear-by-Linear[22] focuses on error-bounded autoencoder compression for scientific/image data with instance-specific guarantees, not probabilistic model-driven lossless compression with prediction mismatch tolerance.

4. Optimization of error-bounded lossy compression for hard-to-compress HPC data

URL: [View paper](#)

Brief Assessment

Error-Bounded HPC[21] focuses on error-bounded lossy compression for scientific HPC data with adaptive partitioning and XOR optimization, not on correctness guarantees for robust compression under prediction mismatch in model-driven lossless compression.

5. Adaptive information-constrained mapping for feature compression in edge AI and federated systems

URL: [View paper](#)

Brief Assessment

Adaptive Information Constrained[16] focuses on entropy-regularised feature compression in edge AI and federated systems, not on prediction mismatch in model-driven lossless compression. The candidate addresses compression quality functionals and projection-gradient optimisation for feature reduction, which is a different technical domain from PMATIC's arithmetic coding robustness guarantees.

6. FedSZ: Leveraging error-bounded lossy compression for federated learning communications

URL: [View paper](#)

Brief Assessment

FedSZ[17] focuses on lossy compression for federated learning model updates with error bounds, not on correctness guarantees for lossless compression under prediction mismatch as in PMATIC.

7. Towards Guaranteeing Error Bound in DCT-based Lossy Compression

URL: [View paper](#)

Brief Assessment

DCT Error Bound[24] focuses on error-controlled lossy compression for scientific simulation data using DCT transforms, not on correctness guarantees for lossless compression under prediction mismatch.

8. ISABELA for effective in situ compression of scientific data

URL: [View paper](#)

Brief Assessment

ISABELA[20] focuses on B-spline-based lossy compression for scientific data with error-bounded guarantees, not on prediction mismatch robustness in model-driven compression. The theoretical frameworks address fundamentally different problems.

9. On the horizontal compression of dag-derivations in minimal purely implicational logic

URL: [View paper](#)

Brief Assessment

DAG Derivations[25] focuses on horizontal compression of dag-derivations in minimal purely implicational logic, not on probabilistic compression algorithms or prediction mismatch robustness.

10. Significantly improving lossy compression for scientific data sets based on multidimensional prediction and error-controlled quantization

URL: [View paper](#)

Brief Assessment

Multidimensional Prediction[18] focuses on lossy compression for scientific datasets using prediction-based quantization with error bounds on reconstruction accuracy (RMSE, NRMSE, PSNR), not on lossless compression with correctness guarantees under prediction mismatch between encoder and decoder.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Synchronizing Probabilities in Model-Driven Lossless Compression [View paper](#)
- [1] Lossless Compression of Large Language Model-Generated Text via Next-Token Prediction [View paper](#)
- [2] Universally quantized neural compression [View paper](#)
- [3] Robust 1-bit compressed sensing via hinge loss minimization [View paper](#)
- [4] Mixture of Quantized Experts (MoQE): Complementary Effect of Low-bit Quantization and Robustness [View paper](#)
- [5] Robust 2-bit Quantization of Weights in Neural Network Modeled by Laplacian Distribution [View paper](#)
- [6] Lossless Image Compression Using Context-Dependent Linear Prediction Based on Mean Absolute Error Minimization [View paper](#)
- [7] Digital Image Compression [View paper](#)
- [8] DNACoder: a CNN-LSTM attention-based network for genomic sequence data compression [View paper](#)
- [9] Variable-Bitrate Neural Compression via Bayesian Arithmetic Coding [View paper](#)
- [10] Fast Point Cloud Geometry Compression with Context-based Residual Coding and INR-based Refinement [View paper](#)
- [11] Climate science data can be compressed efficiently by dual-stage extreme compression with a variational auto-encoder transformer [View paper](#)
- [12] Enhanced Color Palette Modeling For Lossless Screen Content Compression [View paper](#)
- [13] Deep Lossless Compression Algorithm Based on Arithmetic Coding for Power Data [View paper](#)
- [14] Lossless Image Coding Using Non-MMSE Algorithms to Calculate Linear Prediction Coefficients [View paper](#)
- [15] Adaptive Context Modeling for Arithmetic Coding Using Perceptrons [View paper](#)
- [16] Adaptive information-constrained mapping for feature compression in edge AI and federated systems [View paper](#)
- [17] FedSZ: Leveraging error-bounded lossy compression for federated learning communications [View paper](#)
- [18] Significantly improving lossy compression for scientific data sets based on multidimensional prediction and error-controlled quantization [View paper](#)
- [19] Improving prediction-based lossy compression dramatically via ratio-quality modeling [View paper](#)
- [20] ISABELA for effective in situ compression of scientific data [View paper](#)
- [21] Optimization of error-bounded lossy compression for hard-to-compress HPC data [View paper](#)
- [22] Nonlinear-by-Linear: Guaranteeing Error Bounds in Compressive Autoencoders [View paper](#)
- [23] Compressive quantum waveform estimation [View paper](#)
- [24] Towards Guaranteeing Error Bound in DCT-based Lossy Compression [View paper](#)
- [25] On the horizontal compression of dag-derivations in minimal purely implicational logic [View paper](#)