Quantum-Enhanced Part-of-Speech Tagging with Gated Recurrent Units

Ashutosh Rai
National Institute of Technology Silchar
ashutoshrai_pg_24@cse.nits.ac.in

Shyambabu Pandey
National Institute of Technology Silchar
shyambabu21_rs@cse.nits.ac.in

Partha Pakray
National Institute of Technology Silchar
partha@cse.nits.ac.in

Abstract

Deep learning models for Natural Language Processing (NLP) tasks, such as Part-of-Speech (POS) tagging, usually have significant parameter counts that make them costly to train and deploy. Quantum Machine Learning (QML) offers a potential approach for building more parameter-efficient models. This paper proposes a hybrid quantum-classical model for POS tagging in code-mixed social media text. By integrating a quantum layer into the recurrent framework, our model achieved an accuracy comparable to the baseline classical model, while needing fewer parameters. Although the cut-off point in the parameters is modest in our setup, the approach is promising when scaled to deeper architectures. These results suggest that hybrid models can offer a resource-efficient alternative for NLP tasks.

1 Introduction

Understanding natural human language, which is a central basis of communication, has been a long-standing goal of artificial intelligence (Russell and Norvig, 2010). Natural language processing (NLP) successfully tackles this problem by developing methods for machines to read, examine, and produce natural language in ways that support tangible real-world applications (Jurafsky, 2000). Today, NLP supports applications such as conversational assistants, automatic translation systems, and opinion mining tools, making it an important part of our daily engagement with digital technology.

The recent success of NLP is mainly attributed to improvements in machine learning (Janiesch et al., 2021). Training models on large amounts of data makes them capable of learning and recognizing patterns in text and making accurate predictions for tasks like translation, sentiment analysis, and sequence labeling. Neural networks, specifically, have brought about significant developments by modeling complex relationships within language data (Sharkawy, 2020). However, as data sets grow larger and architectures deeper, these models become resource intensive, requiring large amounts of memory and computation for both training and inference (Janiesch et al., 2021).

Quantum computing is one avenue that offers a possible way forward by providing a different and more efficient method of computation (Gyongyosi and Imre, 2019). Quantum characteristics such as superposition and entanglement are essential to how information can be represented and operated on with much greater expressive power than classical bits allow. Based on these principles, quantum machine learning (QML) has emerged as a research field that seeks to merge quantum computation with machine learning methods (Schuld and Petruccione, 2021). Although a nascent field, QML has been explored as an alternative to designing more compact models that can capture patterns differently from their classical counterparts.

Putting these principles into practice, this work solves an important NLP task: part-of-speech (POS) tagging in code-mixed text data from social networks (Pandey et al., 2023). POS tagging works by assigning grammatical roles to each word in a sentence and is a crucial step in syntactic and semantic analysis (Basisth et al., 2023). We present a hybrid quantum gated recurrent units (QGRU) model that integrates a quantum layer with classical recurrent layers. To evaluate the performance of the proposed model, we perform POS tagging on a code-mixed dataset. Based on our findings, this approach competes with classical baselines in accuracy but achieves similar performance with fewer trainable parameters, making it parameter efficient. Still, the approach suggests that greater savings could be realized when scaling to larger architectures, where substituting intermediate layers with quantum circuits may yield noticeable efficiency gains.

The structure of this paper is as follows. Section 2 introduces background on quantum computing and QML, Section 3 reviews related research, Section 4 details the proposed model, Section 5 describes the dataset, Section 6 reports results and analysis, and Section 7 concludes with future directions.

2 Background

2.1 Quantum Computing

Quantum computing is a paradigm of computation that uses the principles of quantum mechanics to process information in ways that are not possible with classical systems (Gyongyosi and Imre, 2019). In a classical computer, the basic unit of information is the bit, which can take one of two values, 0 or 1. In quantum computing, the basic unit is the quantum bit, or qubit. A qubit has two basic states, $|0\rangle$ and $|1\rangle$, which are called computational basis states. These basis states are commonly represented in vector form as

$$|0\rangle = \begin{bmatrix}1\\0\end{bmatrix}, \quad |1\rangle = \begin{bmatrix}0\\1\end{bmatrix} \tag{1}$$

Unlike a classical bit, which can only be 0 or 1 at a time, a qubit can exist in a superposition of both states. The state of a single qubit can be expressed as

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle, \quad \text{with } |\alpha|^2 + |\beta|^2 = 1 \tag{2}$$

where α and β are complex amplitudes. The normalization condition ensures that the total probability of measuring the qubit in either state is one. When multiple qubits are combined, they form a joint system described by the tensor product of individual qubit states. For example, the state of two qubits $|\psi\rangle \otimes |\phi\rangle$ can be written as

$$|\psi\phi\rangle = \alpha_{00}|00\rangle + \alpha_{01}|01\rangle + \alpha_{10}|10\rangle + \alpha_{11}|11\rangle \tag{3}$$

This shows that a two-qubit system can represent all four possible basis states at the same time. In general, a n-qubit system can represent 2^n states in parallel, which provides exponential representational power compared to classical bits (Pandey and Pakray, 2023). Another important property is entanglement. Entangled qubits are correlated in such a way that the state of one qubit cannot be described independently of the other. For instance, an entangled two-qubit system may be described as

$$|\Phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{4}$$

In this state, measuring the first qubit immediately determines the outcome of the second. Entanglement enables forms of information processing that are not possible with classical systems.

Quantum operations are carried out using quantum gates, which are unitary matrices that transform qubit states while preserving normalization. For example, the Hadamard gate H creates a superposition state:

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle), \quad H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \tag{5}$$

By combining such gates into circuits, quantum computers can implement a wide variety of computations. At the end of a computation, qubits are measured, and the superposition collapses into one of the basis states, with probabilities determined by the amplitudes.

Together, these basic elements, qubits, superposition, entanglement, quantum gates, and measurement, form the foundation of quantum computing. They allow quantum systems to process and represent information in fundamentally different ways than classical systems, opening possibilities for speed-ups in certain computational tasks.

2.2 Quantum Machine Learning

Quantum machine learning (QML) is an emerging area of research that combines the principles of quantum computing with machine learning (Schuld and Petruccione, 2021). The goal is to take advantage of the unique properties of quantum computation to help improve the process of learning from data. While classical machine learning relies on algorithms that run on conventional hardware, QML explores how quantum states and operations can be used to represent and process information.

In general, a QML model makes use of quantum circuits whose parameters can be adjusted during training, similar to how weights are updated in a neural network. After a computation, quantum systems are measured, and the results are expressed as expectation values of observables. The outcome of a quantum measurement is typically expressed as the expectation value of an observable. For a quantum state $|\psi\rangle$ and an observable Z, the expectation value is defined as

$$\langle Z \rangle = \langle \psi | Z | \psi \rangle \tag{6}$$

The output expectation values can now be used in the same way that the output of a classical model would be used, for instance, in calculating a loss function during training.

The major advantages of QML include fewer parameters, high-dimensional solution spaces, and the possibility of forming correlations through entanglement that is not possible while using classical models. Quantum methods can also provide performance gains for specific computation-related tasks. However, these gains are highly dependent on the application at hand and the current constraints of quantum hardware. Currently, most QML methods are implemented in a hybrid manner, where quantum circuits are merged with classical machine learning components and trained using standard optimization methods (Sweke et al., 2020).

Recent work shows that there is growing interest in applying QML to domains such as optimization, quantum chemistry, and NLP (Pandey et al., 2023). NLP tasks, in particular, are challenging due to their heavy dependence on large datasets and complex models with deep architectures, making them a viable area of exploration for possible benefits of QML. This motivates exploring QML in problems such as POS tagging, where both efficiency and performance are important considerations.

3 Related Work

POS tagging is a fundamental task in NLP, serving as a foundational step for many downstream applications. The classical state-of-the-art for sequence labeling tasks such as POS tagging has been dominated by recurrent neural networks, particularly Bidirectional Long Short-Term Memory (BiLSTM) and Gated Recurrent Unit (GRU) architectures, often paired with a Conditional Random Field (CRF) layer (Lample et al., 2016). However, such models are typically parameter heavy and their application to noisy code-mixed social media text presents many challenges (Jamatia et al., 2015).

Coecke et al. (Coecke et al., 2020) introduced a grammar-aware compositional DisCoCat framework that maps the sentence structure to quantum circuits. This addresses directly the growing computational demands of traditional machine learning and deep learning models by leveraging quantum circuits for language tasks. Our work, in contrast, integrates a variational quantum algorithm in the form of a parameterized circuit directly into a deep learning model.

Many studies have shown that such a hybrid approach is valid for NLP tasks Pandey et al. (2024). Another work by Shi et al. Shi et al. (2023) details a quantum-inspired neural network that uses complex-valued embeddings to capture better semantic information.These works showcase the potential of using quantum principles to enhance classical NLP architectures.

A more recent development with a direct relation to our task, the application of quantum circuits to POS tagging, is demonstrated by Di Sipio et al. Di Sipio et al. (2022). The authors introduced a Quantum Long-Short-Term Memory (QLSTM) model applied to a sequence tagging task. This foundational work was extended by Pandey et al. Pandey et al. (2022) in a low-resource language. The same group later modified the QLSTM model specifically for code-mixed social media data (Pandey et al., 2023) and advanced it by making a bidirectional variant (BiQLSTM) (Pandey and Pakray, 2023)).

4 Architecture

This section discusses the architectures of the two models that are compared in our study, a fully classical model, which serves as our baseline, and the proposed hybrid quantum-classical model.

4.1 Classical Model

We chose to use a standard architecture for our baseline model. It is built using Gated Recurrent Units (GRU) (Cho et al., 2014). The model takes as input sequences 100-dimensional word embeddings and processes them via two bidirectional GRU layers with a hidden state dimension of 16. Bidirectionality allows the GRU layers to capture context-aware representations by processing the information from both preceding and succeeding tokens in the sequence. The output obtained from the GRU layers is passed through a fully connected classification head, which helps map the hidden states to a dimension corresponding to the number of POS tags.

The output of the fully connected layer is passed to a Conditional Random Field (CRF) layer which produces the final tag sequence (Lample et al., 2016). CRF is a statistical modeling method that learns transition probabilities between adjacent tags to support sequence tagging tasks. This helps the model to consider the context of neighboring predictions based on which the model can penalize grammatically unlikely tag sequences, thereby improving the accuracy and coherence of the output.

4.2 Hybrid Model

Our proposed hybrid model uses the same core layers as the baseline models. Embedding, Bidirectional GRU and the CRF layers are used in the hybrid model as well. The only distinction is the quantum layer that replaces the fully-connected classification head. The quantum layer receives its input from the fully connected layer attached to the GRU layers. The main purpose of this fully connected layer is to downsize the output from the GRU layers to match the input size of the quantum layer. It was included in the baseline model to ensure consistency.

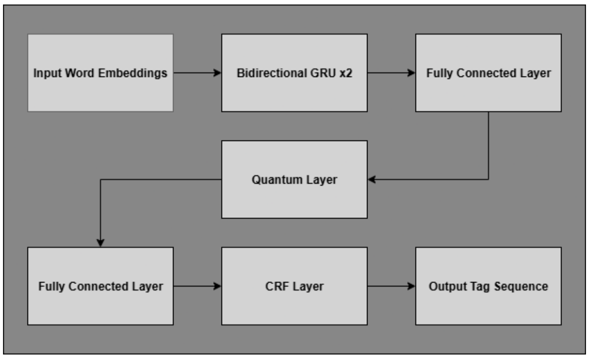

Figure 1: Architecture of the Hybrid Model.

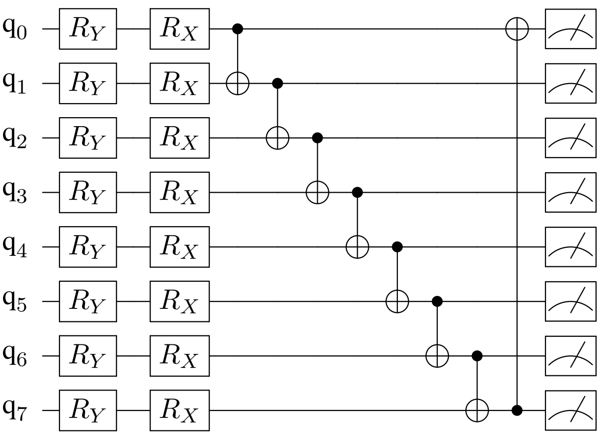

Figure 2: The 8-qubit variational quantum circuit. The initial R_Y gates are parameterized by input features, and the R_X gates are parameterized by trainable weights. This entire entangling block is repeated 6 times.

Algorithm 1 Quantum Circuit Layer

Input: Classical feature vector $x \in \mathbb{R}^8$, quantum circuit weights W.
Output: Expectation values vector $E \in \mathbb{R}^8$.
Initialize 8-qubit state to $|0\rangle^{\otimes 8}$.
Encode x into the state using AngleEmbedding.
Apply the variational BasicEntanglingLayers circuit parameterized by weights W.
for $i = 0$ 7 **do**
    Measure $\langle \sigma_z \rangle$ on qubit i.
    $E_i \leftarrow \langle \sigma_z \rangle_i$.
end for
return E.

The input to the quantum layer is an 8 dimensional vector This vector is encoded and processed by a variational quantum circuit. The quantum circuit consists of two main components. Each element in the input vector is first encoded on a qubit using an AngleEmbedding layer, a standard method for mapping feature vectors into qubit rotations. Following this, a BasicEntanglingLayers circuit is used with trainable parameters which applies one-parameter single-qubit rotations on each qubit followed by a ring of Controlled-Not (CNOT) gates, where each qubit is entangled with its neighbor, and the last qubit is connected back to the first, forming a closed chain. This circuit architecture was chosen for the quantum layer to strike a practical balance between circuit expressibility and parameter efficiency. Methods for evaluating the effectiveness of such circuits are an active area of research (Sim et al., 2019).

The operation of the quantum circuit computation is discussed in Algorithm 1. After applying the basic entanglement layer, we measure the qubits to output classical values. The resulting 8-dimensional output vector of Pauli-Z expectation values is then mapped to the tag space by a final linear layer, which provides the input logits for the CRF layer for the final tag prediction.

5 Dataset and Preprocessing

5.1 Corpus Description

The data set used in our experiments is a social media corpus of code-mixed Hindi-English text. It was originally collected and annotated by Jamatia et al. (2015). The corpus consists of messages from the IIT Bombay Facebook Confession page, which contains informal posts and chat-like comments. This type of data presents unique challenges for NLP tasks due to non-standard grammar, transliterated spellings, and informal language.

The data set used is a component of a larger corpus that also includes WhatsApp and Twitter data and covers other pairs of Indian languages such as Bengali-English and Telugu-English (Pandey et al., 2023). However, this study focuses exclusively on the Hindi-English Facebook portion. The language distribution at the token level, as reported by the original authors, is shown in Table 1. It highlights that the text is predominantly English, with a significant presence of Hindi and language-independent universal tokens, such as punctuation. The data set is annotated with a coarse-grained POS tagset, which combines universal tags with categories specific to the text of social networks. This tagset, which comprises 11 unique tags, is described in Table 2. Our data set loading process yielded a total of **1069** sentences.

5.2 Preprocessing and Data Representation

For feature representation, each word in the corpus was mapped to a **100-dimensional vector** using precomputed embeddings for this data set. Any word not present in the embedding vocabulary was represented by a zero vector. To handle variable sentence lengths for batch processing, all sequences were standardized to a uniform length of **62 tokens** by padding shorter sequences and truncating longer ones. This length was determined on the basis of the 95th percentile of sentence lengths in the corpus. Following these preprocessing steps, the data set was partitioned into training sets (60%), validation (20%) and testing (20%), resulting in **641 samples for training, 214 for validation and 214 for testing**.

Token Language	Distribution (%)
English	75.61
Hindi	4.17
Universal	16.53
Named Entity	2.19
Acronym	1.46
Mixed	0.02
Undefined	0.01

Table 1: Token-level language distribution for the Facebook portion of the corpus, as reported by Jamatia et al. (2015).

Tag	Description
G_N	Noun
G_V	Verb
G_PRP	Pronoun
G_J	Adjective
G_R	Adverb
PSP	Pre- or Post-position
G_PRT	Particle
CC	Conjunction
G_SYM	Quantifier / Symbol
DT	Determiner
G_X	Residual / Other

Table 2: Coarse-grained POS tagset used in the dataset.

6 Experiment and Results

6.1 Experimental Setup

To evaluate our proposed model, we conducted a series of experiments to benchmark its performance against a purely classical counterpart. The models were implemented using PyTorch, with the quantum components built in Pennylane and executed on the **default qubit** simulator. The experiments compare a **classical GRU** based model against the proposed hybrid model. To ensure a fair comparison, a consistent set of hyperparameters was used to train both models, as detailed in Table 3.

Both models utilize a final Conditional Random Field (CRF) layer and were trained by minimizing its negative log-likelihood. Performance was evaluated using token-level accuracy on the held-out test set. Training was performed for a maximum of 300 epochs, with early stopping triggered if validation loss did not improve for 5 consecutive epochs.

Parameter	Value
Embedding Dimension	100
GRU Hidden Dimension	16
GRU Layers	2
Dropout Rate	0.3
Optimizer	Adam
Learning Rate	0.001
Batch Size	32
Number of Qubits	8

Table 3: Hyperparameters used for training.

6.2 Results

The final performance of both models was determined by evaluating the best-performing checkpoint, selected based on the peak validation accuracy observed during training. A summary of these results, alongside the final test accuracy and total parameter counts, is presented in Table 4. The proposed Hybrid QGRU model achieved a final test accuracy of **78.13%**, a result comparable to the **80.29%** accuracy achieved by the fully Classical GRU baseline. The central finding, however, lies in the model's efficiency. The hybrid model required only **16,682** trainable parameters to achieve this result, a modest but clear reduction of approximately **5.7%** compared to the **17,690** parameters of the classical model.

Model	Params	Val. Acc. (%)	Test Acc. (%)
Classical GRU	17,690	81.80	80.29
Hybrid QGRU	16,682	77.77	78.13

Table 4: Performance comparison of the baseline and hybrid models.

To provide a more granular analysis, Table 5 details a per-tag comparison of the F1-scores for both models on the test set. This breakdown reveals a nuanced performance landscape. For high-support, core grammatical categories such as **G_N** (Noun), **G_V** (Verb), and **DT** (Determiner), the hybrid model's performance is nearly identical to the classical baseline. Notably, it performs slightly better on **G_PRP** (Pronoun) tags. However, the hybrid model struggles with certain low-frequency tags, showing a significant performance drop for **CC** (Conjunction) and struggling significantly with **G_SYM** (Symbol) tags, failing to correctly classify any instance, likely due to their very low support in the test set. This suggests that while the quantum layer is effective at learning representations for common classes, it may be less robust on sparse data categories compared to its classical counterpart in this configuration.

Tag	Support	Cls. F1	Hyb. F1	Δ (Hyb-Cls)
CC	118	0.52	0.19	-0.33
DT	238	0.89	0.91	+0.02
G_J	199	0.62	0.54	-0.08
G_N	755	0.82	0.81	-0.01
G_PRP	336	0.83	0.86	+0.03
G_PRT	142	0.51	0.42	-0.09
G_R	188	0.63	0.56	-0.07
G_SYM	31	0.43	0.00	-0.43
G_V	697	0.85	0.82	-0.03
G_X	478	0.96	0.95	-0.01
PSP	339	0.81	0.78	-0.03

Table 5: Per-tag F1-score comparison on the test set. Δ indicates the change in F1-score for the hybrid model.

7 Conclusion

In this work, we addressed the challenge of high parameter counts in deep learning models for NLP by proposing and evaluating a hybrid quantum-classical Gated Recurrent Unit (QGRU). We applied this model to the task of POS tagging on code-mixed social media text, a domain characterized by noisy and non-standard language. Our findings indicate that the hybrid model achieves a test accuracy of **78.13%**, which is comparable to the **80.29%** accuracy of its classical counterpart, while requiring approximately **5.7%** fewer trainable parameters. This outcome serves as a successful proof-of-concept, demonstrating that the integration of variational quantum circuits into recurrent architectures is a viable strategy for reducing model complexity. Our work contributes to the growing field of quantum NLP by illustrating a practical approach to develop more compact and parameter-efficient models.

References

Nihar Jyoti Basisth, Tushar Sachan, Neha Kumari, Shyambabu Pandey, and Partha Pakray. 2023. An automatic pos tagger system for code mixed indian social media text. In *International conference on computational intelligence in communications and business analytics*, pages 273–286. Springer.

Kyunghyun Cho, Bart Van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. Learning phrase representations using RNN encoder-decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1724–1734, Doha, Qatar.

Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. 2020. Foundations for near-term quantum natural language processing. *arXiv preprint arXiv:2012.03755.*

Riccardo Di Sipio, Jia-Hong Huang, Samuel Yen-Chi Chen, Stefano Mangini, and Marcel Worring. 2022. The dawn of quantum natural language processing. In *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8612–8616. IEEE.

Laszlo Gyongyosi and Sandor Imre. 2019. A survey on quantum computing technology. *Computer Science Review*, 31:51–71.

Anupam Jamatia, Björn Gambäck, and Amitava Das. 2015. Part-of-speech tagging for code-mixed

English-Hindi Twitter and Facebook chat messages. In *Proceedings of Recent Advances in Natural Language Processing*, pages 239–248, Hissar, Bulgaria.

Christian Janiesch, Patrick Zschech, and Kai Heinrich. 2021. Machine learning and deep learning. *Electronic Markets*, 31(3):685–695.

Dan Jurafsky. 2000. *Speech & Language Processing*. Pearson Education India.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California.

Shyambabu Pandey, Nihar Jyoti Basisth, Tushar Sachan, Neha Kumari, and Partha Pakray. 2023. Quantum machine learning for natural language processing application. *Physica A: Statistical Mechanics and its Applications*, 627:129123.

Shyambabu Pandey, Pankaj Dadure, Morrel V. L. Nunsanga, and Partha Pakray. 2022. Parts of speech tagging towards classical to quantum computing. In *2022 IEEE Silchar Subsection Conference (SILCON)*, pages 1–6. IEEE.

Shyambabu Pandey and Partha Pakray. 2023. Bi-quantum long short-term memory for part-of-speech tagging. In *Proceedings of the 19th International Conference on Natural Language Processing (ICON)*, pages 272–278, Goa, India.

Shyambabu Pandey, Partha Pakray, and Riyanka Manna. 2024. Quantum classifier for natural language processing applications. *Computación y Sistemas*, 28(2):695–700.

Stuart J. Russell and Peter Norvig. 2010. *Artificial Intelligence: A Modern Approach*. Pearson Education, Inc.

Maria Schuld and Francesco Petruccione. 2021. *Machine Learning with Quantum Computers*. Springer.

Abdel-Nasser Sharkawy. 2020. Principle of neural network and its main types. *Journal of Advances in Applied & Computational Mathematics*, 7:8–19.

Jinjing Shi, Zhenhuan Li, Wei Lai, Fangfang Li, Ronghua Shi, Yanyan Feng, and Shichao Zhang. 2023. Two end-to-end quantum-inspired deep neural networks for text classification. *IEEE Transactions on Knowledge and Data Engineering*, 35(4):4335–4345.

Sukin Sim, Peter D Johnson, and Alán Aspuru-Guzik. 2019. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies*, 2(12):1900070.

Ryan Sweke, Frederik Wilde, Johannes Meyer, Maria Schuld, Paul K. Faehrmann, Barthélemy Meynard-Piganeau, and Jens Eisert. 2020. Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum*, 4:314.