Attention Mechanisms for Broadband Feature Prediction for Electromagnetic and Photonic Applications

Ergun Simsek[a], Masoud Soroush[a], Gregory Moille[b,c], Kartik Srinivasan[b,c], and Curtis R. Menyuk[a]

[a]University of Maryland Baltimore County, Baltimore, MD 21250, USA.
[b]National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA.
[c]Joint Quantum Institute, NIST/University of Maryland, College Park, MD 20742, USA.

ABSTRACT

We present a study on the accuracy of three neural network architectures, namely fully-connected neural networks, recurrent neural networks, and attention-based neural networks, in predicting the coupling response of broadband microresonator frequency combs. These frequency combs are crucial for technologies like optical atomic clocks. Optimizing their spectral features, especially the dispersion in coupling to an access waveguide, can be computationally demanding due to the large number of parameters and wide spectral bandwidths involved. To address this challenge, we employ machine learning algorithms to estimate the coupling response at wavelengths not present in the input training data. Our findings demonstrate that when trained with data sets encompassing the upper and lower limits of each design feature, attention mechanisms achieve over 90% accuracy in predicting the coupling rate for spectral ranges six times wider than those used in training. This significantly reduces the computational burden for numerical optimization in ring resonator design, potentially leading to a six-fold reduction in compute time. Moreover, devices with strong correlations between design features and performance metrics may experience even greater acceleration.

Keywords: Ring resonators, machine learning, deep learning, attention mechanism, broadband prediction

1. INTRODUCTION

We have witnessed a surge of interest in the use of artificial intelligence, specifically neural networks, to enhance the understanding and application of light-matter interactions.[1–8] Neural networks have shown promise in various photonics applications, including inverse photonic design, material and device characterization, optical sensing, image processing and classification, and optical communication.

For forward problems in photonics, such as predicting optical properties or device performance metrics, fully connected and recurrent neural networks (FCNNs and RNNs, respectively) are commonly used due to their high accuracy.[1–8] Neural networks significantly reduce computation time compared to traditional simulation-based approaches, especially when dealing with weakly non-linear problems. Our focus in this work is on the behavior of coupling quality factor (Q_c) in ring resonators, which are critical elements in optics and photonics.[9] Calculating Q_c using coupled mode theory (CMT) becomes laborious and computationally intensive when considering a wide range of frequencies and waveguide geometries. To address this challenge, we explore the use of neural networks, specifically FCNNs, RNNs, and neural networks with an attention mechanism.[10,11]

We conduct a systematic study comparing the performances of these three neural network implementations in two scenarios: interpolation and extrapolation problems. In the interpolation scenario, where the training and testing data sets are randomly split, we achieve high accuracy (96%) and significantly reduce (6 times) computation time. In the extrapolation scenario, where the neural networks predict Q_c values for new physical parameters, larger training data is required, but we still achieve notable accuracy (90%) and computational efficiency. Our study demonstrates that neural networks, particularly with the integration of attention mechanisms, are valuable tools for reducing computation time and resources in the design and optimization of photonic devices. This approach can also be extended to estimate the characteristics of moderately non-linear devices by training on a broad data set covering a wide frequency range.

Further author information: (Send correspondence to E.S.) E-mail: simsek@umbc.edu, Telephone: +1 410 455 3540.

2. CONFIGURATION AND ATTENTION MECHANISM

2.1 Configuration

The configuration investigated in this study is depicted in Fig. 1. It involves a Si_3N_4 microring with specific dimensions, including an outer radius (R) and ring width (RW), which is separated from the coupling waveguide by a gap (G) and has a pulley coupling design. The pulley length is denoted as L_c and is utilized to enhance the interaction between the waveguide and the ring. The chosen dimensions aim to achieve an octave-spanning frequency comb for optical clock applications. The ring has a fixed height of 410 nm and a ring width of approximately 855 nm, while the ring radius is set at 80 μm to target a repetition rate of 275 GHz, which can be detected using advanced high-speed photodetection techniques. A total of 5750 distinct devices are examined, systematically varying the dimensions. Table 1 provides the minimum, maximum, and step size values for RW, W, G, and L_c. The substrate is SiO_2, serving as the lower optical cladding, while the upper and side claddings consist of air.

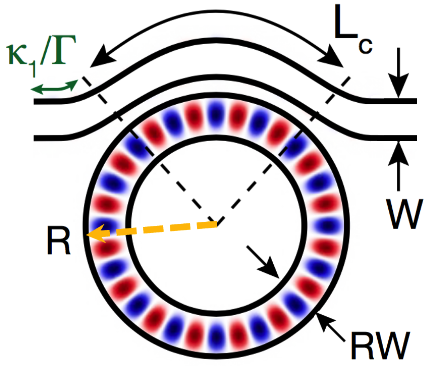

Figure 1. Schematic illustration of the pulley-coupled microring resonator.

Table 1. Simulation parameter space includes 5 unique values of RW, W, and G, and 46 distinct L_c values, for a total of 5750 unique devices.

Parameter	Minimum	Step Size	Maximum
RW	855 nm	10 nm	895 nm
W	550 nm	25 nm	650 nm
G	100 nm	50 nm	300 nm
L_c	5 μm	1 μm	50 μm

The resonator-waveguide coupling is calculated using the spatial coupled mode theory (CMT) formalism in an integrated planar geometry by considering only the region over which their fields interact.[8,9] The complete set of equations can be found in literature.[8]

In this study, our focus is primarily on the fundamental transverse-electric polarized modes of the waveguide and ring. However, it's important to note that our approach, which includes both CMT calculations and neural network (NN) predictions, can be applied to other cases as well. The primary objective of our work is to develop a highly accurate deep model that takes into account the geometric features and Q_c values at lower frequencies as input, and predicts the corresponding Q_c values at higher frequencies. When we employ traditional machine learning techniques such as linear regression or random forests to achieve this objective, we find that the accuracy is relatively low, approximately 70%. On the other hand, by utilizing fully-connected or recurrent neural networks, we are able to achieve significantly higher accuracy, 90% or higher, as discussed in the Numerical Results section. The data set and all the codes to build and test the neural networks studied in this work can be found at https://github.com/simsekergun/QcPrediction.

2.2 Attention Mechanism

In our study, we employ attention mechanisms as effective tools for constructing accurate deep models when dealing with sequential features. We assume that the features are presented as a sequence $\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \rangle$, where $\boldsymbol{x}_i \in \mathbb{R}^d$ for all i. The attention-based model consists of two neural networks. The first network, referred to as the backbone network $F_b(\cdot)$, takes the input sequence $\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n \rangle$ and generates a rich representation of it, producing a new sequence $\langle \boldsymbol{h}_1, \boldsymbol{h}_2, \cdots, \boldsymbol{h}_n \rangle$. The second network, the score network denoted as $F_s(\cdot)$, assigns scores to the elements of the input sequence based on their significance in the final output. The scores are then transformed using the softmax function to obtain a discrete probability distribution. The attention scores can be expressed as a vector $\vec{\alpha}$ obtained from the softmax function. The output of the attention-based model is calculated as the weighted sum of the backbone network outputs, where the weights are determined by the attention scores.

Score networks can be categorized into simple attention-based networks and attention-based networks with a context vector. The latter performs better as it incorporates all elements of the input sequence in determining the attention scores. The context vector $\boldsymbol{h}_c$ is computed as the average of all elements of the sequence. In our problem, we found that the additive attention score network yields the strongest performance. It involves concatenating the context vector $\boldsymbol{h}_c$ with each element of the sequence and passing it through a fully-connected linear layer followed by an activation function. To apply the attention mechanism, we organize the input data in a sequential form. The features are collected in subintervals, and for each subinterval, the corresponding $\log_{10} Q_c$ values are concatenated with the geometric features. The input variables are arranged as a sequence, which is represented as a tensor $\mathbb{X}$. We feed $\mathbb{X}^\mathsf{T}$ to the attention-based model as it yields stronger performance. The backbone network $F_b(\cdot)$ in our study consists of four hidden layers with varying numbers of neurons. Each layer includes batch normalization and a ReLU activation function.

3. NUMERICAL RESULTS

3.1 Interpolation

First, we consider the following scenario: we are given four geometric parameters $\langle RW, W, G, L_c \rangle$ and Q_c values for the frequency range 200 THz to 400 THz as the input to the model. The goal is to predict the Q_c values for 160 distinct frequencies ranging from 420 THz to 500 THz. The data set is split into two subsets, with 4750 designs used for training and 1000 designs used for testing.

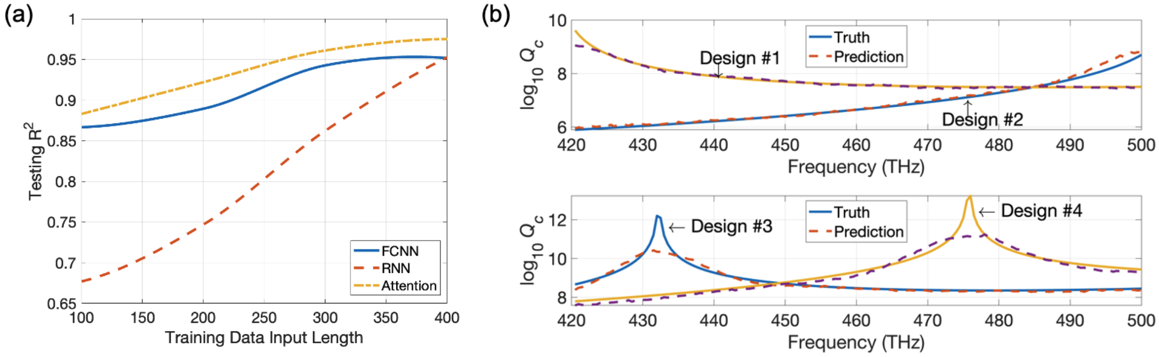

Figure 2. (a) The R^2 scores for the FCNN (blue), RNN (red), and the attention-mechanism (orange) models for the interpolation problem with different frequency bandwidths for the feature variables. (b) True (CMT-determined) vs. predicted Q_c values for two devices without (top) and with (bottom) a peak (a coupling anti-resonance) in the frequency range of interest. Here, the geometric parameters defining the resonator-waveguide coupler and the Q_c values between 200 THz and 400 THz are given as inputs in order to predict Q_c between 420 THz and 500 THz.

To evaluate the accuracy of the attention-based neural network, the mean squared error is used as the loss function. The learning rate is set to 0.001, and the activation function is chosen as ReLU. After 100 epochs, the test R^2-score reaches 96.7%. The comparison between true and predicted Q_c values for four randomly chosen devices is shown in Fig. 2(b).

In the upper plot of Fig. Fig. 2(b), there are no peaks in the true Q_c values within the selected frequency range. The attention-based model accurately predicts the Q_c values for these designs with high precision. However, in the second plot of Fig. 2(b), an anti-resonance is observed in the high-frequency range. The attention-based model correctly identifies the position of the peak but typically predicts a shorter height for the peak. This behavior can be attributed to the extreme nonlinearity of Q_c, which neural networks, incorporating well-behaved activation functions, struggle to capture. Nevertheless, the model is capable of estimating when Q_c exceeds the intrinsic loss by more than an order of magnitude, which is important for photonic design.

We studied how the accuracy changes with optimizers using Adam,[12] AdamW,[13] NAdam,[14] and RAdam[15] as the optimizers. The performance of these optimizers is comparable, but AdamW achieves slightly higher R^2-scores for the train and test subsets. Hence, AdamW is selected as the preferred optimizer.

The accuracy of the model is analyzed in relation to the frequency bandwidth of the feature variables and plotted in Fig. 2(a). The attention-based neural network, as well as FCNN and RNN models, are implemented. As the sequence length decreases, the accuracy of the RNN model drops faster compared to FCNN and attention-based NN. Even with the shortest interval, the attention-based NN achieves a reasonable accuracy of approximately 88%. The reduction in computing time provided by the attention-based model can be significant for computationally intensive studies and applications. Overall, the implementation of an attention mechanism improves the accuracy of the RNN-type model in predicting Q_c values for photonic designs.

3.2 Extrapolation

To study extrapolation, we changed our approach to creating training and test data sets. Instead of random splitting, we specifically selected values of RW and W for training, and for testing, we chose values of RW and W that were beyond the ranges used in training. The training data set consisted of 3680 devices, while the testing data set had 230 devices. Our analysis showed that the accuracy of the model was high when there was no anti-resonance peak, and it accurately estimated the width of the anti-resonance when it was present, see Fig. 3(b).

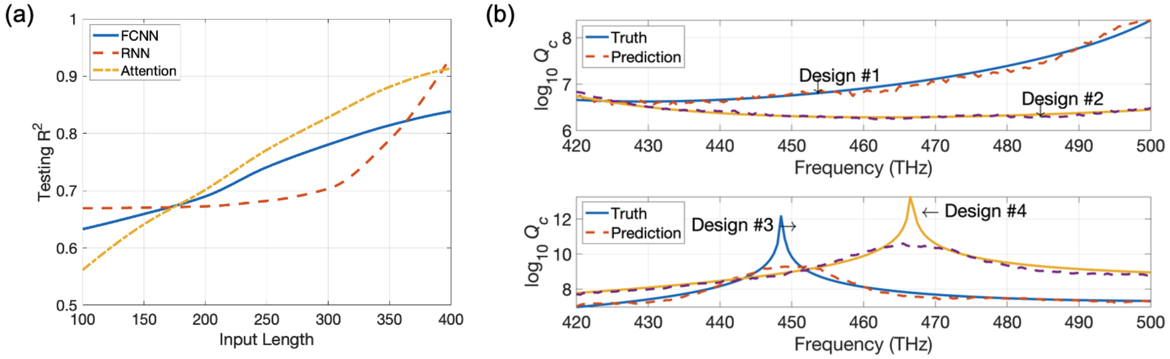

Figure 3. Follows Fig. 2 for the extrapolation problem.

We compared the performance of different models in terms of accuracy using the test R^2-score, as shown in Fig. 3(a). The FCNN and the attention-based neural network consistently improved their accuracy as the training data set size increased. The attention-based neural network outperformed the FCNN when the training data set was sufficiently large. However, the performance of the RNN did not show the same steady improvement and only started to improve when the training data set included at least 300 samples. For extrapolation problems, achieving high accuracy requires a large number of samples in the training data. Due to its stable performance, we recommend using attention-based neural network models for extrapolation applications that demand high accuracy.

In terms of computational efficiency, we observed a reduction in computation time of 33% when using the trained model to predict Q_c values for the 400 THz to 500 THz range, with an accuracy of 91%. This reduction was not as significant as observed in the interpolation problem. When choosing geometric parameters far from those used in training, the accuracy is likely to decrease, depending on the correlation between input parameters and output (in this case, the Q_c value). To maintain accuracy, it is advisable to prepare training data that spans the entire testing design space and samples parameters at moderate rates. The optimal sampling rate selection is beyond the scope of this study and can be explored in related literature. The reduction in computing time while maintaining accuracy depends heavily on the non-linearity level of the design under investigation. Training data needs to be large enough to capture the influence of each feature on the output and wide enough in the frequency domain to capture the unique signature of each device. Smaller fractions of an octave may be sufficient for devices with weaker nonlinearity, leading to even higher acceleration in computation.

4. CONCLUSION

In this study, we explore the use of various deep learning architectures, including fully-connected neural networks, recurrent neural networks, and attention mechanisms, to create predictive models for the behavior of the frequency-dependent coupling quality factor of pulley-coupled microring resonators. The attention mechanism, a more recent deep learning architecture, proves to be the most accurate and stable in predicting the coupling quality factor. We provide a detailed explanation of how attention mechanisms are implemented in constructing predictive deep learning models. The models are used for both interpolation and extrapolation problems, depending on how the training and testing data sets are arranged. We investigate the precision of the predicted results in relation to the bandwidth of the input frequencies. The accuracy of the attention-based models consistently improves as the size of the training data set increases. For the interpolation problem, the attention mechanisms achieve over 90% accuracy in predicting the coupling efficiency for spectral ranges six times wider than those used in the training data. This indicates the potential for a significant reduction in computational time during large-scale numerical optimization studies. Based on our findings, we conclude that deep learning models, once trained with sufficiently large data sets, offer a promising approach to accelerate spectral studies in electromagnetics and photonics. The physics-agnostic nature of the approach allows for its application to a wide range of problems.

ACKNOWLEDGMENTS

Gregory Moille and Kartik Srinivasan acknowledge support from the NIST-on-a-chip and DARPA APHI programs. C. R. M acknowledges support from the AFOSR grant (FA9550-19-S-0003).

REFERENCES

[1] Christensen, T., Loh, C., Picek, S., Jakobović, D., Jing, L., Fisher, S., Ceperic, V., Joannopoulos, J. D., and Soljačić, M., "Predictive and generative machine learning models for photonic crystals," *Nanophotonics*, **9** (13), 4183–4192 (2020).

[2] Chugh, S., Gulistan, A., Ghosh, S., and Rahman, B. M. A., "Machine learning approach for computing optical properties of a photonic crystal fiber," *Opt. Express* **27**, 36414–36425 (2019).

[3] Singh, R., Agarwal, A., and Anthony, B. W., "Mapping the design space of photonic topological states via deep learning," *Opt. Express* **28** (19), 27893–27902 (2020).

[4] Li, R., Gu, X., Li, K., Li, Z., and Zhang, Z., "Predicting the Q factor and modal volume of photonic crystal nanocavities via deep learning," in *Proc. SPIE* **11903**, 1190305 (2021). DOI: 10.1117/12.2597618

[5] Sharabi, Y., Patsyk, A., Ziv, R., and Segev, M., "Deep Learning Method for Quantum Efficiency Reconstruction," *CLEO*, paper STh4J.7 (2021). DOI: 10.1364/CLEO_SI.2021.STh4J.7

[6] Simsek, E., Mahabadi, S. E. J., Carruthers, T. F., and Menyuk, C. R., "Photodetector Performance Prediction with Machine Learning," in *Proc. FiO+LS*, paper FTu6C.4 (2021). DOI: 10.1364/FIO.2021.FTu6C.4

[7] Simsek, E. "Determining Optical Constants of 2D Materials with Neural Networks from Multi-Angle Reflectometry Data," *IOP Machine Learning: Science and Technology* bf1 (2020).

[8] Soroush, M., Simsek E., Moille G., Srinivasan K., and Menyuk, C. R., “Predicting broadband resonator-waveguide coupling for microresonator frequency combs through fully connected and recurrent neural networks and attention mechanism,” *ACS Photonics* **10** (6), 1795–1805 (2023).

[9] Moille, G., Li, Q., Briles, T. C., Yu, S.-P., Drake, T., Lu, X., Rao, A., Westly, D., Papp, S. B., and Srinivasan, K., “Broadband Resonator-Waveguide Coupling for Efficient Extraction of Octave Spanning Microcombs,” *Opt. Lett.* **44** (19), 4737–4740 (2019).

[10] Bahdanau, D., Cho and K., Bengio, Y., “Neural Machine Translation by Jointly Learning to Align and Translate,” *arXiv preprint arXiv:1409.0473* (2014). Accessed on 2022-12-01.

[11] Raff, E., *Inside Deep Learning, Math, Algorithms, Models*; Manning Publications (2021).

[12] Kingma, D. and Ba, J. Adam, “A Method for Stochastic Optimization,” *arXiv preprint arXiv:1412.6980* (2014). Accessed on 2022-12-01.

[13] Loshchilov, I. and Hutter, F., “Decoupled Weight Decay Regularization,” *arXiv preprint arXiv:1711.05101* (2017). Accessed 2022-12-01.

[14] Dozat, T., “Incorporating Nesterov Momentum into ADAM,” in *Proc. ICLR Workshop,* San Juan, Puerto Rico (2016).

[15] Liu, L., Jiang, H., He, P., Chen, W., Liu, X., Gao, J., and Han, J., “On the Variance of the Adaptive Learning Rate and Beyond,” *arXiv preprint arXiv:1908.03265* (2019). Accessed on 2022-12-01.