Machine Learning Compensation of Fiber Nonlinear Noise

Marina Medhat Melek (Mmrmelek@uwaterloo.ca)
University of Waterloo https://orcid.org/0000-0003-1665-3383
David Yevick
University of Waterloo

Research Article

Keywords: Optical fiber, Optical Communication, Nonlinear Noise, Kerr Effect, Artificial Intelligence, Machine Learning, Optical Nonlinearity

Posted Date: April 4th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1448350/v1

Machine Learning Compensation of Fiber Nonlinear Noise

Marina M. Melek[1], and David Yevick[1]

[1]*Department of Physics, University of Waterloo, ON N2L 3G1, Canada*

Abstract:

This paper examines artificial intelligence (AI) methods for compensating the distortion experienced in optical communication systems resulting from fiber nonlinearity. To identify the resulting degree of improvement afforded by machine learning methods, the procedures are applied to a model of a typical single frequency optical communication system with, a 3200km fiber length, double polarization, and a 16-QAM modulation format. The performance of transmitters and receivers that incorporate Neural Networks (NNs) are in particular examined for different values of the nonlinear coefficient (γ). Both of these are found to improve the system Q-factor for all values of γ although the degree of enhancement is dependent on the signal to noise ratio. The structures studied include Siamese neural networks (SNN) implemented at the receiver end and two-stage architectures that employ NNs at the transmitter together with a classifier at the receiver side. Here classifiers ranging from simple decision tree structures to boosting, forests, extra trees, and Multi-layer perceptron (MLP) were examined and found to provide significant enhancement for $\gamma > 4W^{-1}km^{-1}$. The optimal performance for highly nonlinear systems was achieved with two-stage systems with random forest or extra tree AI methods at the receiver. Empirical equations are also presented for each AI technique that relates the Q-factor enhancement to γ and the computational resource requirements (the number of included triplet terms).

Keywords: Optical fiber, Optical Communication, Nonlinear Noise, Kerr Effect, Artificial Intelligence, Machine Learning, Optical Nonlinearity

1. Introduction:

High-capacity optical communication single-mode optical fiber systems are often limited by nonlinear distortion in the optical fiber transmission medium. This distortion is a phase and polarization rotation noise that is determined by the field intensity propagating along the optical fiber. Prior work has demonstrated that the mean phase variation resulting from cross-phase modulation (XPM) can be compensated by a judicious choice of the averaging window in carrier phase estimation (CPE) from the observed XPM amplitude. Hence CPE can significantly improve system performance when the nonlinearity accumulation during transmission is sufficiently small. However, for strong nonlinear signal distortion, the benefit of CPE diminishes because simply compensating for the average phase rotation does not address the rapidly varying nonlinear phase shifts associated with self-phase modulation (SPM) (van den Borne et al. 2007; Lin et al. 2012). Several compensation techniques have accordingly been applied at both the receiver and transmitter (Fisher et al. 1983; Tkach 2010). The most common

procedure, digital backpropagation (DBP), employs the coupled nonlinear Schrödinger equation (CNLSE) to model the propagation of the two orthogonal polarizations through the fiber (Ip and Kahn 2008; Liga et al. 2014). Typically, the split-step Fourier method (SSFM) (Agrawal 2007) is applied to the Manakov equation as an approximation to the CNLSE, namely,

$$\frac{\partial}{\partial z}E_x(t,z) = -\frac{\alpha}{2}E_x(t,z) + \frac{i\beta_2}{2}\frac{\partial^2}{\partial t^2}E_x(t,z) - i\gamma\frac{8}{9}\left(|E_x(t,z)|^2 + \left|E_y(t,z)\right|^2\right)E_x(t,z) \quad (1a)$$

$$\frac{\partial}{\partial z}E_y(t,z) = -\frac{\alpha}{2}E_y(t,z) + \frac{i\beta_2}{2}\frac{\partial^2}{\partial t^2}E_y(t,z) - i\gamma\frac{8}{9}\left(\left|E_y(t,z)\right|^2 + |E_x(t,z)|^2\right)E_y(t,z) \quad (1b)$$

where $E_\sigma(t,z),\ \sigma = x,y$ are the two optical field components and α, β_2, and γ represent the fiber attenuation, nonlinear, and chromatic dispersion coefficients, respectively. By reversing the sign of the chromatic dispersion and nonlinear terms as shown in Eq.(1), the undistorted input pulse train at the transmitter can then be estimated from the received signal.

While a DBP analysis enables the nonlinear noise imparted to the signal during propagation to be largely compensated, the computational complexity is $O(N_{step}N_{FFT}\log_2(N_{FFT}))$ which depends on the total number of steps along the fiber length and the total number of symbols fed to each fast Fourier transform (FFT) block. In contrast, the complexity of small structured NN as an AI method is $O(N_{inputs})$ where N_{inputs} is the number of inputs to the NN (Melek and Yevick 2020a), illustrating that AI techniques are considerably more computationally efficient than the DBP.

Previously, AI methods such as k-nearest neighbors(Catanese et al. 2019; Gao et al. 2013; Kamalov et al. 2018; Mata et al. 2018; Wang et al. 2016b) as well as support vector machines (SVM) (Wang et al. 2016a), and dynamic deep neural networks (DDNN) (Sidelnikov et al. 2018) were employed to compute optimum decision boundaries for the received constellations in nonlinear optical communication systems. However, in the presence of nonlinear noise, many of these procedures required averaging the error over multiple transmission blocks before applying the AI algorithm, which limits their applicability to long haul propagation. For example, in the DDNN based method of (Sidelnikov et al. 2018), in order to avoid the effect of the error floor, each BER point was obtained by averaging the error rate over 15 signal block transmissions but the resulting large number of transmitted blocks considerably increases the system complexity.
In contrast, our previous studies (Melek and Yevick 2020a; Melek and Yevick 2020b), combined AI with a perturbation-based compensation technique by adding the most significant self-phase modulation (SFM) and cross-phase modulation (XPM) noise terms to the NN inputs. In the perturbation based technique, the second-order nonlinearity acting on a symbol at t = 0 could be approximated by a series of terms of the form (Tao et al. 2011),

$$\Delta E_x(z,0) = P_0^{\frac{3}{2}}\sum_{m,n}(X_nX_{m+n}^*X_m + Y_nY_{m+n}^*X_m)C_{mn} \quad (2a)$$

$$\Delta E_y(z,0) = P_0^{\frac{3}{2}} \sum_{m,n} (Y_n Y_{m+n}^* Y_m + X_n X_{m+n}^* Y_m) C_{mn} \qquad (2b)$$

where P_0 denotes the pulse peak power at the launch point, X_α, Y_α represent the symbol sequences in the x- and y-polarization channels, respectively, and the nonlinear perturbation coefficients, C_{mn}, depend on the link parameters as well as the duration and shape of the signal pulse. As shown in Eq.(2), the nonlinear noise triplets terms are of the form $X_n X_{m+n}^* X_m$, $Y_n Y_{m+n}^* X_m$, and $Y_n Y_{m+n}^* Y_m$, $X_n X_{m+n}^* Y_m$. These respectively represent intra-self-phase modulation (ISPM) when $m = n = 0$, intra-channel cross-phase modulation (IXPM) when $m \; or \; n = 0$, and intra-channel four-wave mixing (IFWM) when $m \; and \; n \neq 0$. However, the resulting procedure requires a somewhat larger number of complex multiplications compared to the proposed NN procedure of (Melek and Yevick 2020a) as indicated by the black and grey bars in Fig. (1), respectively.

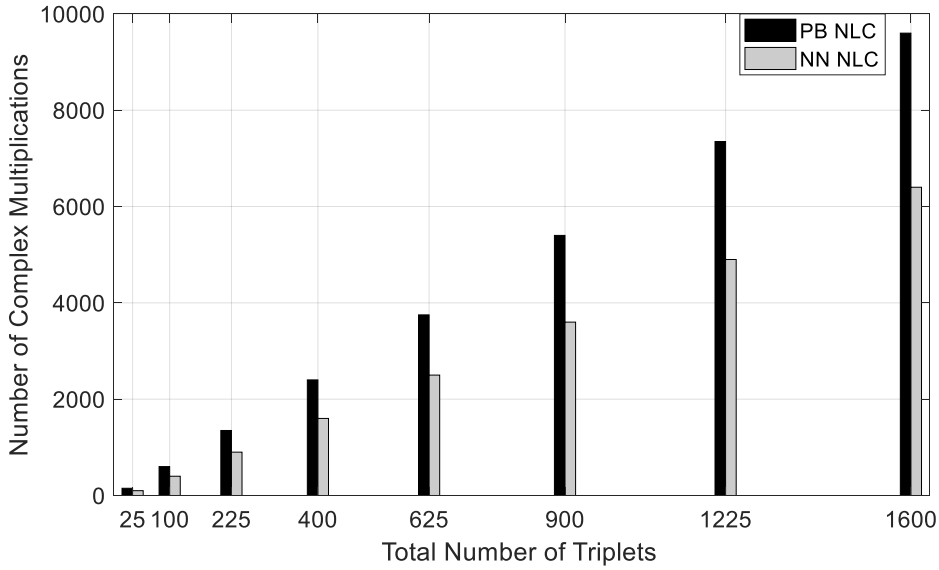

Figure1. The number of complex multiplications required by the perturbation-based compensation method (black bars) versus the proposed neural network approach in (Melek and Yevick 2020a) (grey bars)

In previous work, we investigated the degree to which nonlinear noise prediction and compensation can be enhanced by including the dominant nonlinear noise triplets terms in Eq.(2), defined as those terms for which $20\log(|\frac{C_{mn}}{C_{00}}|)$ exceeds a certain threshold value, into both the standard and Siamese NN (SNN) at the receiver side (Melek and Yevick 2021; Melek and Yevick 2020a), as well as into the NNs at the transmitter side (Melek and Yevick 2020b). This approach was also applied to two-stage AI techniques in which the fiber nonlinearity was

compensated at both a transmitter stage, through a NN, and at a second receiver stage, where additionally various classifier strategies were examined (Melek and Yevick 2020b).

Since the Machine learning (ML) performance improvement depends on the characteristics of the dataset and in particular, the signal-to-noise ratio, we here examine the performance of different AI techniques at high nonlinear noise levels. This provides additional insight into the robustness of different AI techniques to nonlinear noise and aids in determining the most appropriate compensation technique for a given set of fiber properties. The results of this analysis could potentially decrease the dependence of the system performance on the doping profile which significantly affects the nonlinear coefficient in single-mode fibers (Oguama et al. 2005), which could reduce manufacturing cost and enable additional optimization of the fiber properties. As well, the predictions are further expected to be indicative of wavelength division multiplexing (WDM) system behavior as multichannel systems are considerably more sensitive to the presence of nonlinear noise.

2. TRANSMISSION SYSTEM MODEL

TABLE 1. TRANSMISSION PARAMETERS

Parameter	**Value**
Symbol Rate	$32GB$
Modulation Format	16QAM
Number of Polarization Waves	2
Pulse Shape	Root-raised cosine
Roll-off Factor	0.01

TABLE 2. LINK PARAMETERS

Parameter	**Value**
Fiber Attenuation (α)	$0.2\ dB/km$
Fiber Dispersion (D)	$17\ ps/(km.nm)$
EDFA Noise Figure (NF)	$4.5\ dB$
Simulation Wavelength (λ)	$1550\ nm$

Span Length	$80\ km$
Link Length	$3200\ km$

The single frequency system models in this paper employ two simulated 2^{17} amplitude-modulated symbol blocks generated according to the parameters in Table 1 and transmitted through the link specified in Table 2. As indicated in Fig. (2), for nonlinear pre-compensation, a negative shift corresponding to the noise predicted by a NN should be applied to the data symbols at the transmitter side. The encoded data is otherwise modulated and propagated through the fiber without additional compensation. After propagation, the signals are coherently detected and demodulated. The demodulated data either first passes through an AI stage or is inserted without compensation into the decoder. The AI stage either employs a regression or a SNN to predict and compensate the propagation noise or assigns the received symbol to one of the 16 classes using a classifier. Finally, the symbols are decoded, and the Q-factor is evaluated by comparing the received and transmitted bits according to the standard formula

$$Q = \sqrt{2} * \mathrm{erfc}^{-1}(2 * BER) \tag{3}$$

in which erfc^{-1} is the inverse of the error function, and BER is the bit error rate.

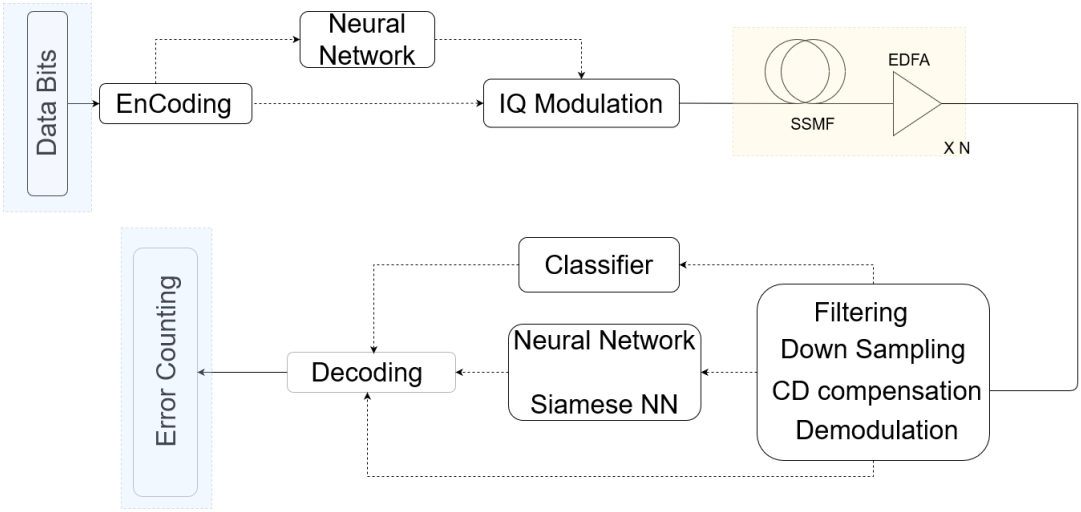

Figure 2. Block diagram for the data path and triplet calculations

3. Methodology and Numerical results
Part I: One stage AI techniques

Neural Networks (NN)

According to our parametric study in previous work (Melek and Yevick 2020a; Melek and Yevick 2020b), the optimum NN design at the receiver side of the system, as shown in Fig.(3), comprises 1 hidden layer with 2 neurons, employing "Relu" activation functions. The target function associated with this structure combines the target functions, $y = f(\sum(w_i\, x_i + b_i))$, of the individual neurons. The output of the NN target function is nearly identical to the result for the nonlinear distortion generated by Eq.(2). However, the nonlinearity of the NN hidden layer neural activation functions generates better predictions for the nonlinear noise while the data-driven stochastic optimization of the weights and biases yields improved performance compared to the perturbation-based nonlinear compensation technique without adding the additional complexity of a deep NN. Indeed, increasing the number of neurons or hidden layers increases the likelihood of overfitting and does not noticeably improve the NN performance . Here the NN inputs are the symbol of interest and the co-polarized symbols in shared time slots. The triplet terms in Eq. 2 that are input into the NN are chosen such that their C_{mn} (Tao et al. 2011) possess values of $20\log(|C_{mn}|/|C_{00}|\,)$ that exceed a threshold value $-22dB$, that yields optimal performance with minimum computational overhead. The NN is trained on ~80,000 received symbols with a given polarization at 2dB above the optimum launch power (Melek and Yevick 2020a; Melek and Yevick 2020b; da Silva et al. 2019; Zhang et al. 2019) and tested on the remaining 50,000 symbols with the identical polarization. The outputs are the Q and I values of the added nonlinear symbol distortion at the time slot of interest. To attain the minimum mean square error (MSE) between the observed and predicted output, the "Adam" first-order gradient-based optimization algorithm is applied to the stochastic objective functions (Kingma and Ba 2017). During the execution stage, the trained receiver NN post-compensates the received data, from both polarization and at all launch powers, according to the following formula,

$$Tx_{symbol} = Rx_{symbol} - \alpha(NN)_{output} \tag{4}$$

where Tx_{symbol}, and Rx_{symbol} are the predicted transmitted and received symbols, α is the power scaling factor that adjusts the predicted noise according to the transmitted signal power, and $(NN)_{output}$ signifies the output perturbations from the NN. If the transmitted symbols are instead pre-compensated at the transmitter side where the same trained NN is employed,

$$Tx_{precomp} = Tx_{symbol} - \alpha(NN)_{output} \tag{5}$$

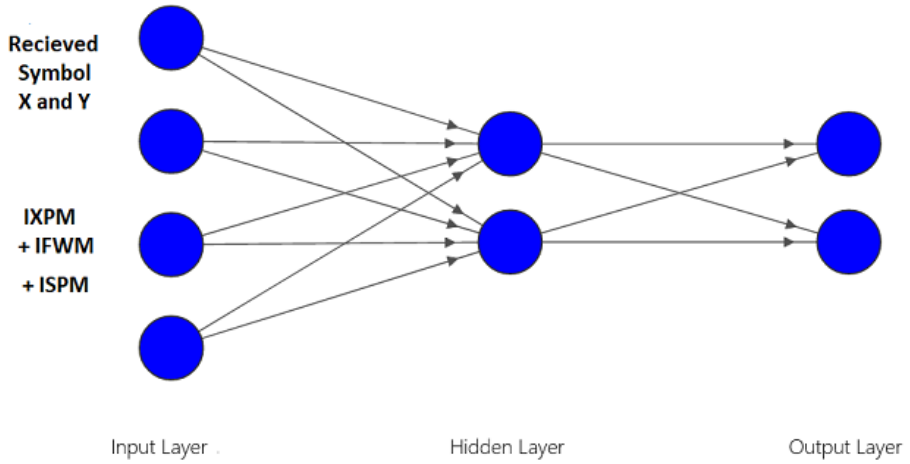

Figure3. Neural Network Structure

To determine the NN performance for different fiber nonlinearities, the nonlinear noise generated by fiber nonlinearity is quantified by the parameter γ. Here, since the optimum power level is γ-dependent the Q-factor enhancement values reported below were evaluated at the optimum system power for each γ. As shown in Fig. (4), the NN performance decreases with increasing γ for $\gamma > 2\ W^{-1}km^{-1}$ independently of the number of triplets input to the NN and hence the system complexity. Therefore, a curve fit to the average of the results for different thresholds with the same γ value can be employed to describe the NN performance. Below the R-squared values associated with the curve fit are employed to determine the most appropriate function for each technique. For a NN placed at the receiver and transmitter sides the R-squared values for $-\gamma$ are 96.7% and 92%, respectively when the results are fit to an exponential function. Indeed, since the curves of Fig.(4), are similar in shape, they can be parameterized in terms of γ and the magnitude of the threshold, θ, according to,

$$\Delta Q = \frac{\theta}{21.6}\exp(-0.03\gamma) \tag{6}$$

when a NN is placed at the receiver side and

$$\Delta Q = \frac{\theta}{21.6}\exp(-0.024\gamma) \tag{7}$$

when a NN is placed at the transmitter side. Eqs. 6 and 7 further indicate that the NN performance is effectively independent of whether the NN is located at the receiver or the transmitter. Indeed, the curves generated from the algebraic expressions nearly coincide with the data and are nearly identical in the two cases, as evident from Fig.(4a), where the values obtained from the above formula (solid lines) and the optimal fit (dotted lines) are compared with the data for 25dB and 20dB threshold values, respectively.

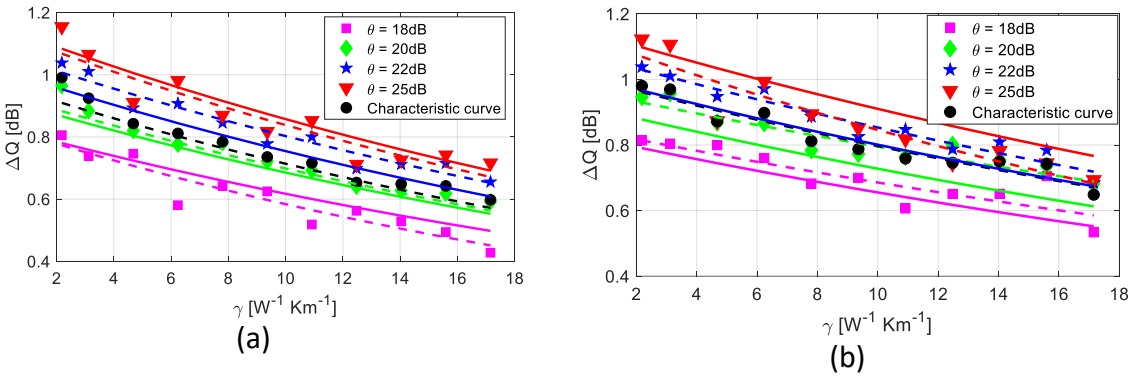

Figure 4. The performance of NN (a) at the receiver side (b) at the transmitter side versus the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.

Siamese Neural Networks (SNN)

In (Melek and Yevick 2021), the two SNN designs, shown in fig.(5) incorporating the nonlinear perturbation terms, were employed at the receiver side of the system to mitigate fiber nonlinearity. In the two branch case, the symbols of interest together with the corresponding co-polarization symbols are input into the first branch while the dominant triplets terms in Eq. (2) are processed by the second branch. An optimum SNN architecture was employed in which each of the first and second NNs possesses 1 hidden layer of 2 neurons with a "RELU" activation function while the output NN layer contains 2 neurons with "linear" activation functions according to fig. (3). In the three branch SNN, the symbols of interest are inserted into the first SNN branch while two different groups of triplets are input into the second and third branches. In the model of (Melek and Yevick 2021), each branch consisted of a two neuron hidden layer with a "Relu" activation function followed by a second two neuron output layer with a "linear" activation function. The maximum Q factor improvement was obtained when the real and imaginary parts of 748 triplets are input into the second branch while the remaining 1697 triplets are inserted into the third branch.

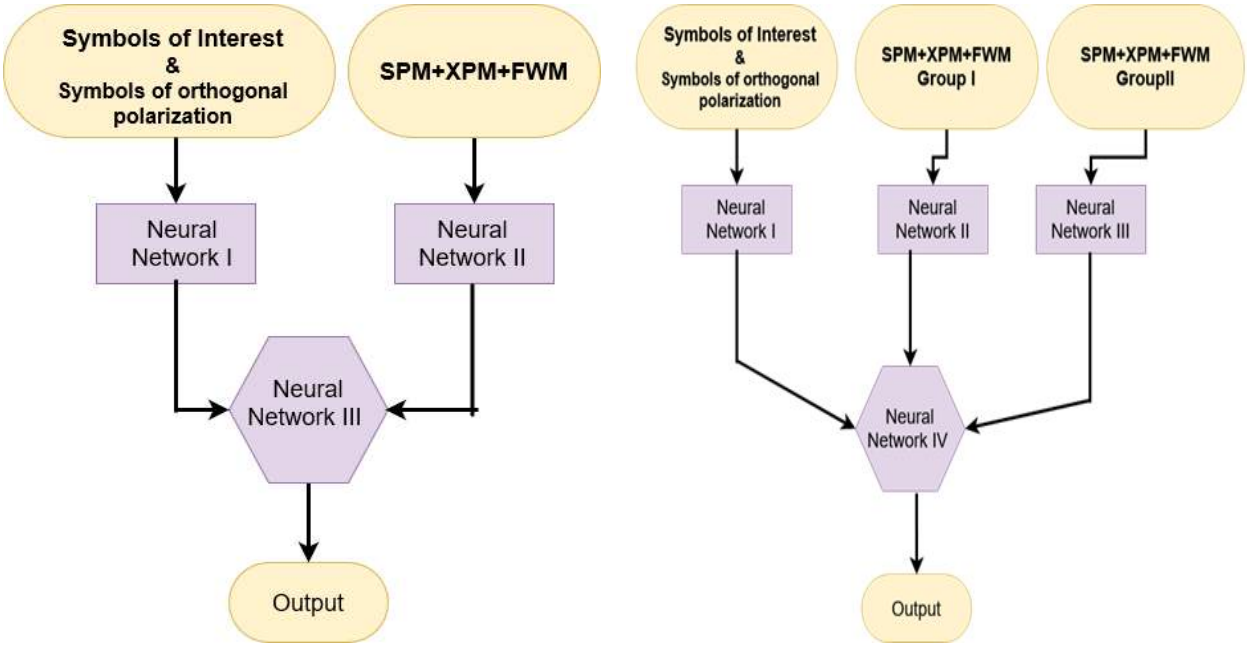

Figure 5. Siamese NN (a) two branch design (b) three branch design

The networks are trained on the 80,000 single-polarization data symbols launched at 2 dB above the optimum launch power associated with each value of γ to amplify the effect of the nonlinear distortions. To suppress overfitting, the remaining 50,000 symbols were employed for validation. The trained network was subsequently also applied to both orthogonally polarized training symbols and symbols at other signal powers. In the latter case, the output of the trained SNN was simply multiplied by a scaling factor α as illustrated in Eq.(4).

To investigate the performance of the SNN for different signal-to-noise ratios, the proposed SNN designs were also applied to systems with different values of γ, as shown in Fig.(6). While the Q-factor decreases monotonically with γ similarly to systems in which a standard NN is applied, either at the transmitter or the receiver sides, the magnitude of the Q-factor is less than that of the NN for all γ which is consistent with the results of (Melek and Yevick 2021). Indeed, for either SNN implementation, the improvement in the Q factor is approximately given by

$$\Delta \mathrm{Q} = \frac{\theta}{26} \exp(-0.026\gamma) \tag{8}$$

which is clearly smaller than the corresponding quantities in Eqs. (6) and (7) since the exponents in the equations are nearly identical. The R-squared values associated with these fits to the averaged curves are 93% and 96%, respectively.

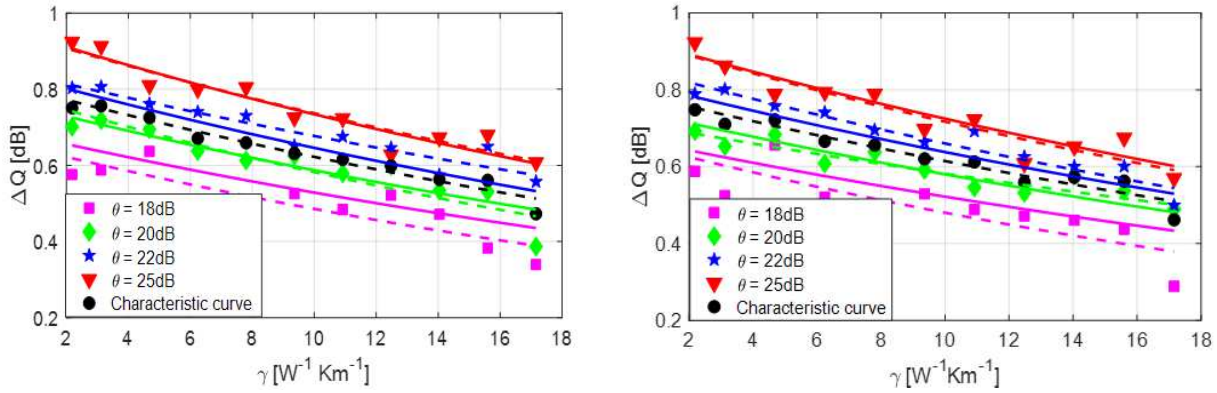

Figure 6. The Q value improvement associated with the SNN for (a) two and (b) three branches at the receiver side as a function of the nonlinearity coefficient. The dotted lines are the optimal algebraic curves while the solid lines are the characteristic equation representation.

Part II: Two stage AI techniques:

The two-stage AI technique for enhancing the Q-factor employs the trained NN design at the transmitter side while a classifier is instead located at the receiver side as indicated in Fig. (2). Similarly to (Melek and Yevick 2020b), the received data is categorized by 16 labels corresponding to the 16QAM constellation points, while the classifiers are trained to predict the appropriate label for each received symbol by employing 80,000 single-polarization symbols at the optimum power. The trained classifiers then predict the remaining 50,000 symbols with identical polarization as well as 130,000 symbols with orthogonal polarization.

A. Decision Tree

In (Melek and Yevick 2020b), a system with $\gamma = 1.4\ W^{-1}km^{-1}$ achieved a smaller Q-factor enhancement when a decision tree was employed as a classifier at the receiver compared to an optimized system with a NN placed at the transmitter. In this reference, which employed the scikit-learn package the trees did not have a specified maximum depth (the length of the longest path from a root to a leaf) since the evolution of the tree was governed by the input data. Fig.(7) demonstrates that such a decision tree in fact exhibits only a slightly improved performance relative to the NN implantation as γ increases. Accordingly, decision trees are relatively ineffective classifiers for nonlinearity mitigation. The continuous curve shown in Fig.(7) corresponds to the formula,

$$\Delta Q = \frac{\sqrt{\theta}}{4.4}\gamma^{-0.14} \tag{9}$$

indicating that the Q-factor enhancement is described by a negative power of γ with a 90% R-squared value.

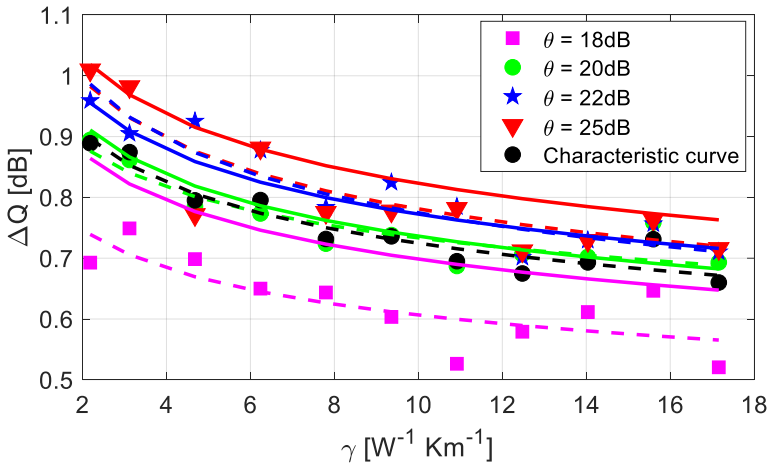

Figure 7. Two-stage AI technique performance with a NN at the transmitter and a decision tree at the receiver as a function of the nonlinearity coefficient. The dotted lines represent the optimal fit while the solid lines are generated from an algebraic formula.

B. AdaBoosting and GBoosting

To improve the decision tree results, the classifier performance can be enhanced through boosting, as demonstrated for Adaptive boosting (AdaBoosting) in (Melek and Yevick 2020b). The AdaBoosting algorithm continuously updates the probability distribution of the input variables of an initially weak and inadequately performing classifier by multiplying the weight of each input variable by a weight updating parameter, resulting in the next classifier iteration. The parameter value decreases when the input is correctly identified by the previous classifier. As shown in Fig. (8a), a two-stage AI technique with Adaboosting, applied to the decision tree classifier with a maximum depth of 8 at the receiver, yields improved performance relative to an uncompensated system for $\gamma < 9\ W^{-1}km^{-1}$. However, the probability of misclassified data increases with γ. This negatively affects the performance since the accuracy of the weight updating parameter depends on the ratio of incorrect to correct classifications, therefore it degrades rapidly as the noise increases (Shrestha and Solomatine 2006).

Strong Gboosting classifiers, which minimize the classification error by combining several weak classifiers have proved effective in compensating high nonlinear noise levels (Son et al. 2015). As shown previously (Melek and Yevick 2020b), in the context of this paper Gboosting is optimally applied to decision trees with a depth of 3. Indeed, Fig. (8b) demonstrates that the system performance enhancement decreases with γ, although for small γ the performance of Gboosting is less than that of Adaboosting for which the Q-factor enhancement can be approximated by

$$\Delta Q = 2.6e^{-0.089\gamma} - 1 \qquad (10)$$

with a fitting parameter of $R^2 = 0.973$ which is almost independent of the threshold level. The corresponding Gboosting Q-factor enhancement in Fig. (8b) is approximated with $R^2 = 0.95$, as,

$$\Delta Q = \frac{\sqrt{\theta}}{3.3}\gamma^{-0.171} \tag{11}$$

Thus Ada-boosting is most advantageous for low fiber nonlinearity and limited computational resources as the number of triplets can be considerably reduced compared to G-boosting. Two-stage G-boosting was further found to be preferable to an isolated NN over the range of γ examined.

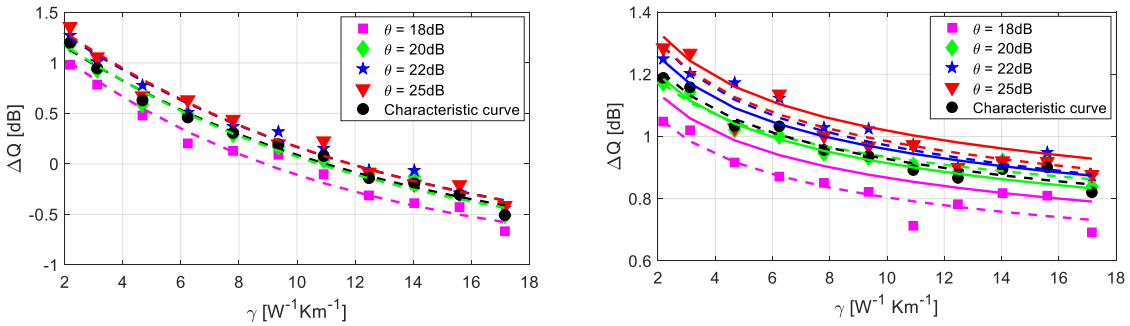

Figure 8. Two-stage AI performance with a NN at the transmitter and (a) Ada-boosting and (b) G-boosting at the receiver as a function of the nonlinearity coefficient. The dotted lines are the algebraic approximation for the results.

C. Random Forest and Extra trees

The random forest ensemble method is based on the voting average for each class of a group of decision tree classifiers running in parallel where each tree of the forest independently samples random vectors containing identically distributed random numbers. In contrast to the extra trees method in which each decision tree in the forest is constructed from the original training sample, which consists of a set of features (dimensions). In the extra trees procedure, each tree is then given a random sample at each test node, from which the decision tree chooses the best feature to classify the data depending on specific mathematical criteria (Geurts et al. 2006).

Although the decision tree technique yields only a limited improvement in the Q-factor relative to the results of the previous section, Fig.(9) demonstrates that the tree ensembles associated with the random forest and extra trees techniques enable significant further improvements. Moreover, Fig.(9) shows that the random forest and extra trees techniques compensate for nonlinear noise more effectively than competing algorithms, especially at high levels of

nonlinearity while the computational time is typically limited as indicated in (Melek and Yevick 2020b). As shown in the previous procedures, to model the dependence of the Q-factor enhancement in the random forest and extra trees methods on γ, an average is performed over of all the results for different thresholds with the same γ value. Unlike the previous methods, the slope of the curves in the random forest and extra trees procedures is dependent on the threshold value as given in the following function,

$$\Delta Q = \frac{\sqrt{\theta}}{3.8}\gamma^{\frac{2.94}{\theta}} \tag{12}$$

Fig.(9) compares Eq. (12) (solid lines) to the optimum fit to the results (dotted lines). The agreement is especially pronounced for the 22dB and 20dB threshold curves.

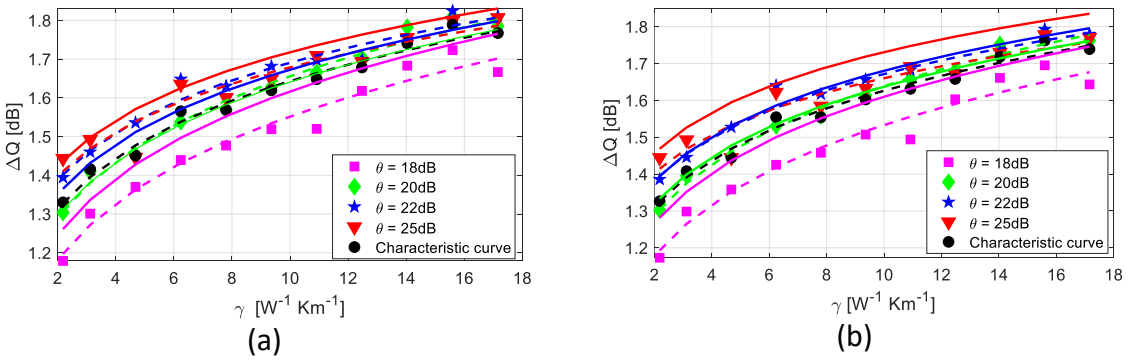

Figure 9. The Q-factor improvement as a function of γ for a two-stage AI technique with a NN at the transmitter side and (a) a random forest and (b) extra trees at the receiver side. The dotted lines represent the best fit while the solid lines correspond to the algebraic approximation.

D. Multi-layer perceptron (MLP) classifier

Employing a MLP classifier containing a single 4 neuron hidden layer with a 'Relu' activation function at the receiver, trained as indicated at the beginning of this section, yields the curves in Fig.(10). While this architecture is near-optimum, the Q-factor improvement is effectively identical to that of a standard NN placed at the receiver end. Further, the equation

$$\Delta Q = \frac{\theta}{21.6}\exp(-0.019\gamma) \tag{13}$$

that parameterizes the curves is almost identical to that associated with a transmitter side NN. This is identical to the result in (Melek and Yevick 2020b), which predicted a $0.03dB$ enhancement for $\gamma = 1.4\, W^{-1}km^{-1}$. Indeed, from Fig. (11), which displays the averaged

results for the Q-factor associated with the AI configurations analyzed in this paper, it is evident that all techniques perform nearly identically for small γ but differ increasingly for larger values of γ. This figure further establishes that the most appropriate AI technique for the system under consideration is a two-stage structure with either a random forest or extra trees at the receiver side.

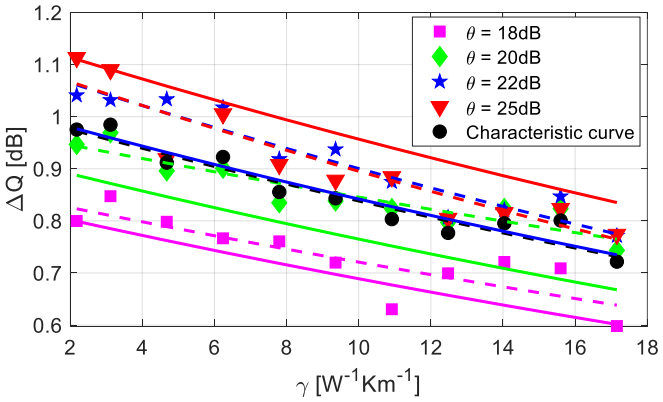

Figure 10. The performance for a transmitter side NN and a receiver side MLP as a function of the nonlinearity coefficient. The dotted lines are the best fit to the data while the solid lines are generated with Eq.(13).

Finally, Fig.(11) compares the system Q-factor for the different AI methods with those of the analytical perturbation-based nonlinear compensator (PB NLC) procedure, shown in Eq. (2), for a fixed number (2445) of triplets and a -22dB threshold value. Evidently, for a fixed number of triplet terms the accuracy of the PB NLC method falls rapidly as the nonlinear coefficient increases for $\gamma > 2W^{-1}\,km^{-1}$ such that the Q-factor values even becomes negative for $\gamma > 5.8W^{-1}\,km^{-1}$. In effect, if the number of triplets in the PB NLC calculations is insufficient to model the nonlinear noise, the transmitted symbols are incorrectly interpreted. Increasing the number of triplets, however, increases the complexity of the model and the required computational resources. In contrast, the AI techniques, which provide a significant system Q-factor for values of γ as large as $17\,W^{-1}\,km^{-1}$ can adapt to large system nonlinearities, presumably as a result of the inherent nonlinearity of the neuron activation functions.

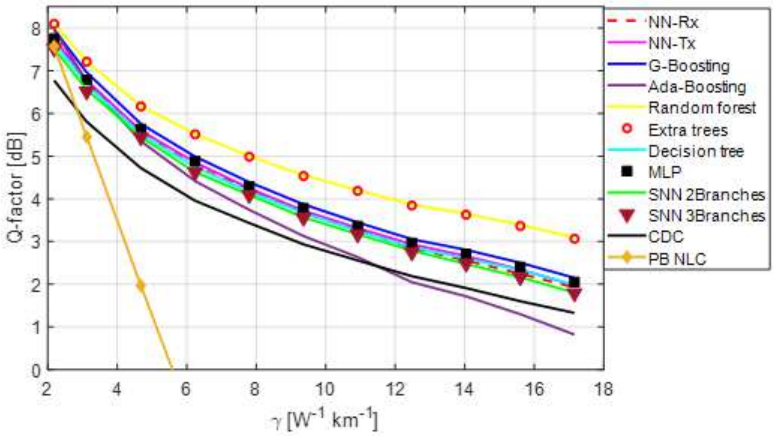

Figure 11. The system Q-factor improvement for different AI techniques and perturbation-based nonlinear compensation (PB NLC) method.

Conclusion:

To benchmark the performance of different AI techniques in optical communication systems affected by high levels of fiber nonlinearity, we investigated nonlinear compensation in a single frequency, $3200km$, double polarization, and 16-QAM optical system. The results demonstrate that a NN can be employed either at the receiver or the transmitter side over a wide range of nonlinear noise levels with identical Q-factor enhancements. Employing instead a SNN at the receiver side leads to slightly reduced performance. On the other hand, two-stage AI with classifiers such as extra trees and random forests at the receiver can significantly compensate for high levels of nonlinear noise, while decision trees do not afford any noticeable advantage over the standard NN procedure. Additionally, Ada-boosting improves the performance for small nonlinear coefficients, even if a reduced number of triplets are used as input into the transmitter NN, but its effectiveness decreases rapidly with nonlinearity. For each topology, an empirical algebraic equation was generated for the system Q-factor enhancement in terms of the triplet selection threshold and γ. The enhancement associated with each technique at any nonlinear noise value can therefore be rapidly estimated, which should be useful in communication system design. Finally, while the AI designs examined in this paper will be applied to WDM systems with differing γ values and number of channels in a subsequent paper, the results are expected to be largely identical to the single frequency channel case.

Acknowledgments

The Natural Sciences and Engineering Research Council of Canada (NSERC) is acknowledged for financial support.

References:

Agrawal, G.P.: Nonlinear fiber optics. Elsevier / Academic Press, Amsterdam ; Boston (2007)

van den Borne, D., Fludger, C.R.S., Duthel, T., Wuth, T., Schmidt, E.D., Schulien, C., Gottwald, E., Khoe, G.D., de Waardt, H.: Carrier phase estimation for coherent equalization of 43-Gb/s POLMUX- NRZ-DQPSK transmission with 10.7-Gb/s NRZ neighbours. In: 33rd European Conference and Exhibition on Optical Communication - ECOC 2007. pp. 723–723. IEE, Berlin, Germany (2007)
Catanese, C., Triki, A., Pincemin, E., Jaouen, Y.: A Survey of Neural Network Applications in Fiber Nonlinearity Mitigation. In: 2019 21st International Conference on Transparent Optical Networks (ICTON). pp. 1–4. IEEE, Angers, France (2019)
Fisher, R.A., Suydam, B.R., Yevick, D.: Optical phase conjugation for time-domain undoing of dispersive self-phase-modulation effects. Opt. Lett. 8, 611 (1983). https://doi.org/10.1364/OL.8.000611
Gao, Y., Cartledge, J.C., Downie, J.D., Hurley, J.E., Pikula, D., Yam, S.S.-H.: Nonlinearity Compensation of 224 Gb/s Dual-Polarization 16-QAM Transmission Over 2700 km. IEEE Photonics Technol. Lett. 25, 14–17 (2013). https://doi.org/10.1109/LPT.2012.2227111
Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. 63, 3–42 (2006). https://doi.org/10.1007/s10994-006-6226-1
Ip, E., Kahn, J.M.: Compensation of Dispersion and Nonlinear Impairments Using Digital Backpropagation. J. Light. Technol. 26, 3416–3425 (2008). https://doi.org/10.1109/JLT.2008.927791
Kamalov, V., Jovanovski, L., Vusirikala, V., Zhang, S., Yaman, F., Nakamura, K., Inoue, T., Mateo, E., Inada, Y.: Evolution from 8QAM live traffic to PS 64-QAM with Neural-Network Based Nonlinearity Compensation on 11000 km Open Subsea Cable. In: Optical Fiber Communication Conference Postdeadline Papers. p. Th4D.5. OSA, San Diego, California (2018)
Kingma, D.P., Ba, J.: Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs. (2017)
Liga, G., Xu, T., Alvarado, A., Killey, R.I., Bayvel, P.: On the performance of multichannel digital backpropagation in high-capacity long-haul optical transmission. Opt. Express. 22, 30053 (2014). https://doi.org/10.1364/OE.22.030053
Lin, C.-Y., Asif, R., Holtmannspoetter, M., Schmauss, B.: Nonlinear mitigation using carrier phase estimation and digital backward propagation in coherent QAM transmission. Opt. Express. 20, B405 (2012). https://doi.org/10.1364/OE.20.00B405
Mata, J., de Miguel, I., n, R. n J.D., Merayo, N., Singh, S.K., Jukan, A., Chamania, M.: Artificial Intelligence (AI) Methods in Optical Networks: A Comprehensive Survey. Opt. Switch. Netw. 28, 43–57 (2018). https://doi.org/10.1016/j.osn.2017.12.006
Melek, M.M., Yevick, D.: Nonlinearity mitigation with a perturbation based neural network receiver. Opt. Quantum Electron. 52, 450 (2020)(a). https://doi.org/10.1007/s11082-020-02565-5
Melek, M.M., Yevick, D.: Machine learning two stage optical fiber nonlinearity mitigation. J. Mod. Opt. 67, 1072–1077 (2020)(b). https://doi.org/10.1080/09500340.2020.1810347
Melek, M.M., Yevick, D.: Fiber nonlinearity mitigation with a perturbation based Siamese neural network receiver. Opt. Fiber Technol. 66, 102641 (2021). https://doi.org/10.1016/j.yofte.2021.102641
Oguama, F.A., Johnson, A.M., Reed, W.A.: Measurement of the nonlinear coefficient of telecommunication fibers as a function of Er, Al, and Ge doping profiles by using the photorefractive beam-coupling technique. J. Opt. Soc. Am. B. 22, 1600 (2005). https://doi.org/10.1364/JOSAB.22.001600
Shrestha, D.L., Solomatine, D.P.: Experiments with AdaBoost.RT, an Improved Boosting Scheme for Regression. Neural Comput. 18, 1678–1710 (2006). https://doi.org/10.1162/neco.2006.18.7.1678
Sidelnikov, O., Redyuk, A., Sygletos, S.: Equalization performance and complexity analysis of dynamic deep neural networks in long haul transmission systems. Opt. Express. 26, 32765 (2018). https://doi.org/10.1364/OE.26.032765
da Silva, E.P., Yankov, M.P., Da Ros, F., Morioka, T., Oxenløwe, L.K.: Perturbation-based FEC-assisted Iterative Nonlinearity Compensation for WDM Systems. J. Light. Technol. 37, 875–881 (2019). https://doi.org/10.1109/JLT.2018.2882638

Son, J., Jung, I., Park, K., Han, B.: Tracking-by-Segmentation with Online Gradient Boosting Decision Tree. In: 2015 IEEE International Conference on Computer Vision (ICCV). pp. 3056–3064. IEEE, Santiago, Chile (2015)
Tao, Z., Dou, L., Yan, W., Li, L., Hoshida, T., Rasmussen, J.C.: Multiplier-Free Intrachannel Nonlinearity Compensating Algorithm Operating at Symbol Rate. J. Light. Technol. 29, 2570–2576 (2011). https://doi.org/10.1109/JLT.2011.2160933
Tkach, R.W.: Scaling optical communications for the next decade and beyond. Bell Labs Tech. J. 14, 3–9 (2010). https://doi.org/10.1002/bltj.20400
Wang, D., Zhang, M., Cai, Z., Cui, Y., Li, Z., Han, H., Fu, M., Luo, B.: Combatting nonlinear phase noise in coherent optical systems with an optimized decision processor based on machine learning. Opt. Commun. 369, 199–208 (2016)(a). https://doi.org/10.1016/j.optcom.2016.02.029
Wang, D., Zhang, M., Fu, M., Cai, Z., Li, Z., Han, H., Cui, Y., Luo, B.: Nonlinearity Mitigation Using a Machine Learning Detector Based on k -Nearest Neighbors. IEEE Photonics Technol. Lett. 28, 2102–2105 (2016)(b). https://doi.org/10.1109/LPT.2016.2555857
Zhang, S., Yaman, F., Nakamura, K., Inoue, T., Kamalov, V., Jovanovski, L., Vusirikala, V., Mateo, E., Inada, Y., Wang, T.: Field and lab experimental demonstration of nonlinear impairment compensation using neural networks. Nat. Commun. 10, 3033 (2019). https://doi.org/10.1038/s41467-019-10911-9

Statements & Declarations

- *“The authors have no relevant financial or non-financial interests to disclose.”*
- *“Marina Melek is responsible for the design of the work; the collecting, analysis, and interpretation of data. David Yevick is responsible for supervision and revision. “*
- *This work is part of a PhD thesis submitted to the University of Waterloo. That pre-print can be found at https://uwspace.uwaterloo.ca/bitstream/handle/10012/17780/Melek_Marina.pdf*