A Simplified Approach to Estimate Bounds on Fading Channel Unavailability

Anjana Jain P. D. Vyavahare L. D. Arya*

***Abstract*—A simplified approach is presented to evaluate fading channel unavailability. A fixed degree of certainty can be established and bounds on the same can be estimated. The slow fading channel is considered and is modeled as a discrete state Markov process. For such fading channel average value of outage time can be computed either using simulation or with the help of measured data set. But there is no indication of certainty that can be placed in its value. The paper presents an approach to estimate the range of outage rate and outage probability. Finally, bounds on channel unavailability are presented with the variation in outage time and number of outages. Proposed analysis can also be applied to fast fading channels when outage occurs frequently. Results may be used to provide higher order protocol performance evaluation and optimization. Examples are given to illustrate the results derived in the paper and to justify the approach presented.**

***Index Terms*—Channel unavailability, Fading channel, Markov Model, Outage time, Transition rate.**

I. INTRODUCTION

Literature shows that Finite State Markov Models are appropriate to represent fading channel behavior [1], [2]. Each state of the model corresponds to a specific channel quality given by constant bit error probability [3], [4], [5], and [6]. The received signal envelope experiences deep and shallow fades, occasionally and frequently respectively. Hence the channel can be considered as slow or fast fading channel. A quantitative evaluation of this phenomena is the level crossing rate defined as the rate at which fading envelope crosses level R in positive (or negative) going direction.

The average outage time is defined as the average time for which the received envelope remains below a specified level after crossing that threshold level [7], [8].

When Channel suffers with outage it is unavailable to user and said to be in bad state. When channel is above threshold it will be available to user and will be in good state. Extensive research has been devoted to model the fading channel as first order Markov process. However the issue of channel availability while accounting for the outage times need to be resolved. The essence of the paper is to estimate channel availability and upper and lower bounds on channel unavailability. Practically a situation may occur in fading environment when outage starts and channel becomes unavailable. It leads to starting the precautionary measures such as change in system parameters e.g. transmission reception policies, coding, interleaving, and diversity and equalization algorithms. The need of these alterations can be avoided, if the range in which the outage probability and outage rate falls can be estimated with certain confidence level. This can be represented in terms of channel availability and its bounds. Present paper estimates channel availability while accounting for the average outage time. Comparison of channel availability is then demonstrated with the variation in transition rates using numerical examples. Finally bounds on channel unavailability are computed in terms of upper and lower limits. Two state Markov model is considered with exponentially distributed intervals for both good time and bad state. Channel is assumed as a slow fading channel [1], [2]. Rest of the paper is organized as follows.

Section 2 presents the Markov Model of the fading channel. Section 3 analyzes the channel availability and bounds on channel unavailability. Section 4 describes the establishment of outage times required for analysis purpose. Numerical examples to justify the approaches are illustrated in Section 5 and results are shown in the form of plots. Finally, paper is concluded in Section 6.

II. THE CHANNEL MODEL

The fading channel is modeled as a first order two state Markov model with certain assumptions [9], [10], [13], [14]. At any time, the channel has been in the one of the two possible states. Whenever the received signal crosses the threshold, outage is considered and channel is assumed to be in bad state. Duration of good times has negative exponential distribution with parameter λ and bad time has negative exponential distribution with parameter m [11], [12].

The state space is defined for the process is
s = {0, 1}

The process follows a continuous time Markov process which starts from state 1. Stochastic process is defined as

$$\{X(t); t \geq 0\} \text{ Where}$$

$$X(t) = \begin{cases} 1, \ \textit{TheC hannel is} \text{ in } \textit{good State} \\ 0, \ \textit{TheChannel is} \text{ in } \textit{bad State} \end{cases}$$

Numbers of consecutive outage and success times are defined as a sample size. When sample size is small means the channel is slow, single point estimation of outage time can be established. For the fast fading channel sample size can be increased when more outages occur.

V (t) denotes instantaneous availability of the channel for the calculated transition rates [1, 2].

For the first, order two state Markov process

Department of Electronics and Telecommunication, Department of Electrical Engineering*
23 Park Road, Shri G.S. Institute of Technology and Science, M.P., India

$$V(t) = \frac{m}{l+m} + \frac{l}{l+m} e^{-(l+m)t} \qquad (1)$$

λ and μ are the Birth Rate and Departure Rate respectively.

III. ESTIMATION OF CHANNEL AVAILABILITY AND BOUNDS ON CHANNEL UNAVAILABILITY

Accuracy inherent in the calculation process must be verified for the to quantitative evaluation of the model developed and analysis presented. This is related with the validity of the data used. The results obtained using a valid model can be precise but may be highly inaccurate if the data is incorrect for the given purpose. The range in which estimated results fall within a fixed degree of certainty is required to establish. Limits and the bounds, are therefore, appreciated for the same. The approach used for developing the bounds as upper and lower limits of unavailability can be applied to the channel, modeled as a Markov process [11]. For the mentioned process the channel unavailability is defined as

$$\mathrm{U} = \frac{l}{l+m} \qquad (2)$$

The average outage time and satisfactory time are evaluated from the recorded data or using simulation results. An estimate of U at a given confidence level can be evaluated as

$$P[C_{a,b} \le F_{2a,2b} \le \bar{C}_{a,b}] = P\left[\frac{t_d}{t_d + \bar{C}T_s} \le \frac{l}{l+m} \le \frac{t_d}{t_d + CT_s}\right] \qquad (3)$$

Where a, b are number of outage times and satisfactory times. For the proposed analysis a and b are taken as constant [11].

Ts are the satisfactory time. t_d is the outage time.

C and other constant depending up on confidence level with a specific probability Y. These constants are derived from the help of the F Distribution table.

F= Snedecor's F statistics with 2a and 2b degree of freedom in numerator and denominator.

$$P\,[F \ge \bar{C}] = \frac{1-Y}{2} \quad \text{OR} \quad P[F \le C] = \frac{1-Y}{2} \qquad (4)$$

$$\text{Upper limit } U_u = \frac{t_d}{t_d + \bar{C}T_s} \qquad (5)$$

$$\text{Lower limit } U_L = \frac{t_d}{t_d + CT_s} \qquad (6)$$

For the less and frequent outages the limits on channel Unavailability's are computed. The results can be applied to slow and fast fading channel characterization [11].

IV. OUTAGE TIME ESTABLISHMENT

In the paper Rayleigh fading channel is simulated with Doppler shift of 100Hz, Path Delay of 1 µs, Sampling period of 10000 samples/s, and Binary DPSK with carrier Frequency of 900MHz. Based on the simulation results Birth Rate, Departure Rates and Outage time are computed [9], [10], [11], [16]. Channel availability is computed for the given outage times using (2). Variation in Birth Rate and Departure Rates are also considered for estimation of channel availability.

Parameter configuration is as follows

Birth Rate – 21.47 transition per unit time.

Departure Rate – 18.87 transition per unit time.

Outage time – **0.0078, 0.0038 and 0.0019 s.**

Fade Depth – 10 dB.

Fading channel suffers with less and frequent outages. Number of outages considered in the paper varies from 10 to 50.

V. NUMERICAL EXAMPLES AND RESULTS

Initially based on the simulation results following datas are used

Outage Time = 0.0078, 0.0038 and 0.0019

Channel availability with the different departure rates are plotted in the Fig. 1. It has been shown that availability of the channel is increasing function of departure rate. An increase in the departure rate from 26.02 to 128.2 channel availability increases with 25 to 30 %. Further channel unavailability's are computed with 0.90 confidence level and with variation in number of outages. For the computation purpose following parameters are considered

Number of outages = 10, 20 and 50.

Outage time = 0.0038s.

Success Time = 0.038s.

Results are demonstrated in Fig. 2. It shows the variation in Upper and lower limits on Channel Unavailability's with variation in Number of Outage times. It can be concluded that the unavailability lies between 0.0456 to 0.18 for slow fading channel with 10 outages. For 50 outages the channel unavailability is in the range of 0.0679 to 0.123. Fig. 3 shows that if confidence level is increased from 90 to 98%, lower limits changes from 0.0679 to 0.035 and upper limits changes from 0.123 to 0.25. Upper limit indicates that the true unavailability is less than this value. While lower limit shows that true unavailability is more than this value.

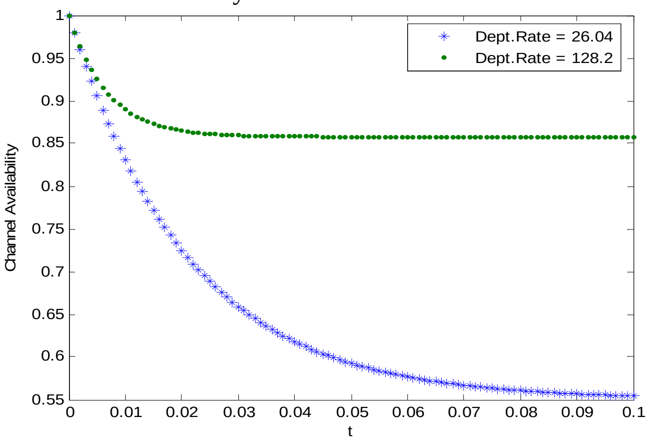


Fig. 1 - Variation of instantaneous channel availability with variation in Departure Rate.

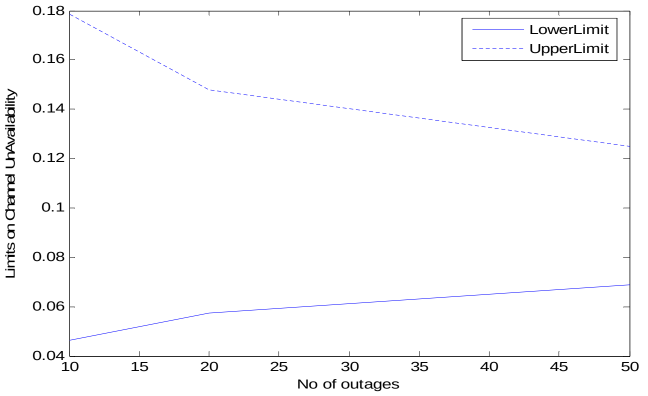


Fig. 2 - Upper and Lower limits on Channel Unavailability with variation in Number of Outage times.

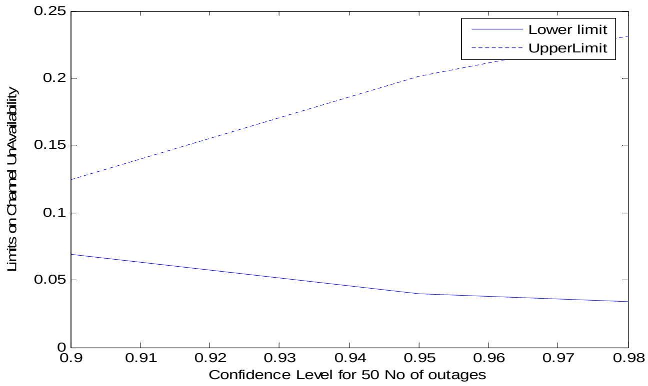


Fig. 3 – Limits on Channel Unavailability with variation in confidence level from 90% to 98 %

Average Fade duration $'t'$ can be considered as a non-negative random variable with possible distribution as exponential, Beta or Weibull. Simulation tool MATLAB is used in deriving results.

VI. CONCLUSION

In the present paper, the fading channel is modeled as first order two state Markov process. Outage times are computed and instantaneous and steady state values of channel availability are estimated. Upper and lower limits on channel Unavailability with variation in number of outage times are then shown. It is observed that range of unavailability limit decreases with number of outages. It is also demonstrated that Unavailability can be further estimated with more certainty. With increase in confidence level channel becomes narrower in term of its availability. The focus of the proposed paper is to develop methodology and algorithms for estimation on channel unavailability and bounds on it. It is also suggested that if certain short outage times are tolerated results in improved availability. The methodology will result in efficient characterization of fading channel to alleviate the multipath effects and to provide higher order protocol performance evaluation and optimization. Further research work may study availability with tolerable outage times and partial availability during nontolerable outage times. Markov model of higher state and higher order can also be built.

REFERENCES

[1] Z.Zheng, L.Cuil, A.G. Hawkes, “A study on single unit Markov repairable system with repair time omission”, IEEE transaction on Reliability, vol.55, no.2, June 2006, pp. 182-188.
[2] R. Perzocon, D. Montoro Cozorla, “Transient analysis of a repairable system using phase type distribution and geometric processes” vol. 53, June 2004, pp. 185-192.
[3] B. Skalar, Digital Communication, Fundamentals and Applications, Second Edition, Pearson Education, 2001.
[4] W.H. Trenter, S. Shanmugan, T.S.Rapaport, K.L.Kosbar, Principles of Communication System Simulation with Wireless Applications, First Edition, Pearson Education, 2004.
[5] E. Gilbert, “Capacity of a burst-noise channel”, Bell System. Tech. Journal, no. 39, September 1960, pp. 1253-1266.
[6] E. O. Elliott, “Estimates of error rates for
[7] Codes on burst-error channels”, Bell System. Tech. Journal 42, September 1963, pp. 1977 1997.T. S. Rappport, “Wireless Communication”, Pearson education, Second Edition 2002.
[8] N.C. Beaulieu, X. Dong, “Level Crossing Rate & Average Fade Duration of MRC (maximum ratio combiner) and EGC (Equal gain combiner) Diversity in Rican Fading”, IEEE Transaction on Communication, vol. 51, no. 5, May 2003, pp. 722-726.
[9] S. M. Ross, Introduction to Probability Models, Elsevier Publication, And Ninth Edition 2007.
[10] K. S. Trivedi, Probability and statistics with reliability, queuing and computer applications” PHI publication, Third Edition, 1992.
[11] Roy Billinton, R.N. Allan, Reliability Evaluation of Engineering Systems, concepts and techniques, Springer, Second Edition 1992.
[12] A. M. Law, Simulation, Modeling and Analysis, TataMcGrawHill, Fourth Edition, 2008.
[13] H.S. Wang, N. Moayeri, “Finite State Markov Channel- a Useful Model for Radio Communication Channels”, IEEE Transaction of Vehicular technology, vol. 44, no. 1, Feb. 1995, pp. 163-171.
[14] F. Babich, O.E. Kelly, G. Lombardi, “Generalized Markov Modeling for Flat Fading”, IEEE Transaction on Communication, vol. 48, no, April 2000, pp. 547-555.
[15] S. Day, J. Evans, “Outage Capacity and Optimal Power Allocation for Multiple Time Scale Parallel Fading”, IEEE Transaction on Wireless Communication Systems, vol. 6, July 2007, pp. 2369-2373.
[16] A. Jain, R. Upadhyay, P.D. Vyavahare, L.D. Arya, “Stochastic Modeling and Performance Evaluation of Fading Channel for Wireless Network Design”, IEEE Proc. AINA 07, PAEWN, pp 893-897.

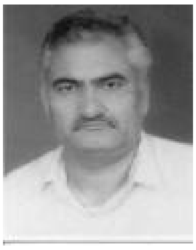
Mrs. Anjana Jain born on 2nd Feb 1967 in India. She received B.E degree (1989) in electronics and telecommunication from M.A.C.T., India and M.E degree (1996) in electrical engineering from Shri G.S. Institute of Technology and Science India in 1996. Her major field of study is electronics and communication.

She has teaching experience of 18 years and currently Reader in Department of Electronics and Telecommunication in Shri G.S. Institute of Technology and Science, India. She has published papers in National and International Journal, Conferences and Workshops organised by IEEE, IETE, ISTE and IE. Her reaserch interest are Charecterisation and Parameter Estimation of Mobile Radio Channel, Reliability and Availability computation of Fading Channel and Stochastic Modeling.

Mrs.Jain is member of IEEE, TWAWS, IACSIT, IETE, ISTE, IE (I). She Got Nokia Award for her paper presented in AINA, PAEWN 2007.

Dr. P.D. Vyavahare born on 6 th Oct. 1953 in India. He obtained M.Tech (1976) and Ph.D (1994) degrees from IIT Bombay. His major field of study is electronics and communication. He is presently Professor and Head of the Department of Electronics and Tele communication in Shri G.S. Institute of Technology and Science, India. He worked as Communication Engineer in TATA Institute of Fundamental Reasearch Bombay during1976 -1982. He has more than 40 articles to his credit, out of which more than 15 are in National and International Journal. His reaserch interests are Cherecterisation of Mobile Radio Channel, Computer Networking and Channel Coding Techniques.

Dr. Vyavahare is member of IEEE, CSI, ISTE and Fellow member of IETE and IE (I). He was Hindu Hitachi scholar during 1981-82 and an Associate member of ICTP – UNESCO, Italy since 1998.



Dr. L.D. Arya born on 18 th Jan 1951 in India. He obtained M.Tech (1974) from BITS Pilani and Ph.D (1991) from IIT Delhi. His major field of study is electrical engineering.

He is presently Professor in the Department of Electrical Engineering in Shri G.S. Institute of Technology and Science, India. His areas of interest include Power Systems, Reliability Engineering and Control Engineering. He has published more than 150 research papers in different International and National journals and Conferences.

Dr. L.D.Arya is the Fellow member of IE (I). He has to his credit 5 national awards including Pt. Madan Mohan Malviya Memorial Medal for year 2001-02, Power Medal of Union Ministry of Energy in 1995-96 etc.

International Journal of Computer and Electrical Engineering, Vol. 1, No. 2, June 2009
1793-8163

Neural Network Based Handwritten Digits Recognition- An Experiment and Analysis

M.J. Islam, Q. M.J. Wu, M. Ahmadi, M.A. Sid-Ahmed

Abstract— Handwritten digit recognition has become very useful in endeavors of human/computer interaction. Reliable, fast, and flexible recognition methodologies have elevated the utility. This paper presents an experiment and analysis of the Neural Network classifier to recognize handwritten digits based on a standard database. The experimental setup implemented in Matlab determines the ability of a Multi-Layer Neural Network to identify handwritten digit samples 5-9. This network is the representative for recognition of remaining digits 0-4. We consider not only accurate recognition rate, but also training time, recognition time as well as the complexity of the networks. The Multi-Layer Perceptron Network (MLPN) was trained by back propagation algorithm. Network structures vary with the hidden units, learning rates, the number of iterations that seem necessary for the network to converge. Different network structures and their corresponding recognition rates are compared in this paper to find the optimal parameters of the Neural Network for this application. Using the optimal parameters, the network performs with an overall recognition rate 94%.

***IndexTerms*— Handwritten Digits, Multi-Layer-Perceptron Neural Network, Network Architecture.**

I. INTRODUCTION

Speech is a sign system that is more natural than writing to humans. However, writing is considered to have made possible much of culture and civilization. The written form of language is contained in printed documents, such as newspapers, magazines and books, and in handwritten matter, such as found in notebooks and personal letters. Given the importance of written language in human transactions, its automatic recognition has practical significance. Great strides have been achieved in pattern recognition specifically in the area of handwritten digit recognition in recent years. This rapid progress has resulted from a combination of number of developments including the proliferation of powerful, inexpensive computers, the invention of new algorithms that take advantage of these computers, and the ability of large database of characters that can be used for training and testing [1]. Though not a significant enough solution in the area of pattern recognition, one realm in which we think our system would be useful is mail sorting such as automatic zip code reading on letters, which is currently in use by the U.S. Postal Service [2].

The practical application of handwritten digit recognition and our interest motivated us to develop a neural network capable of identifying the digits 5-9. Our main intention is to implement a MLPN and to test the level of accuracy we could achieve with a single hidden layer of neurons. We assumed that back propagation training and a single hidden layer structure would yield high accuracy once we were able to provide the network with inputs containing the maximum amount of information about the digits 5-9. Since the success of a neural network depends on the different parameters like hidden units, learning rates, the number of iterations that seem necessary for the network to converge, optimizing the network structure was the main focus of this work.

The remainder of this paper is organized as follows: Section II explains about the experimental handwritten data set, section III discusses about network architecture and training mechanism. Matlab simulation and performance of the results are discussed in section IV and V respectively and finally, section VI concludes the paper.

II. EXPERIMENTAL DATASET

The handwritten data set was obtained from the university of Wisconsin- Madison central repository [3]. The data file, called digits5-9.dat, is an ASCII file containing 100 examples, one per line. Each example is a comma-separated (without any white space) list of 65 integer values, the first 64 specifying the input and the last value specifying the digit which is the desired output. The input values are integers in the range [0..16]. Some examples of handwritten digits are shown in figure 1. We normalize the input values by converting them to real number in the range [0..1] by dividing every value by 16.0. This is useful because the derivative of the sigmoid function is often very close to 0, which can cause the network to converge very slowly to a good set of weights. The desired output digit is converted to a target output vector for the five output units. For example, if the digit is a "5" then the target vector [0.9 0.1 0.1 0.1 0.1]. This set of target output values is preferred because the sigmoid function cannot produce the exact output values of 0 and 1 using finite weights, and so the weight values may get very large and causing overflow problems. Final numbers are 64 pixels of size 8x8 with 16 gray levels per images. Even in the limited set of examples shown, the variety of written numbers is apparent. This variety is the root of one of the challenges of automatic handwriting recognition [4].

Manuscript received on September 26, 2008.
M.J. Islam is a doctoral student of Electrical and Computer Engineering Department, University of Windsor, Windsor, ON N9B3P4, Canada.
Q.M.J. Wu is the Professor of Electrical and Computer Engineering Department, University of Windsor, Windsor, ON, N9B3P4, Canada
M. Ahmadi is the University Professor of Electrical and Computer Engineering Department, University of Windsor, Windsor, ON, N9B3P4, Canada.
M.A. Sid-Ahmed is the Professor and Head of Electrical and Computer Engineering Department, University of Windsor, Windsor, ON, N9B3P4, Canada

IACSIT International Association of Computer Science and Information Technology WWW.IACSIT.ORG

III. Artificial Neural Network

Artificial Neural Network (ANN) is a powerful classifier that represents complex input/ output relationships. It resembles human brain in acquiring knowledge through learning and storing knowledge within inter-neuron connection strengths. Some part of this section is adapted from [5]. Commonly, ANN's synaptic weights are adjusted or trained so that a particular input leads to a specific desired or target output. Figure 2 shows the block diagram for a supervised learning ANN, where the network is adjusted based on comparing neural network output to the desired output until the network output matches the desired output. Once the network is trained it can be used to test new input data using the weights provided from the training session.

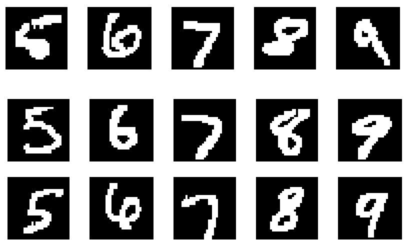


Fig. 1: Sample Handwritten Digits

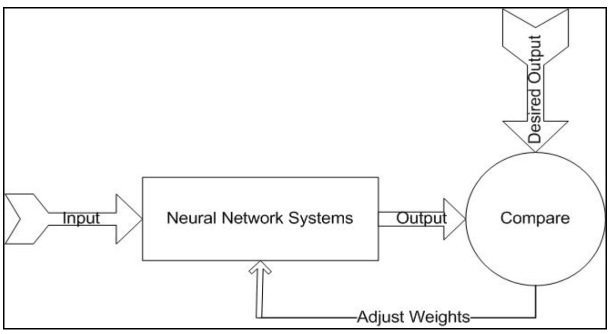


Fig. 2: Supervised Learning of ANN

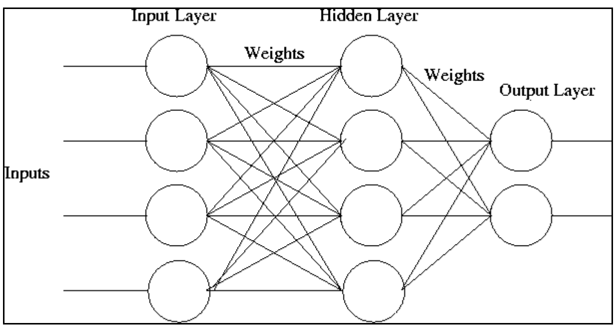


Fig. 3: Multi-Layer Feed-Forward Neural Network

F. The Multi-Layer Feed-Forward Network Model

Figure 3 represents the structure of a multi-layer feed-forward network. The neurons in this model are grouped in layers, which are connected to the direction of passing signal. There are no lateral connections within each layer and no feed-backward connections within the network.

Multi-Layer perceptron (MLP) is the most common neural network model. It is a hierarchical structure of several perceptron that uses supervised training methods to train the network. The training of such a network with hidden layer is complicated. That's why when there exists an output error; it is hard to know how much error comes from the input nodes, other nodes and how to adjust the weights according to their contributions [5]. The problem can only be solved by finding the effect of all the weights in the network. This is solved by the back-propagation algorithm which is a generalization of the least-mean-square (LMS) algorithm.

MLPs contain three layers: the input layer, hidden layers and output layer that is obvious from the figure 3. The input nodes and the hidden nodes are connected via variable weights using feed-forward connections. The output of the hidden layer nodes is connected to the input of the output layer nodes via weights. Details of the back-propagation can be found [6, 7]. The calculated output is compared with the target output. The total mean square error (MSE) is computed using all training patterns of the calculated and target outputs are as follows:

$$MSE = \sum_{j=1}^{m} \frac{1}{2} \sum_{i=1}^{5} \left(T_{ij} - O_{ij}\right)^2 \tag{1}$$

Where m is the number of examples in the training set, T_{ij} is the target output value (either 0.1 or 0.9) of the i^{th} output unit for the j^{th} training example, and O_{ij} is the actual real-valued output of the i^{th} output unit for the j^{th} training example.

IV. Matlab Simulation

To recognize the digits 5, 6, 7, 8 and 9 correctly the input image is provided of size 8x8, given as a row-major ordered sequence of pixel brightness values, which are integers in the range [0..16]. Thus the input layer will have 64 units. The output layer will have five units, one for each of the possible classifications of the input. The goal is to train the network so that if the input image contains a "5", for example, then the first output unit's activation should be near 1 and all the other output unit's activations should be close to 0. Similarly, if the input image contains a "6" then the second output unit's value should be near 1, etc. For example, hidden layer has 10 units. Each unit in the input layer is connected to every unit in the hidden layer, and each unit in the hidden layer is connected to every unit in the output layer that is shown in figure 3. So the entire network contains *((64 + 1) * 10) + ((10 + 1) * 5) = 705* weights. The extra "+1"s in this formula is because each unit in the hidden layer and in the output layer also has an associated bias, which is treated as an extra weight with constant input value -1. Each unit in the hidden layer and the output layer computes the sigmoid function. This function returns a real value in the range [0..1], not a binary value as the linear threshold unit does. Sigmoid function refers to the special case of logistic function and is defined by the equation 2 and is shown in figure 4.

$$f(x) = \frac{1}{1+e^{-x}} \tag{2}$$

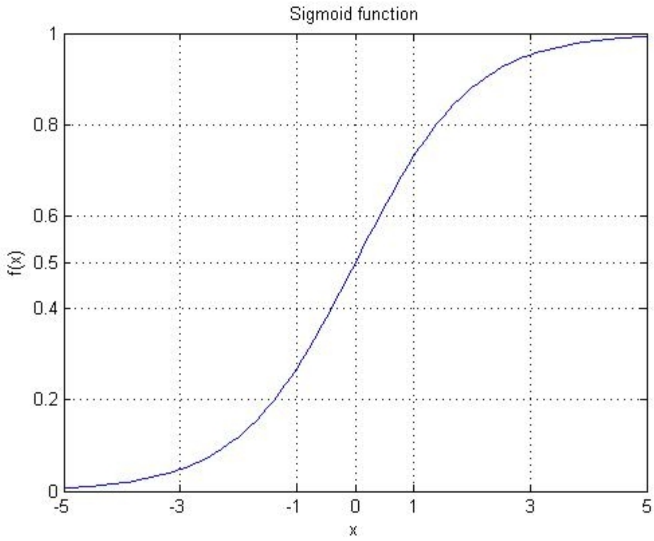


Fig. 4: Sigmoid function

To satisfy the requirements, a 5-fold cross- validation experiment is implemented to evaluate the performance of the network. To do this, the input file is divided into five parts, each containing 100/5=20 examples. For each of the five runs, 4 parts are used as the training set and the 5th part is used as the test set. The training of the network is continued at least for 200 epochs of the training set or until the MSE is 0.2 or less than that [8].

Once the network is trained using the training set of 80 examples in each run, the testing phase begins its recognition. In this phase a set of 20 examples are used to obtain the correct percentage of outputs. To compare the performance and to obtain the optimal set of parameters, the training and testing steps are repeated for different values of hidden units and learning rates. In each case the percentage correct output, training time, testing time, number of iterations is recorded and finally we obtain the optimal set of parameters for this application using the dataset.

V. RESULTS AND PERFORMANCE ANALYSIS

To fulfill the objectives of this work rigorous methodological steps are followed. The first requirement is to implement the back-propagation algorithm to recognize the handwritten digits like 5, 6, 7, 8 and 9. To do this all the weights are initialized in the network to random values in the range [-1.0, 1.0]. The learning rate $a = 0.2$ is selected and the number of hidden units is 10. So the overall structure of the network is shown in Table I.

Table I**:** Neural Network structure

Input units, X_i	64
Hidden units, Z_i	10
Output units, Y_i	5
Learning rate, a	0.2

The above network is trained using 4 parts of the data (total 80 examples) and the 5th part is used for testing. The total training and testing is continued for 5 times. So, total number of training dataset becomes 400 and testing dataset becomes 100. The MSE curve corresponding to the epoch in each run is shown in figure 5. The percentage correct output, training time, testing time, and the number of iterations required to converge to *MSE=0.2* in each run and the average percentage correct output, average training time, average testing time, and average iterations required is shown in Table II from the experiment.

Now, to obtain the optimal set of parameters different network structures are evaluated varying the number of hidden units and learning rate. For example, by keeping the learning rate a =0.2 unchanged, the hidden units are varied and the network is evaluated for the values of 5, 10, 15, 20, 25, and 50. The result for percentage correct output and iterations are shown in Table III and IV respectively.

From the Table III and IV, it is clear that in terms of percentage correct output, and iterations required to converge to *MSE=0.2*, the optimal result is found for hidden units 15. Based on that optimal value of hidden units 15, next step is to select the learning rate for which best recognition rate can be obtained. To do this hidden units are kept unchanged at h=15 and learning rates a are varied in the range of [0.05-0.30], incrementing each time by 0.05. The network is evaluated for each structure. The result for percentage correct output and iterations are shown in table V and VI respectively for the above structure varying the learning rate (a). It was found that if we decrease the learning rate, the network becomes too slow, although in our case it does not show any significant improvement in recognizing correct output. So eventually the training time and the number of iterations increased. At the same time, if the learning rate is increased the network is trained too fast, but there is a chance to skip the global minima.

From Table V and VI, it is obvious that in terms of percentage correct output, and number of iterations required to get the *MSE=0.2* or lower, the learning rate a =0.25 provides the best result. Finally the network is evaluated using the best values for *h=15* and a =0.25 and the corresponding training plot (Epoch vs. MSE) is shown in figure 6 and the outputs are shown in Table VII.

Finally the comparative statement of our proposed optimal network to the other methods like linear classifiers and neural networks [2] with different hidden units are shown in Table VIII.

Table VIII**:** Performance of the Proposed Network Structure

Classifier	Hidden Units	Error Rate (%)
Linear Classifier [2]	-	12.0
Neural Network [2]	300	4.7
Neural Network [2]	1000	4.5
NN (Proposed)	**15**	**6.0**

From table VIII, the performance of the proposed optimal network is obvious. Linear classifier has huge error rate. Although the neural network has the error rate 4.7% and 4.5%, but the number of hidden units are used 300 and 1000 respectively that definitely increases the recognition time. In our proposed optimal network only 15 hidden units are used which is very low compare to other neural network methods and the correct recognition rate is 94%.

VI. Conclusions

Network performance depends on many factors including high accuracy, low run time, low memory requirements, and reasonable training time. In this paper, in addition to accuracy, we look at measures the training time, testing time and number of iterations required to converge to a specific MSE like 0.2 in our case. In this paper, back-propagation learning algorithm was successfully applied to a large, real world task. Our results appear to be at the state of the art in handwritten digit recognition. Back-propagation was implemented for the 64 input units, different hidden units and 5 output units to recognize 5 handwritten digits like 5, 6, 7, 8 and 9. This algorithm can be easily implemented for recognizing 10 digits (0-9) by doing some minor modifications. The network parameters hidden units, learning rate plays an important role for obtaining high percentage of correct output. In this experiment hidden units $h=15$ and learning rate α =.25 is found to be the optimal. Finally, the optimal parameters are applied to the same set of data and 94% recognition rate is achieved. This result is found to be the superior in terms of complexity of the network, training time and testing time.

References

[1] Alpaydin E., An Introduction to Machine Learning, The MIT press, Cambridge, Massachusetts, London, England, 2004.
[2] LeCun Y. et al., Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 1999, pp. 2278–2324.
[3] Dataset [online] Available: http://pages.cs.wisc.edu/dyer/cs540/hw/hw4/digits5-9.dat.
[4] Duda R. et. al. Pattern Classification. Wiley Inter-science, 2nd ed. 2000.
[5] Hamze F., Technical report, Stanford University, March 2000.
[6] Alginahi Y. 2004, Computer analysis of composite documents with non-uniform background. PhD Thesis, Electrical and Computer Engineering Department, University of Windsor, Windsor, ON, Canada.
[7] Sid-Ahmed M.A., Image Processing: Theory, Algorithm and Architectures. McGraw-Hill, 1995.
[8] Wu Q.M.J., Class Notes- Machine Learning and Computer Vision. University of Windsor, Windsor, ON, Canada, 2007.

M.J. Islam was born in Comilla, Bangladesh on June 1, 1975. He received the BSc. (Hon's) and MSc. in Electronics and Computer Science in 1995 and 1996 respectively from Shahjalal University of Science and Technology, Sylhet, Bangladesh. He received MASc in Electrical and Computer Engineering from Ryerson University, Toronto, ON, Canada, in 2003. Currently he is doing PhD in Electrical and Computer Engineering at University of Windsor, Windsor, ON, Canada. The author is the IEEE student member and the member of International Association of Engineers (IAENG). His research interests include but not limited to machine learning and computer vision, image processing and document image analysis.

Q.M.J. Wu received the Ph.D. degree in electrical engineering from the University of Wales, Swansea, U.K., in 1990. In 1995, he joined the National Research Council of Canada where he became a Senior Research Officer and Group Leader. He is currently a Professor in the Department of Electrical and Computer Engineering, University of Windsor, Windsor, ON, Canada. He was recently named as the Canada Research Chair (CRC) in Automotive Sensors and Sensing Systems. He has published over 100 scientific papers in areas of computer vision, neural networks, fuzzy systems, robotics, micro-sensors and actuators, and integrated micro-systems. His current research interests include 3-D image analysis, active video object tracking and extraction, vision-guided robotics, sensor analysis and fusion, wireless sensor networks, and integrated micro-systems.

M. Ahmadi received the B.Sc. degree in electrical engineering from Arya Mehr University, Tehran, Iran, in 1971 and the Ph.D. degree in electrical engineering from Imperial College of London University, London, UK, in 1977.
He has been with the Department of Electrical and Computer Engineering, University of Windsor, Windsor, Ontario, Canada, since 1980, currently as University Professor and Director of Research Center for Integrated Microsystems. His research interests include digital signal processing, machine vision, pattern recognition, neural network architectures, applications, and VLSI implementation, computer arithmetic, and MEMS. He has co-authored the book Digital Filtering in 1 and 2 dimensions; Design and Applications (New York: Plennum, 1989) and has published over 400 articles in these areas. He is the Regional Editor for the Journal of Circuits, Systems and Computers and Associate Editor for the Journal of Pattern Recognition, and the International; Journal of Computers and Electrical Engineering.

M.A. Sid-Ahmed was born in Alexandria, Egypt, in March 1945. He received the BASc. degree in Electrical Engineering from the University of Alexandria, Egypt, in 1968 and the Ph.D. degree in Electrical Engineering from the University of Windsor, Windsor, Ontario, Canada, in 1974.
After graduation, he worked at the Alberta Government telephones and the University of Alexandria, Egypt. He joined University of Windsor in 1978 as a faculty member and is presently the department head for the ECE program. Dr. Sid-Ahmed holds four US patents, published a book on Image processing with McGraw-Hill and authored and co-authored over 100 papers. His areas of interest include image processing, robotic vision, pattern recognition and Improved Definition Television

TABLE II: OUTPUT OF THE NETWORK. H = 10 AND a = 0.2

Output	Run:1	Run:2	Run:3	Run:4	Run:5	Average
Iterations	380	410	290	340	300	344
Correct output (%)	90	100	90	95	90	93
Training time (sec)	92.157	97.390	70.828	83.203	73.828	83.48
Testing time (sec)	0.0310	0.0320	0.0310	0.0310	0.0150	0.0280

TABLE III: CORRECT OUTPUT (%) VS. HIDDEN UNITS, H. a = 0.2

Hidden units (h)	Run:1	Run:2	Run:3	Run:4	Run:5	Average Output (%)
5	95	100	90	100	90	95
10	90	95	90	95	95	93
15	**95**	**100**	**90**	**95**	**95**	**95**
20	95	100	90	100	90	92
25	85	100	90	95	90	92
50	80	95	90	100	95	92

TABLE IV: ITERATIONS VS. H TO CONVERGE TO MSE = 0.2. a = 0.2

Hidden units (h)	Run:1	Run:2	Run:3	Run:4	Run:5	Average
5	200	540	1330	600	1290	792
10	480	360	290	340	250	344
15	**310**	**380**	**320**	**380**	**260**	**330**
20	320	400	280	440	370	362
25	320	360	390	400	330	360
50	280	350	260	330	320	308

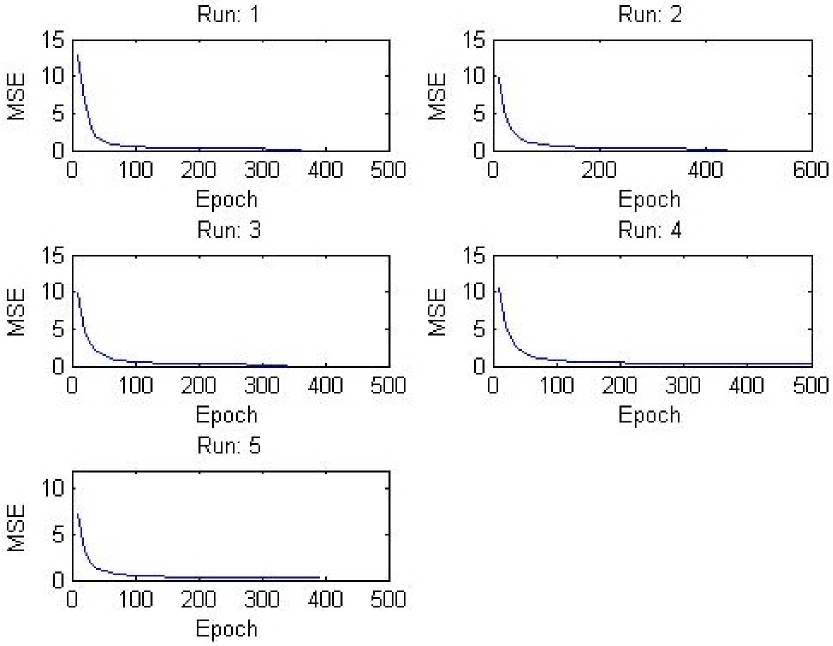


Fig. 5 : Epoch vs. MSE. Hidden units, h= 10 and $a = 0.2$

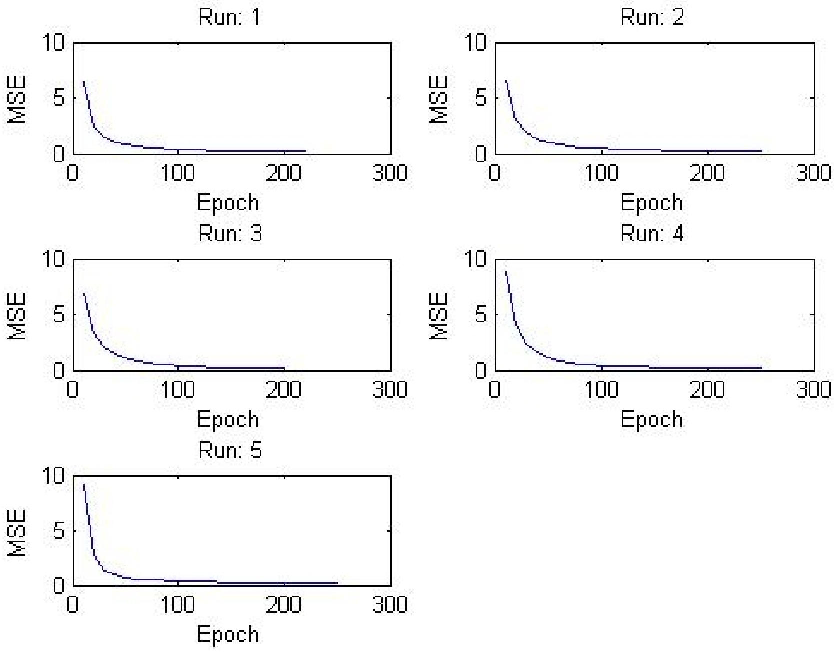


Fig. 6: Epoch vs. MSE. Optimal network for h = 15 and a = 0.25

TABLE V: CORRECT OUTPUT (%) VS a . H = 15

Learning Rate (a)	Run:1	Run:2	Run:3	Run:4	Run:5	Average Output (%)
0.05	85	100	95	100	95	95
0.10	95	100	90	100	95	96
0.15	90	95	90	100	95	94
0.20	85	100	95	95	95	94

0.25	**95**	**100**	**90**	**100**	**95**	**96**
0.30	90	100	90	100	95	95

TABLE VI: ITERATIONS VS. a TO CONVERGE TO MSE = 0.2, H = 15

Learning Rate (a)	Run:1	Run:2	Run:3	Run:4	Run:5	Average Output (%)
0.05	1120	1300	1040	1270	1540	1254
0.10	600	650	790	450	660	630
0.15	380	440	470	420	460	434
0.20	290	300	360	380	320	330
0.25	**280**	**250**	**230**	**320**	**260**	**268**
0.30	220	200	230	220	260	226

TABLE VII: OUTPUT OF THE OPTIMAL NETWORK. H = 15 AND a = 0.25

Output	Run:1	Run:2	Run:3	Run:4	Run:5	Average
Iterations	200	260	200	270	370	**260**
Correct Output (%)	85	100	90	100	95	**94**
Training Time (sec)	59.547	76.375	59.062	79.797	115.344	**78.025**
Testing Time (sec)	0.031	0.047	0.032	0.031	0.031	**0.0344**