

PCA Data Reduction for MIMO Based Channel, Capacity Enhancement Using 16-QAM Modulation

Panem Charan Arur, Deepak Chodankar, Rajendra S. Gad*, G. M. Naik

Altera SoC Lab., Department of Electronics, Goa University, India

Received October 9, 2019; Revised December 4, 2019; Accepted December 10, 2019

Copyright©2020 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract In the Modern Wireless Communication Systems, there is a direct need for the enhancement of the network capacity, in order to support more and more data. This paper describes the data reduction methodology for enhancement of channel capacity based on Principal Component Analysis (PCA) of the Multiple-Input-Multiple-Output (MIMO) based untether communication system. The communication system has employed 16 Quadrature Amplitude Modulation (16-QAM) scheme with LDPC for channel coding method for error correction. The proposed method here utilizes the Principal Component Analysis (PCA) for data reduction which transmits only the feature vectors i.e. Eigen Vector for the set of composite data for large numbers of signals data over time-stamped. The effectiveness of the proposed work is evident through the simulation results of the systems incorporating 1×1, 2×2, 3×3, 4×4 MIMO networks. We have presented RMSE analysis for the PCA regression for verification of the signal fidelity and the BER for confirmation of the recovery of the data over MIMO channel over LDPC coding. Here we observe that our proposed work has achieved much better results.

Keywords PCA, MIMO, 16-QAM, LDPC

1. Introduction

Wireless Sensors Networks (WSNs) are employed to gather variety of information which includes light, humidity, air quality, wind speed, temperature, and other vital signs. Monitoring systems are being developed in health care, industries, ecological system, governmental as well as military applications. These monitoring systems need to transmit data on continuous basis over sufficiently prolonged periods. However, it is not easy to get WSNs to achieve such widespread tasks over a prolonged period as the power resource is a serious limitation. Hence, energy

efficiency becomes the prime consideration as well as the lifetime longevity of WSNs. The major chunk of the collected data usually renders itself redundant and generally can also be easily mined from other observations. Therefore, reducing or preventing altogether, the unnecessary transmissions in a WSN has a very significant influence on reducing the overall energy consumption. These results in lifelong longevity since most of the energy is spent in the data transmissions [1].

The PCA is a technique to reduce the dimensions of the incoming sensor data. Here in this technique one applies an orthogonal transformation to transform a set of observations of interrelated variable values into a set of un-interrelated variables called the Principal Components. PCA causes a transformation of the data into another domain where the data with dominant covariance are held together and is known as the first principal component. Furthermore, there are more fragments known as the second principal component and the third principal part, and so on. Here the first principal component has the prominence [2].

The most dominant error correction code which has a near Shannon limit presentation is the Low Density Parity Check (LDPC) codes. The Shannon Limit determines the ultimate limit to the actual number of users or the maximum data rates which can be supported, within the given limits of specified bandwidth and the power constraints. The LDPC code is a Linear Error Correcting Code, which is a technique of transmitting the data in a broadcasting channel full of noise [3]. LDPC has been built by means of a Sparse Bipartite Graph. LDPC codes are used for enhancing channel capacity which permit the noise threshold to be set very close up to the theoretical maximum (Shannon limit) for symmetric channel system.

The type of Modulation is one of the most significant characteristics often used in identification as well as classification. QAM is both, an analog as well as a digital modulation technique, which has to utilize a limited number of at least two phases and two amplitudes. QAM is

constantly utilized as a modulation technique for computerized telecom systems. High spectral efficiencies may be achieved at random using QAM, by means of setting an appropriate constellation size, limited only by the corresponding noise levels and the linearity of the communication channels [4].

MIMO systems are designed using multiple transmit and receive antennas and are currently under intense research for their striking potential to accommodate a great increase in capacity. MIMO transmission takes advantage of the fact that signals at different antennas experience independent attenuation in their signal levels in a scattering environment, wherein the antennas are distributed and separated in a Fading system with multiple transmit and multiple receive antennas [4,5].

This paper is committed to the study of data at the Sensor Levels, using composite data transmission over PCA features vectors employing 16-QAM modulation schemes and LDPC error correction over MIMO Channel.

This study is carried out as per the following:

Present the sensor signal level data fusion of 45 signals originating from various sources utilizing PCA data reduction methods. Encoded the features extracted through PCA with LDPC codes for channel coding. These encoded signals are modulated with 16-QAM and transmitted over the MIMO Additive White Gaussian Noise (AWGN) channel. We have presented RMSE analysis for the PCA regression of the original composite signals, for verification of the signal fidelity. Also we have presented the BER for confirmation of the recovery of the data over MIMO channel over LDPC coding.

2. Previous Work

Over recent years, there have been immense efforts in the areas of research in data reduction in sensor networks. Among those methods which can be ordered in [6], data expectation and data fidelity are the most important classes to our artistic creations. In data forecast, sensor hubs don't constantly send their measured qualities to the base station and permit the base station in anticipating the predefined values. In this way, the base station consistently keeps up the estimation of an announcement rather than its genuine expense.

The sensor nodes, small cubage, typically deliver confined battery power and in lots of instances, changing the battery is just too costly. The strength consumption of conversation is better than that of the calculation, as an instance, the energy consumption of shifting one-bit information is approximately the same to run a thousand

CPU codes [7]. So the energy efficiency will be decreased if we at once transmit all uncooked measurements accrued through sensors to sink node. In an identical place, there exists high correlation among the sensor measurements of neighboring nodes, so it is essential to lessen the data before facts moving verbal exchange system. Sensors nodes in networks are responsible for four major tasks: data aggregation, sending and receiving data, and in-network data processing. This implies that they must effectively utilise their resources, including memory usage, CPU power and, more importantly, energy, to increase their lifetime and productivity. Besides harvesting energy, increasing the lifetime of sensors in the network by decreasing their energy consumption has become one of the main challenges of using untether communication in practical applications. In order to minimize the energy consumption to the maximum extent, the energy consumption E_{con} of the untether communication physical layer can be indicated as shown in eq(1).

$$E_{con} = ((P_{send}/\eta) + P_{amp} + PSC) * T_{all} \quad (1)$$

whereas P_{send} indicates the sending power, which is determined by the signal-to-noise ratio κ and the specific error rate, here $\kappa = (P_{accept}/2BN_0)$, among which P_{accept} refers to the receiving power, B refers to the signal bandwidth, N_0 refers to the power spectral density of additive white Gaussian noise. The relationship between the signal-to-noise ratio and the bit error rate is different under different coding schemes. We perform researches with QAM encoding mechanism in this paper; therefore, the relationship between the frame error rate and the signal-to-noise rate can be indicated as $P_{send} = Q((4\eta) * 1/2)$, among which η refers to the amplification efficiency of the signal amplifier at the sending end, Q defines numbers of phases in QAM, P_{amp} refers to the energy consumption of the power amplifier, $P_{amp} = \omega P_{send}$, PSC refers to the power consumption of the circuit at the transmitting end, in transmitting the data, consumed by the circuit at the receiving end when receiving the data, and T_{all} refers to the time required for completing data transmission, the completion of data transmission, which equals the time required by sending each datum [8]. This implies that the energy consumption will be proportional to the size of the data i.e. numbers of bits. Hence, problem of digital communication system sending the higher data (bits) automatically increases the transmission power. The present novelty focuses on when transmitting the bits in communication system reduced one can save the transmission power.

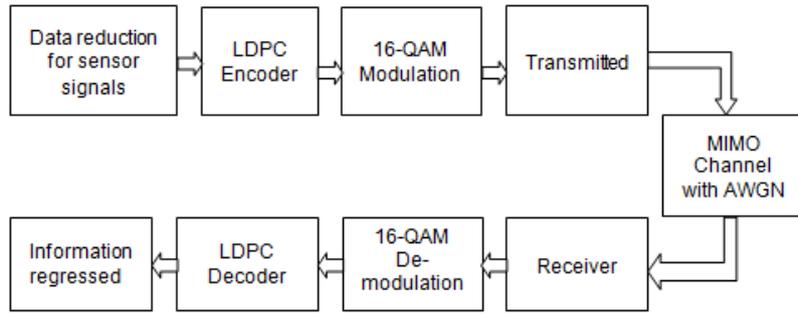


Figure 1. The proposed Sensor's Data Transmission over 16-QAM employing LDPC Error Corrections over MIMO Channels

3. Proposed Methodology

The untether communication system discussed in this paper is to improve the end to end performance regressed signal RMSE and the bit error rate (BER) of the transmitted data over the specific channel coding. The block diagram of the proposed method can be seen in Figure 1. The input block of the system is data reduction methods through PCA. The reduced data is encoded with LDPC technique and further modulated with 16-QAM. Here the 16-QAM is used as it gives high spectral efficiency compared to the 4 and 8QAM. Further the data is transmitted parallelly over the MIMO AWGN channel. As said above further high spectral efficiencies may be achieved using M-QAM, by means of setting an appropriate constellation size as shown in Figure 2.

The data is transmitted here through the channel wherein the noise is added using AWGN. At the receiver end, the inversion of the signal is performed i.e. conversion from parallel to serial, then the demodulation is performed, followed by the LDPC decoding. After the LDPC decoding, the regression of the signals from the feature matrix is performed. The methodology described above is implemented using MATLAB R2018b and this implementation flow is explained below.

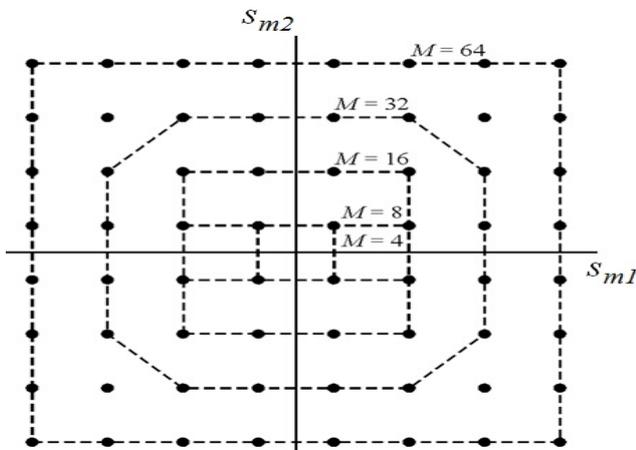


Figure 2. M-QAM constellation diagram

The signal composite grid to be broken down with the

guide of PCA includes I perceptions characterized by method for J factors and it's represented by the $I \times J$ lattice X , each discourse is a cyclic recurrence zone profile $I(\alpha)$. The lattice X has rank L where $L \leq \min(I, J)$ [9]. Constantly, the segments of X may be focused just so the mean of every section is the same as 0. Notwithstanding focusing, while the factors are estimated with explicit gadgets, it's far prominent to institutionalize every factor to unit standard. This is obtained with the asset of isolating every factor through its standard [10]. In this model, the connection framework might be composed as eq(2).

$$C_x = X^T X \quad (2)$$

C_x catches the covariance between every single imaginable pair of perceptions. The covariance esteems mirror the clamor and excess in our estimations. Next, we need to acquire the eigenvalues and eigenvectors [11].

$$C_x a_i = \lambda_i a_i \quad (3)$$

$\{a_1, a_2, \dots, a_r\}$ is the arrangement of orthonormal $J \times 1$ eigenvectors with related eigenvalues $\{\lambda_1, \lambda_2, \dots, \lambda_r\}$ for the symmetric lattice C_x . The eigenvalues are organized from enormous to little, and the eigenvectors comparing to the main $L (L < J)$ eigenvalues of the covariance lattice C_x are chosen to frame another change grid A shown eq(4).

$$A = \{a_1, a_2, \dots, a_1\} \quad (4)$$

The dimensionality reduction data set can be obtained according to the following formula eq(5).

$$Y = A^T \cdot X' \quad (5)$$

This paper utilizes the PCA algorithm to reduce the dimension of sensor data. After sensor data reduction the features are transmission using QPSK through LDPC, MIMO channels as shown in Figure 1. The basic channel capacity of an additive white Gaussian noise (AWGN) channel with B Hz bandwidth and signal-to-noise ratio S/N is shown eq(6).

$$C = B \log_2 (1 + S/N) \quad (6)$$

C is measured in bits per second if the logarithm is taken in base 2, or nats per second if the natural logarithm is used, assuming B is in hertz; the signal and noise powers S and N are measured in watts or volts², so the signal-to-noise ratio

here is expressed as a power ratio, not in decibels (dB), but it is (dBm) due to giving exact power; since figures are often cited in dB, a conversion may be needed. Here bandwidth B increases channel capacity also increases.

Further, the channel capacity can be enhanced using MIMO methods. For MIMO channel capacity is given by eq(7).

$$C = N_t N_r B \log_2 (1 + S/N) \quad (7)$$

Here N_t is the number of transmit antennas and N_r is the number of receive antennas. In this number of transmit and receive antennas are increases channel capacity increases. At the same time after input data reduction using PCA channel capacity given by eq(8).

$$C = (1/K_{dr})(N_t N_r) B \log_2 (1 + S/N); 0 \leq K_{dr} \leq 1 \quad (8)$$

Here K_{dr} is the data reduction factor. The data reduction factor is decreased than the channel capacity also increases.

If the data is transmission performed using LDPC codes, channel capacity can be expressed as eq(9).

$$C = (1/K_{dr}) K_{ldpc} (N_t N_r) B \log_2 (1 + S/N); 0 \leq K_{ldpc} \leq 1 \quad (9)$$

Here K_{ldpc} is the error correction factor, which is 1 in case of zero errors, which is an ideal case in noisy channels. Further one more factor of modulation will be integrated into the capacity of channel, this is because channel modulation factor also affects the channel capacity. So channel capacity after modulation is given by eq(10).

$$C = (1/K_{dr}) K_{ldpc} K_{mod} (N_t N_r) B \log_2 (1 + S/N) \quad (10)$$

Here K_{mod} is the modulation factor. The above equation specified the channel capacity improves by using the bandwidth of the channel (B), number of transmitting and receiving antennas ($N_t N_r$), data reduction factor (K_{dr}), Error correction factor (K_{ldpc}), modulation factor (K_{mod}) shown in eq(11).

$$K_{dr} = (Total\ data\ reduction) / (Total\ input\ data) \quad (11)$$

This performance can be further enhanced by using advanced multiplexing techniques such as Orthogonal Frequency Division Multiplexing (OFDM).

In this study, the performance of the data reduction and subsequent regression at the receiving end is quantified using Root Mean Square Error (RMSE) for the set of composite signal while the data communication is quantified using Bit Error Rate (BER). RMSE is a frequently used measure of the differences between values predicted by a model or an estimator and the values actually observed. The RMSE serves to aggregate the magnitudes of the errors in predictions for various times into a single measure. Let's assume y_i the observed value for the i_{th} observation and \bar{y}_i is the predicted value. It can be positive or negative as the predicted value under or over estimates the actual value. Squaring the residuals, averaging the squares, and taking the square root gives us

the RMSE. Then we can use the r.m.s. error as a measure of the spread of the y_i values about the predicted \bar{y}_i value for n number of observations, and hence the RMSE is given by Equation (12).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\bar{y}_i - y_i)^2}{n}} \quad (12)$$

The BER gives information about the how many bits are corrupted with noise at the receiver circuit. The system is tested by transmitting one-dimensional binary data and the BER analysis is performed. Here we have computed the BER for various signal strength or power over variance of AWGN for noise amplitude. Hence, the power of the signals can be stated as in equation (13).

$$BER = 2Q \left(\sqrt{\frac{P_s}{\sigma^2}} \right) \left(1 - \frac{1}{2} Q \left(\sqrt{\frac{P_s}{\sigma^2}} \right) \right) \quad (13)$$

P_s is the power the signals, the noise variance is σ^2 , and Q is the standard function corresponding to the type of modulation.

4. Simulation Results

From perspective of the capacity of the channel, QPSK performance is better compared to BPSK but overall QAM is most suitable for massive MIMO system. It efficiently reduces the Inter Symbol Interference (ISI) and increases the performance as compared to the conventional CDMA system since the BER is better for QPSK, BPSK and QAM modulation. The work also gives a motivation to scale MIMO for next generation network such scaled MIMO can meet the demands and expectation of telecommunication for future generation [12].

Here we have integrated the channel coding technique using LDPC encoder and decoder with MATLAB Simulink (2018B). Using Simulink library in communication system, we have generated composite data in matrix form of block size (100,45) of 16-bit wordlength corresponding to 45 sensors signal data having 100 time stamped data samples. This is reduced to (45,45) using PCA data transformation and subsequent reduction in form of features matrices or eigen matrix having 16-bit wordlength. This reduces matrix is mapped to LDPC for the code rate of $R = 1/2$ to generate the block length of 64800 bits at the LDPC encoder having 32400 message bits. The encoded data from LDPC encoder is later modulated using 16-QAM modulator with 2 bits per symbol. Channel selected here is AWGN with SNR values ranging from -15dB to 40 dB. The system is further tested (refer to Table 1) Please for the performance over LDPC codes for 50 training iterations. The channel is also configured for 2x2, 3x3 and 4x4 MIMO over untether communication.

Table 1. Parameters used in data communication system.

| Parameter | Value |
|-----------------------------------|----------------------------|
| LDPC code rate | $\frac{1}{2}(32400/64800)$ |
| LDPC Encoder Input Type | Bit |
| LDPC Decoder output Type | Information Part |
| LDPC Decoder Decision Input Type | Hard Decision |
| LDPC Decoder Number Of Iterations | 50 |
| Modulator Type | 16-QAM |
| Modulator Input Type | Bit |
| Modulator Symbol Order | Binary |
| Demodulator Output Type | Bit |
| LLR Algorithm | LLR |
| Channels | SISO&2x2,3x3,4x4 MIMO |
| Channel Noise Factor | SNR |

The use of LDPC (Low Density Parity Check) code as a linear block code with parity-check matrix containing fewer '1' bits compared with '0' bits proved to produce excellent error correction performance [13]. MIMO techniques such as Alamouti-STBC offer some benefits such as greater network access capabilities that help to make transmission more reliable, while LDPC codes give higher transmission reliability because they have good error correction codes and offer an approximation to the limit of capacity established by Shannon [14]. LDPC codes from the obtained parity check matrix and performed BER simulations using massive MIMO, QAM modulation [15].

In figure 3 and 4, we have presented the performance curves for RMSE and BER for various values of SNR and various configuration of MIMO over 16-QAM modulation. From this figure shows that 4x4 MIMO with LDPC good performance with efficiency up to -10 dB compare with 1x1 MIMO.

The RMSE performance (refer to figure 3) of data reduction with PCA over 16-QAM modulation with and without LDPC error correction over AWGN channel having SNR -15dB to 40dB. One can see that the performance of RMSE is better with LDPC channel coding as compared to the without LDPC error corrections for 1x1, 2x2, 3x3 and 4x4 MIMO channels. The RMSE is 1.0 for 1x1 MIMO at +5 dB SNR and similarly the value of RMSE is 1.0 for 4x4 MIMO at -5dB SNR without LDPC channel coding. This indicated that the performance is better for higher MIMO channel. So this indicates that one can have error-free signals transmission up to -5dB without LDPC error correction coding for proposed system. Similarly the RMSE is 1.0 for 1x1 MIMO at +2.5dB SNR and similarly the value of RMSE is 1.0 for 4x4 MIMO at -13.5dB SNR with LDPC channel coding. This implied that the channel coding has better RMSE performance as compared to non-channel coding communication systems.

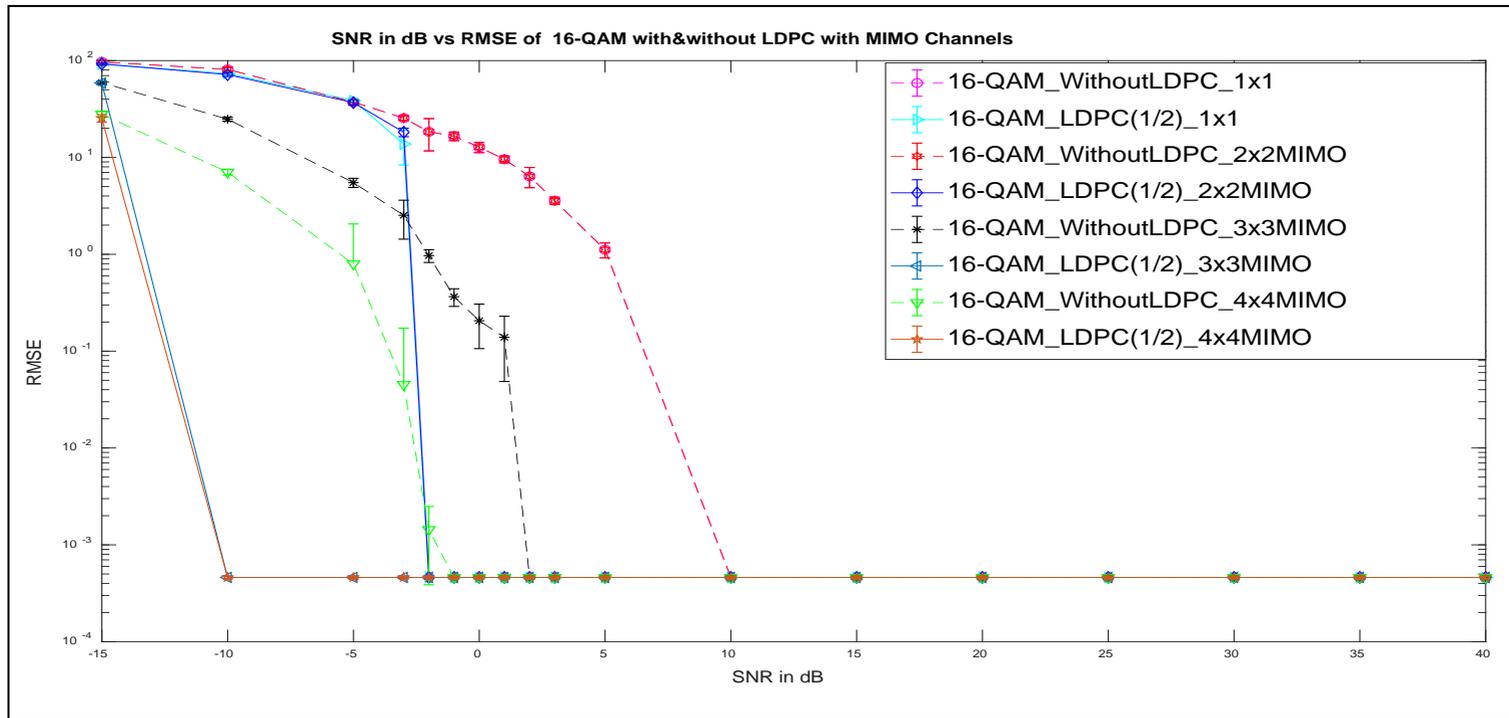


Figure 3. RMSE performance of LDPC code with code rate set to $R = 1/2$ over the assigned AWGN channel using the technique of PCA-16QAM modulation.

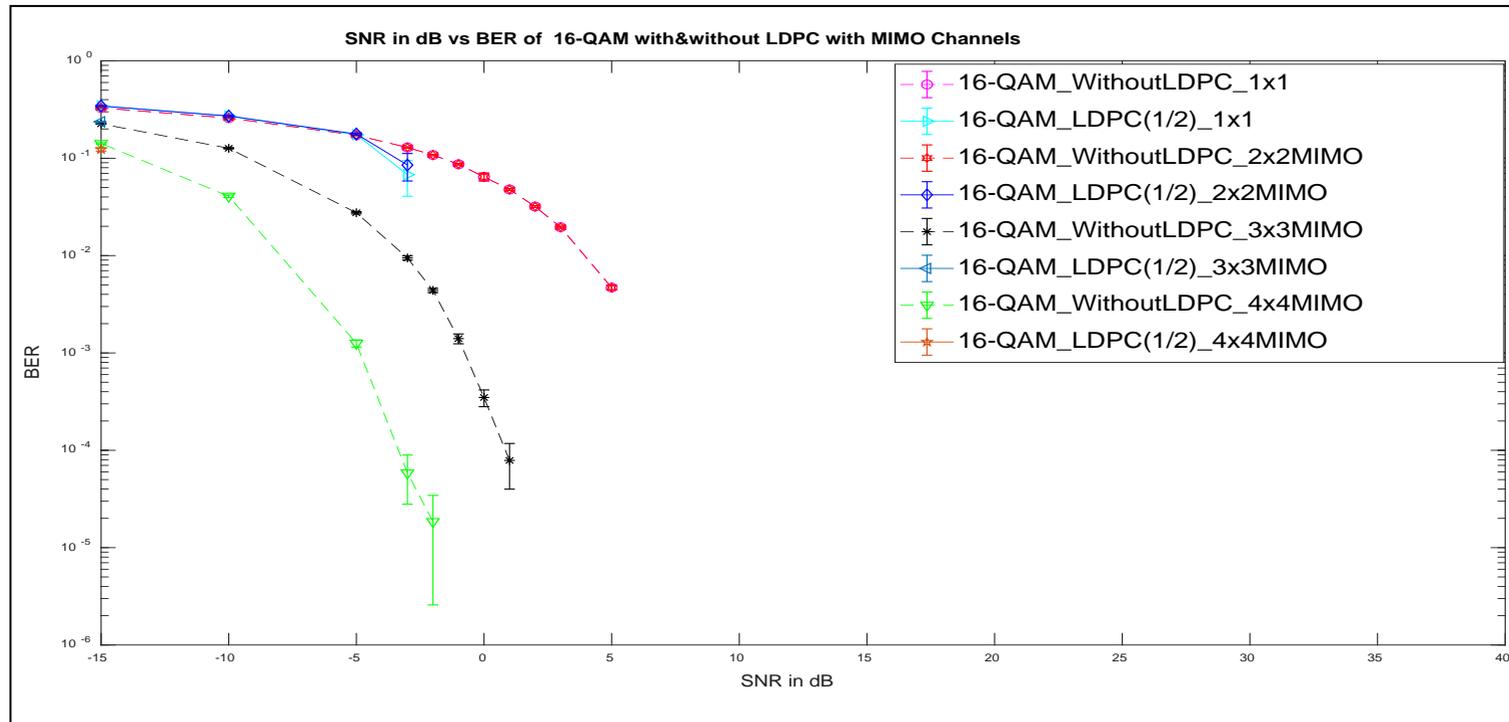


Figure 4. BER performance LDPC code with code rate set to $R = 1/2$ over the assigned AWGN channel using the 16-QAM modulation technique.

We have also presented the BER performance in figure 4. The performance of data reduction methods using PCA over 16-QAM modulation, with and without LDPC error correction for AWGN channel is presented. Performance for the AWGN channel with varying SNR -15dB to 40dB is evaluated. One can see that the performance of BER is better with LDPC channel coding as compared to the without LDPC error corrections. The BER value is 0.01 at +4dB, +4dB, -3dB and -8dB for 1x1, 2x2, 3x3, and 4x4 MIMO channel respectively without LDPC channel coding. This implies that 4x4 MIMO channel is performing better compared to 1x1 MIMO channel for error correction. Similarly the BER value is 0.2 at -5dB, -5dB, -15dB and -15dB for 1x1, 2x2, 3x3, and 4x4 MIMO channel respectively with LDPC channel coding. This implies that 4x4 MIMO channel is performing better compared to 1x1 MIMO channel for error correction with channel coding communication systems. Therefore, the system outperforms significantly the distortion when the antennas number in the channel is increased. The gain obtained with MIMO 4x4 provides the possibility to work with greater capacity in the system without a trade-off of interference and distortion. For this reason is very important to choose the correct modulation index. This proves that LDPC encoding works very well on a 4x4 MIMO even on the small value of SNR, wherein SNR has a larger noise value than the signal itself.

5. Conclusions

Data Reduction Methods are not much prevalent presently in the areas of sensor networks. In some ongoing works, PCA has been utilized broadly to extract the required features out of the Wireless Sensor Data. Typically, the data aggregation method [16] is utilized using PCA that help coordinate the certainties from more than one sensor. Borgne et. al. [17] support a decentralized apportioned PCA in wi-fi sensor systems to blend the transmitted information. The essential problem of PCA is the normal sending of refreshed eigenvectors from a base station to its sensor hubs. In [18] a path alluded to as Distributed PCA (DPCA) is proposed to explain this problem. By utilizing DPCA, each sensor can compute the required components of the eigenvector with no interchanges to the base station. The DPCA utilizes the printed idea of wi-fi correspondences and accepts that sensor hub which isn't inside the radio scope of each extraordinary has no information now.

The system that we have proposed here has been properly evaluated and the performance observations have been presented here in terms of the BER and the RMSE. It is observed that the values of BER/RMSE are enhanced with the use of higher-order MIMO channels. There is plenty of scopes to work towards the improvement of the system to much higher levels by utilizing the most efficient methodologies and techniques of the data transmission in

wireless communication systems. One can apply data prediction methods and study the performance in terms of BER and RMSE over the raw composite input data matrix by dropping some data points and predicting them through some interpolation along with regression methods explain in the paper. This kind of system could be useful in wireless data transmission i.e. 5G, 6G, WSN applications.

REFERENCES

- [1] G. Caire, G. Taricco, and E. Biglieri, "Optimum power control over fading channels," *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp.1468–1489, July 1999.
- [2] J. P. Kermoal and L. Schumacher, "A Stochastic MIMO radio channel model with experimental validation", *IEEE J. Select. Areas in communication*, VOL.20, NO. 6, Aug. 2002.
- [3] H. Sankar, N. Sindhushayana, and K.R. Narayanan, "Design of LDPC codes for high order constellations," in *Proc. IEEE Globecom '04*, Dallas, TX, Nov. 2004, vol. 5, pp. 3113–3117.
- [4] D. W. Bliss, Keith W. Forsythe, and Amanda M. Chan "MIMO Wireless Communication", *Lincoln Lab. Journal*, Vol. 15, No. 1, 2005.
- [5] B. Muquet, E. Biglieri, A. Goldsmith, and H. Sari, "An analysis of MIMO techniques for mobile WiMAX systems," in *Advances in Mobile WiMAX*. New York: Wiley-IEEE Press, 2008.
- [6] G. Anastasi, M. Conti, M. D. Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: A survey," *Ad Hoc Networks*, vol. 7, pp. 537–568, 2009.
- [7] M. Chen and M. L. Fowler, "Data compression trade-offs in sensor networks," *Proceedings of SPIE*, Vol. SPIE-5561, pp. 96-107, 2004.
- [8] Chunliang Zhou, Ming Wang, Weiqing Qu, and Zhengqiu Lu, "A Wireless Sensor Network Model considering Energy Consumption Balance," *Mathematical Problems in Engineering*, vol. 2018, Article ID 8592821, 8 pages, 2018.
- [9] Abdi H, Williams LJ. *Principal Component Analysis*[J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, 2(4): 433-459.
- [10] Y. Yao and G. J., "The cougar approach to in-network query processing in sensor networks," *SIGMOD Record*, vol. 31, no. 3, pp. 9–18, 2002.
- [11] Shlens J. A Tutorial on Principal Component Analysis[J]. *In Systems Neurobiology Laboratory, Salk Institute for Biological Studies*, 2014, 51(3):219-226.
- [12] Arun Kumar, Manisha Gupta, "Design, comparative study and analysis of CDMA for different modulation" *Egyptian Informatics Journal*, 2015(16), pp.351-365
- [13] I. Bestari, I. G. P. Astawa and Arifin, "MIMO-OFDM 2x2 based on single RF using LDPC code and Kalman filter detection," *2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA)*,

Surabaya, 2017, pp. 69-74.

- [14] R. Prieto, A. Abril and A. Ortega, "Experimental Alamouti-STBC using LDPC codes for MIMO channels over SDR systems," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), Windsor, ON, 2017, pp. 1-5.
- [15] Inho Hwang, Han Jin Park and Jeong Woo Lee, "LDPC Coded Massive MIMO Systems", Journal of Entropy, MDPI, 2019, pp.1-21.
- [16] G. Bontempi and Y. L. Borgne, "An adaptive modular approach to the mining of sensor network data," in Proc. SIAM International Conference on Data Mining, 2005.
- [17] Y. L. Borgne and G. Bontempi, "Unsupervised and supervised compression with principal component analysis in wireless sensor networks," in Proc. First International Workshop on Knowledge Discovery from Sensor Data, 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007.
- [18] Y. L. Borgne, S. Raybaud, and G. Bontempi, "Distributed principal component analysis for wireless sensor networks," Sensors Journal, vol. 8, pp. 4821–4850, 2008.