CENTRAL LIBRARY

# SCHOLARLY ARTICLES

## *A CURRENT AWARENESS BULLETIN OF ARTICLES BY FACULTY, STUDENTS AND ALUMNI*

***~ APRIL 2013 ~***


DELHI TECHNOLOGICAL UNIVERSITY

(FORMERLY *DELHI COLLEGE OF ENGINEERING)*

GOVT. OF N.C.T. OF DELHI

SHAHBAD DAULATPUR, MAIN BAWANA ROAD

DELHI 110042

# PREFACE

This is the Fourth Issue of Current Awareness Bulletin for the year 2013, started by Delhi Technological University Library. The aim of the bulletin is to compile, preserve and disseminate information published by the Faculty, Students and Alumni for mutual benefits. The bulletin also aims to propagate the intellectual contribution of DTU as a whole to the academia. It contains information resources available in the internet in the form of articles, reports, presentation published in international journals, websites, etc. by the faculty and students of Delhi Technological University in the field of science and technology. The publication of Faculty and Students, which are not covered in this bulletin, may be because of the reason that the full text either was not accessible or could not be searched by the search engine used by the library for this purpose. To make the bulletin more comprehensive, the learned faculty and Students may provide their uncovered publication to the library either through email or in CD, etc.

This issue contains the information published during April 2013. The arrangement of the contents is alphabetical wise starting from A-Z. The Full text of the article, which is either subscribed by the University or available in the web, is provided in this Bulletin.

# CONTENTS

13. Low-Voltage MOS Current Mode Logic Multiplexer ***Kirti GUPTA**, *Neeta PANDEY*, *Maneesha GUPTA*

14. NGFICA based Digitization of Historic Inscription Images ***Indu Sreedevi, *Rishi Pandey, *N. Jayanthi, *Geetanjali Bhola*** and Santanu Chaudhuryy

15. Operational Trans-Resistance Amplifier Based Tunable Wave Active Filter *Mayank BOTHRA, ***Rajeshwari PANDEY, * Neeta PANDEY**, Sajal K. PAUL*

16. Photon-Photon Collision: Simultaneous Observation of Wave-Particle Characteristics of Light ***Himanshu Chauhan, *Swati Rawal and *R.K. Sinha (corresponding author)***

17. SOFTWARE ARCHITECTURE BASED REGRESSION TESTING ***Harsh Bhasin,*** Ankush Goyal and Deepika Goyal

18. Spectrum Management Models for Cognitive Radios Prabhjot Kaur, Arun Khosla, and ***@Moin Uddin***

19. The competitiveness of SMEs in a globalized economy Observations from China and India ***Rajesh K. Singh***, ****Suresh K. Garg*** and S.G. Deshmukh

**@ Pro-Vice Chancellor**
*** Faculty**
**# Alumni**

pISSN 1976-913X
eISSN 2092-805X

J Inf Process Syst, Vol.9, No.1, March 2013 http://dx.doi.org/10.3745/JIPS.2013.9.1.089

# A Fuzzy Impulse Noise Filter Based on Boundary Discriminative Noise Detection

Om Prakash Verma* and Shweta Singh*

**Abstract**—The paper presents a fuzzy based impulse noise filter for both gray scale and color images. The proposed approach is based on the technique of boundary discriminative noise detection. The algorithm is a multi-step process comprising detection, filtering and color correction stages. The detection procedure classifies the pixels as corrupted and uncorrupted by computing decision boundaries, which are fuzzified to improve the outputs obtained. In the case of color images, a correction term is added by examining the interactions between the color components for further improvement. Quantitative and qualitative analysis, performed on standard gray scale and color image, shows improved performance of the proposed technique over existing state-of-the-art algorithms in terms of Peak Signal to Noise Ratio (PSNR) and color difference metrics. The analysis proves the applicability of the proposed algorithm to random valued impulse noise

**Keywords**—Impulse Noise, Decision Boundaries, Color Components, Fuzzy Filter, Membership Function

## 1. INTRODUCTION

Digital images can be contaminated by different types of noises. Impulse noise is one of such noises, which affect images at the time of acquisition due to noisy sensors or at the time of transmission due to channel errors, or in storage media due to faulty hardware. Various detection and filtering algorithms have been proposed. Among them, median filters [1-3] proved to be most prominent in their ability to suppress noise. Presence of three domains of filtering, i.e. spatial, frequency and fuzzy, gives ample opportunity for researchers in image processing to exploit the field of impulse noise filtering. Several linear and non-linear filters have been proposed in the literature. Linear filters are not able to effectively eliminate impulse noise as they can blur the edges. The intuitive solution to overcome the problem of linear filters is to implement an impulse-noise detection mechanism prior to filtering; hence, only those pixels identified as corrupted would undergo the filtering process, while those identified as uncorrupted would remain unchanged. With the incorporation of such a noise detection mechanism into the median filtering framework, the switching median filters [4-6] have shown significant performance improvement. Many non-linear filters based on classical and fuzzy techniques have emerged in recent years. The noise adaptive soft switching median (NASM) filter [7] proposed by Eng et al. consists of a three-level hierarchical noise detection process.

Manuscript received September 28, 2012; accepted November 3, 2012.
**Corresponding Author: Shweta Singh**
* Delhi Technological University, Delhi, India (opverma.dce@gmail.com, singh.shweta1210@gmail.com)

The NASM achieves a fairly robust performance in removing impulse noise, while preserving signal details across a wide range of noise densities, ranging from 10% to 50%.

The quality of NASM recovered images degrades as noise density increases above 50%. The Boundary Discriminative Noise Detection (BDND) [8], a switching median filter, is proposed for the detection of impulse noise based on a large difference between the noisy pixel and the noise free pixel. This paper claims to achieve better results than NASM by passing the pixel from two sizes of window (21x21 and 3x3) to confirm whether it is noisy or not. It gives acceptable results up to 80% of noise density. Srinivasan et al. [9] also presents a new decision based algorithm for restoration of images that are highly corrupted by impulse noise. It removes only the corrupted pixel by the median value or by its neighboring pixel value. Very recently, Tripathi et al. [10] presented a switching median filter which is an advanced boundary discriminative noise detection algorithm. It is also a two stage detection algorithm. Smail Akkoul et al. presented an adaptive switching median filter (ASWM) [11] which requires no prior threshold as required by a classical switching median filter. Threshold is computed locally from image pixels intensity values in a sliding window. Duan and Zhang presented a two iteration algorithm [12] for impulse noise detection for switching median filter.

In the field of fuzzy domain, the fuzzy median filter [13, 14] is a modification to the classical median filter. The Fuzzy Inference Rules by Else action (FIRE) filters [15-17] are a family of non-linear operators that adopt fuzzy rules to remove impulse noise from images. Androutsos et al. [18] designed a new class of filters called fuzzy vector rank filters, based on a combination of different distance measures. Stefan *et al*. [19] presented a fuzzy two-step color filter for the reduction of impulse noise. This filter utilizes the fuzzy gradient values and fuzzy reasoning for the detection of noisy pixels. Verma et al. [20] presents two stage fuzzy filter for reduction of both impulse and Gaussian noise in color images. It also considers the interactions between the color components. Toh at al. presents a cluster based adaptive fuzzy switching median filter [21] for universal impulse noise reduction i.e. random valued or fixed valued impulse noise. Madhu et al. proposed a new efficient fuzzy-based decision algorithm (FBDA) [22] for the restoration of images that are corrupted with high density of impulse noises. Another two phase process of fuzzy logic based impulse noise filtering technique [23] is presented by Aborisade.

Most of the above reported schemes work well under salt and pepper noise but fail under random valued impulse noise, which is more realistic when it comes to real world applications. These schemes generally use a threshold value for identification of pixel as corrupted or uncorrupted. The partitions obtained are thus strict and boundaries are rigid. The fuzzy based boundary discrimination detection algorithm presented here considers smooth boundaries which are computed for each candidate pixel. In the present paper, a two step impulse noise filter is proposed. This approach exploits the advantages of switching median filter and a fuzzy rule based system. The filter works in the stages of detection, filtering and color correction. For gray scale images only the first two are applicable whereas for the color images, the correction term is added after filtering to remove the residue impulse noise remaining in the color components. Simulation results are carried out for random impulse noise, and comparative analysis shows that fuzzifying the decision boundaries gives better performance than the existing techniques in terms of various image metrics.

The paper is divided into four sections. Section 2 discusses the noise models considered for result evaluation. Section 3 presents our Fuzzy Impulse Filter. Section 4 gives the results and comparisons of our approach with existing algorithms. Finally, section 5 concludes the paper.

## 2. NOISE MODELS

Four impulse noise models are presented in [8]. Out of these four models only two noise models, namely model 3 and 4, are selected in present case. *I(i,j)* represents the intensity value at location *i,j* . For the pixel, *y(i,j)*, of noisy image at location *i,j,* the noise models (renamed) are represented as follows:

1) *Noise Model 1:* An impulse noise in this model is represented with two fixed ranges of length *m* at both the ends of the gray scale. For example, if *m* is 20 then the impulse noise used to corrupt an image will have the intensity value in [0, 19] and [236, 255] with equal probability. The probability density function for *y(i,j)* in this noise model given as [8]

$$f(y) = \begin{cases} \frac{p}{2m}, & 0 \le y < m \\ 1-p, & y = I(i,j) \\ \frac{p}{2m}, & 255-m < y \le 255 \end{cases} \quad (1)$$

where *p* is the noise density.

2) *Noise Model 2:* It is an extension of above model with unequal densities of the low intensity and high intensity impulse noises and is given as

$$f(y) = \begin{cases} \frac{p_1}{2m}, & 0 \le y < m \\ 1-p, & y = I(i,j) \\ \frac{p_2}{2m}, & 255-m < y \le 255 \end{cases} \quad (2)$$

where $p=p_1 + p_2$ is the noise density.

## 3. FUZZY IMPULSE FILTER

This section discusses a fuzzy based approach to boundary discriminative noise detection. The proposed algorithm presents an impulse filter which is applicable to both gray scale and color images. The algorithm is a multi step process of detection, filtering and noise correction. Noise correction is specifically designed for color images to remove the residue noise presents in color components by exploiting the interactions between them.

The basic strategy of BDND [8] is to examine each pixel in its neighborhood from coarse to fine. The pixel under consideration is examined in two stages of different window sizes to be marked as corrupted or uncorrupted. The most critical part of this is the determination of the decision boundaries that classifies the pixels as

$$Class(y(i,j)) = \begin{cases} \text{Low intensity noise}, & y(i,j) \le b_1 \\ \text{Uncorrupted}, & b_1 < y(i,j) \le b_2 \\ \text{High intensity noise}, & y(i,j) > b_2 \end{cases} \quad (3)$$

where *y(i,j)* is the intensity of the pixel being considered, and $b_1$ and $b_2$ are two decision boundaries. The major weakness of BDND resides in selecting the strict boundaries, which is a reason for misclassification. The issue is resolved by taking the smooth boundaries for the classes (Eq. 3) in our improved algorithm for detection. Overlapped regions are treated by creating fuzzy rules to obtain a correct degree of noisyness for the candidate pixel. The detection procedure starts by subjecting a candidate pixel to 21x21 window centered on it. Decision boundaries are formed similar to [8]. Fuzzification is incorporated in the boundary discriminative noise detection with the help of three membership functions. Functions are described in detail later in this section. Different fuzzy rules are designed to consider all the possibilities of the existence of a pixel in any membership function. Each rule is given a weight, according to which a degree of noisyness is calculated which will decide the pixel is corrupted or uncorrupted. The first part of detection process ends with generating a decision map, with "0" representing the corrupted and "1" represents the uncorrupted pixel. The second part of the detection process starts by confirming the pixel's class by imposing it to a 5x5 window. The decision map formed after this step is given to the filtering stage to filter out the corrupted pixels by replacing them with median of uncorrupted ones only. The output thus obtained is a filtered image and can be considered an output image of fuzzy based filter in case of gray scale images. As mentioned, for color images another step is applied in order to remove leftover impulse noise in color components by checking the difference between the color components and devising a membership function to calculate a correction term to be added to a pixel. The complete algorithm is presented in the following subsections.

### 3.1 Noise Detection for gray scale images

The main aim of the detection step is to classify the pixels as corrupted (high or low intensity) or uncorrupted. Therefore, it is carried out by first finding the decision boundaries. These boundaries are themselves calculated in two iterations i.e. first the pixel are checked on global statistics (21x21 window) and then with local statistics (5x5 window) just to confirm the classification. The boundaries selected are fuzzified in order to avoid the rigidness of the strict thresholds. The output of the detection step is a decision map.

Steps for detection are as follows:

1) Impose 21x21 window centered on *y(i,j)*
2) Sort the pixels of the window to an ordered vector $v_o$ and find the median *med.*
3) Compute the differences between each pair of adjacent pixels in vector $v_o$. The new vector is denoted as $v_D$.
4) Find the pixels which correspond to maximum differences in $v_D$ corresponding to intervals [0, med] and (med, 255] in $v_o$. And set these two pixels' intensities as the decision boundaries $b_1$ and $b_2$ respectively.
5) With $b_1$, $b_2$ and *med* values, three membership functions are formed. This step is known as fuzzification of decision boundaries.

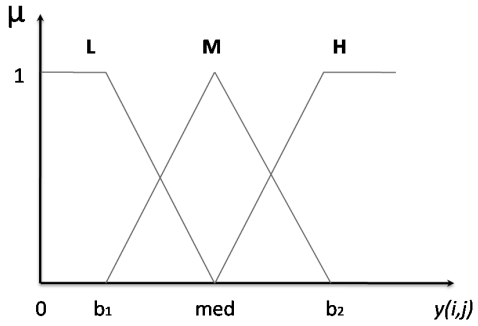



Fig. 1. The membership functions

Three membership functions $\mu_L$, $\mu_M$ and $\mu_H$ are devised corresponding to a pixel as shown in the Fig. 1. The membership function, $\mu_L$ represents a fuzzy set Low (L), indicating the pixels belonging to low intensity corrupted class. Similarly $\mu_H$ represents fuzzy set High (H) for high intensity corrupted class. And $\mu_M$ is for fuzzy set Medium (M) for pixels which are uncorrupted. The closer the value of a pixel to the boundaries $b_1$ and $b_2$, the higher is the possibility of a pixel to lie in one of the corrupted classes (low or high).

$$\mu_H = \begin{cases} 1, & y(i,j) \le b_1 \\ \dfrac{med}{med - b_1} - \dfrac{y(i,j)}{med - b_1}, & b_1 < y(i,j) \le med \\ 0, & otherwise \end{cases} \tag{4}$$

$$\mu_M = \begin{cases} 1, & b_1 < y(i,j) \le b_2 \\ 0, & otherwise \end{cases} \tag{5}$$

$$\mu_H = \begin{cases} 1, & y(i,j) \le b_2 \\ \dfrac{y(i,j)}{b_2 - med} - \dfrac{med}{b_2 - med}, & med < y(i,j) \le b_2 \\ 0, & otherwise \end{cases} \tag{6}$$

The degree of noise present in the pixel is ascertained by forming fuzzy inference rules. These are used to validate the existence of a pixel in the particular class of output. For a pixel P, the examples of rules are as follows:

**Rule 1:** *If P lies in L* ***and*** *not in M* ***and*** *not in H, then P is corrupted (low intensity).*
**Rule 2:** *If P lies in L* ***and*** *P lies in M* ***and*** *not in H, then P is corrupted (low intensity) or uncorrupted.*
**Rule 3:** *If the pixel lies in M* ***and*** *P lies in H* ***and*** *not in L, then P is uncorrupted or corrupted (high intensity).*

Each rule has a participation weight, $N_x$ assigned to it. The expression of $N_x$ is obtained by modifying the weight presented in [26].

$$N_x = \frac{l+m+h}{n(n-1)} \tag{7}$$

where *l, m* and *h* represents the index numbers of $\mu_L$, $\mu_M$ and $\mu_H$ in each rule and *n* is 3 in present case. The pixels are classified into corrupted and uncorrupted classes by obtaining the degree of noisyness given as [27]

$$N = \frac{\sum_z N_x \mu^x_{prem}}{\sum_z \mu^x_{prem}} \tag{8}$$

where *z* is the number of rules (here 27), $N_x$ is the weight of each premise and $\mu^x_{prem}$ is a certainty of premise for the $x^{th}$ rule. The certainty of premise is obtained by applying the Mamdani fuzzy rule on membership functions

$$\mu^x_{prem} = \mu_L \wedge \mu_M \wedge \mu_H \tag{9}$$

The pixel is classified into the particular output class by following the equation

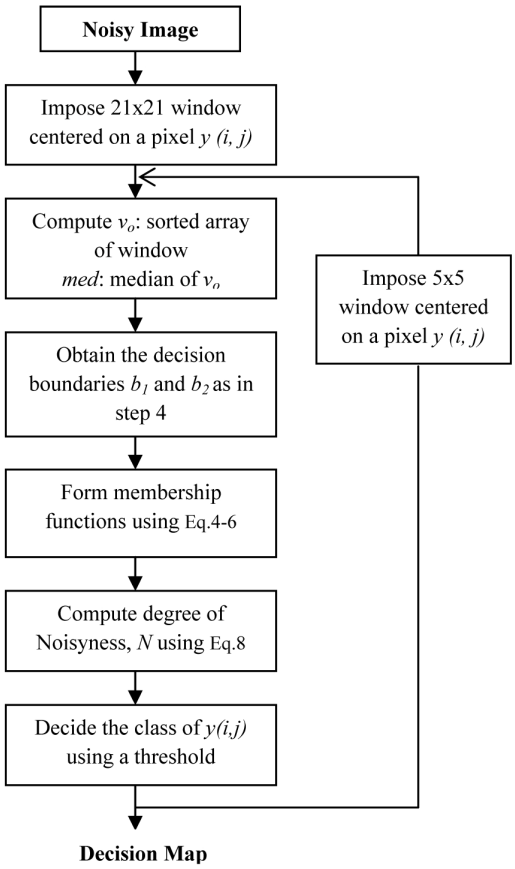



Fig. 2. Flow chart for detection stage

$$class = \begin{cases} uncorrupted, & N \le Threshold \\ corrupted, & N > Threshold \end{cases} \tag{10}$$

The value of threshold is experimentally computed.

6) The validation of noisy candidates is confirmed by imposing a 5x5 window and repeating the steps 2-5.

Fig. 2 shows the complete flow chart to depict the above steps.

### 3.2 Filtering

The decision map obtained in the previous section is given as an input to this stage. Only the uncorrupted pixels are selected for filtering i.e. the pixels represented as "0" in the decision map. The important change in the filtering technique is the approach of incremental window size [7, 8]. The window size is increased from 3x3 to 7x7 depending on the criterion that number of uncorrupted pixels should be more than or equal to half of the total number of pixels in that particular window i.e.

$$No.\,of\ uncorrupted\ \ pixels\,in\,a\,window \ge \frac{1}{2}(W \times W) \tag{11}$$

where W is window size. If the above condition holds true then the median value is assigned otherwise W is incremented by 1.

The window size is limited to 7x7 because for larger windows severe blurring takes place in high density noise. The median of the particular window is assigned to the pixel. The filtered image ***F*** obtained is considered as an output of the gray scale version whereas for the color version, ***F*** is subjected to noise correction.

### 3.3 Noise Correction for Color Images

The noise detection in color images is performed on the same lines as that of gray scale images. The membership functions are formed for each color component. The filtered image ***F*** obtained from the filtering step is subjected to this correction step. This filter invokes the interaction between the color components [20] to remove further left-over impulse noise present in color components. The most widely used RGB color space will be used in our work. Here the pixel is represented as ***F(i,j,z)*** where z ranges from 1 to 3 representing 1 for the red, 2 for the

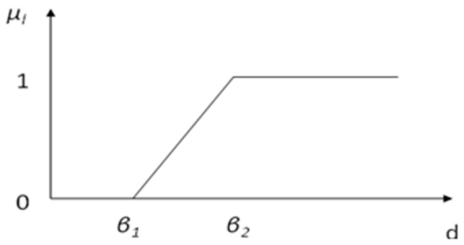



Fig. 3. The membership function "large"

green, and 3 for the blue component. The correction term to be added to the components is decided by examining the differences in each color pair and forming a membership function. The steps [20] are as described below.

1) Differences between the color pairs is calculated to check for any residual impulse noise in individual color components:

$$\begin{aligned} d_{rg}(i,j) &= \left|F(i,j,1) - F(i,j,2)\right| \\ d_{rb}(i,j) &= \left|F(i,j,1) - F(i,j,3)\right| \\ d_{gb}(i,j) &= \left|F(i,j,2) - F(i,j,3)\right| \end{aligned} \tag{12}$$

where $d_{rg}(i,j)$, $d_{rb}(i,j)$, $d_{gb}(i,j)$ represent differences between red-green, red-blue and green-blue components for the same pixel of the filtered image F.

2) A fuzzy rule system is framed to compute the degree of noise present in the color component of the pixel concerned. For each component the rule of the following form is formed:

**Rule:** *If $d_{rg}(i,j)$ is **Large** and $d_{gb}(i,j)$ is **Large**, then the green component is **noisy**.*

Similar fuzzy rules are coined for other color components. The significance of this rule holds only when there is an impulse noise left in the color components. The application of this stage nullifies when there is region of same color as the differences will be large for that region. Let us say there is a green region present in an image; this step will not consider this area as noisy as the median of the region will also be green. Thus it makes the filter more efficient with respect to noises present in the color components. The definition of "Large" is expressed by the membership function $\mu_l$ with the parameters $\beta_1$ and $\beta_2$ as shown in Fig. 3. For every difference computed above (generalized as d), $\mu_l$ is given as

$$\mu_l = \begin{cases} 0, & d < \beta_1 \\ \dfrac{d}{\beta_2 - \beta_1} - \dfrac{\beta_1}{\beta_2 - \beta_1}, & \beta_1 < d \le \beta_2 \\ 1, & d \ge \beta_2 \end{cases} \tag{13}$$

3) The degree of noise, $n_d$ in the color component is obtained by the minimum value of the membership functions corresponding to the differences. The correction term is given by:

$$\Delta(i,j,z) = n_d(i,j,z) \times (med(i,j,z) - F(i,j,z)) \tag{14}$$

4) The final output of the impulse filter is given by:

$$O(i,j,z) = F(i,j,z) + \Delta(i,j,z) \tag{15}$$

Fig. 4 shows the flow chart representing the second stage of the proposed algorithm.

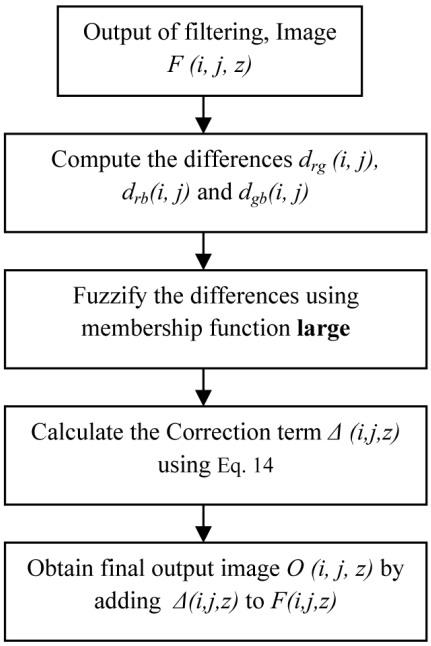



Fig. 4. Flow chart of Noise Correction

## 4. RESULTS AND COMPARATIVE ANALYSIS

The simulation results are obtained both on gray scale and color images. The proposed approach is compared with four existing algorithms, ASWM [11], BDND [8], SMF [10] and DBAIN [9]. The bases for comparisons are image metrics which are selected according to the type of an image. All the simulations are carried out with Matlab 2010b on Intel Core 2 Duo CPU of 2.53 GHz speed and 1.99 GB RAM machine.

### 4.1 Gray Scale Images

The performance evaluation of the filtering operation is quantified by PSNR. For an image of size MxN

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right)dB \tag{16}$$

The filtered image ***F*** of each algorithm is compared with that of proposed approach. For experiments, gray scale image of Lena of size 256x256 is selected. Two noise models are implemented as discussed in section II to compare the results. To study the quantitative performance of algorithms, an input image is corrupted with impulse noise with different densities. Here the results are shown to compare the technique with ASWM (initial alpha=20, iterations=7), BDND (two stages of 21 x 21 and 3x3 with adaptive filtering), SMF (21x21 window with four directional kernels using the thresholds values as 5/255 and 1/255) and DBAIN (3x3 window). Table 1 (a) and (b) shows the values of PSNR for all the five algorithms for noise density 50%. Fig 5(a) compares the algorithms graphically.

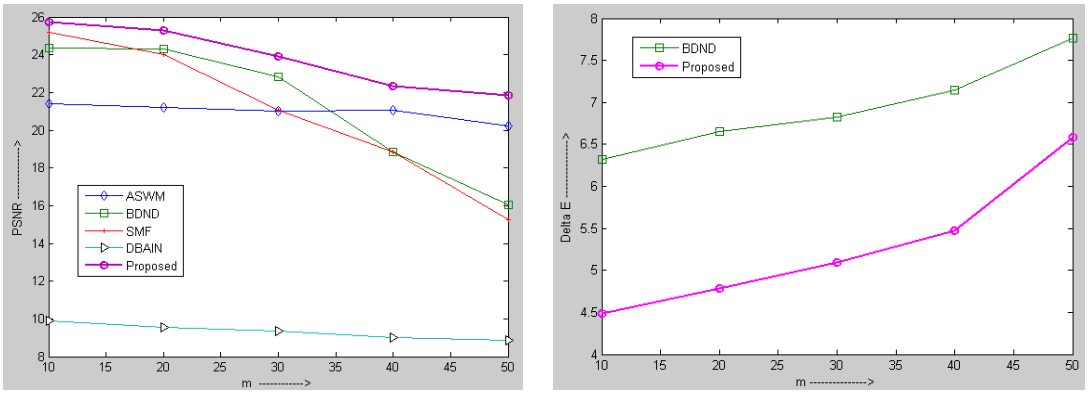
Table 1. Quantitative performance (in terms of PSNR values) on Lena (gray scale)

(1) Noise Model 1

| Noise=50% | | ASWM | BDND | SMF | DBAIN | Proposed |
|---|---|---|---|---|---|---|
| **Low Range** | **High Range** | | | | | |
| **[0,9]** | **[246,255]** | 21.42 | 24.34 | 25.20 | 9.90 | 25.72 |
| **[0,19]** | **[236,255]** | 21.19 | 24.31 | 23.99 | 9.53 | 25.31 |
| **[0,29]** | **[226,255]** | 21.01 | 22.85 | 21.05 | 9.37 | 23.91 |
| **[0,39]** | **[216,255]** | 21.06 | 18.84 | 18.86 | 9.03 | 22.32 |
| **[0,49]** | **[206,255]** | 20.22 | 16.02 | 15.26 | 8.86 | 21.85 |

(2) Noise Model 2

| Noise=50% | | m=10 | | | | | m=30 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Low Density** | **High Density** | **ASWM** | **BDND** | **SMF** | **DBAIN** | **Proposed** | **ASWM** | **BDND** | **SMF** | **DBAIN** | **Proposed** |
| **10** | **40** | 20.01 | 25.95 | 26.97 | 10.81 | 28.56 | 20.70 | 24.02 | 22.32 | 11.46 | 25.66 |
| **20** | **30** | 16.53 | 24.17 | 26.05 | 9.55 | 26.73 | 18.84 | 23.53 | 21.92 | 10.29 | 24.17 |
| **30** | **20** | 15.44 | 25.59 | 26.06 | 8.83 | 26.37 | 16.62 | 20.01 | 20.00 | 9.40 | 22.01 |
| **40** | **10** | 18.80 | 25.88 | 27.69 | 9.90 | 28.13 | 20.58 | 22.92 | 22.80 | 10.23 | 23.19 |




(1) m versus PSNR for gray scale image (2) m versus ΔE for color image

Fig. 5. Graphical comparison for noise model 1 on Lena

In the Noise Detection step, the membership functions are formed adaptively according to the value of pixel in hand. That pixel is classified, after fuzzification, into a corrupted or uncorrupted class using a threshold value of Eq. 10. Experimentally the value of threshold for which the technique gives best results is 0.5.

## 4.2 Color Images

PSNR is just a numerical value more applicable to the monochrome images, thus we need a metric which can appropriately measure the differences between the color images. In response to this need,, the perceptually uniform color space CIELAB, standardized by the Commission Internationale de l'Eclairage (CIE) [24], is more accurate for defining quantitative measurements

Table 2. Quantitative performance (in terms of mean value of ΔE) on Lena (Color)

(1) Noise Model 1

| Noise=40% | | BDND | Proposed |
|---|---|---|---|
| **Low Range** | **High Range** | | |
| **[0,9]** | **[246,255]** | 6.32 | 4.49 |
| **[0,19]** | **[236,255]** | 6.65 | 4.78 |
| **[0,29]** | **[226,255]** | 6.82 | 5.09 |
| **[0,39]** | **[216,255]** | 7.15 | 5.47 |
| **[0,49]** | **[206,255]** | 7.77 | 6.58 |

(2) Noise Model 2

| Noise=40% | | m=10 | | m=30 | |
|---|---|---|---|---|---|
| **Low Density** | **High Density** | **BDND** | **Proposed** | **BDND** | **Proposed** |
| **10** | **30** | 5.84 | 3.73 | 6.58 | 4.86 |
| **20** | **20** | 10.10 | 5.09 | 11.80 | 6.18 |
| **30** | **10** | 6.15 | 3.758 | 7.20 | 4.40 |

of perceptual error between the two color vectors. In CIELAB color space the color difference is calculated in terms of ΔE. It is given by [25]:

$$\Delta E = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \tag{17}$$

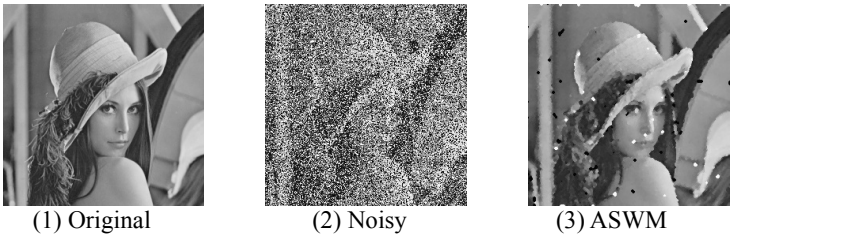
where ($L_1$, $a_1$, $b_1$) and ($L_2$, $a_2$, $b_2$) are lab transform of the RGB image. Here we will consider the average value of ΔE for complete image.

Table 2 (a) and (b) shows the comparison of BDND with the proposed approach in terms of color difference on color image of Lena of size 256x256. According to the definition of ΔE, the value of 1 means there are almost no perceptible differences or variations between two colors. The value ranging from 2-5 represents minute and 6-10 represents noticeable color differences or variations in high quality imaging systems. According to the values obtained for different noise models, our approach shows an acceptable range as compared to BDND. Fig 5(b) shows the comparison graphically. The incorporation of the second stage i.e. to add correction term to color components by exploiting the interactions between them proves to be beneficial. The parameters $\beta_1$ and $\beta_2$ used are found experimentally to be 0.5 and 0.6 respectively.

## 4.3 Visual Performances

As a final illustration and comparative analysis, Fig 6 and 7 shows the filtered and output images for gray scale and color versions of the input Lena image respectively for specified noise densities. The algorithm performs well for generalized models of impulse noise.

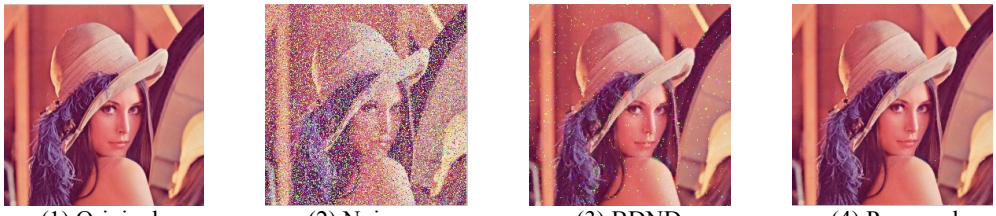
For gray scale, the proposed algorithm is being compared with all the four existing algorithms. As shown ASWM, BDND and SMF give comparable results for noise model 1 with m=10 but DBAIN does not perform well as it was only defined for salt and pepper noise. In the color version, the proposed algorithm is compared with BDND only.



(1) Original (2) Noisy (3) ASWM

(4) BDND (5) SMF (6) DBAIN (7) Proposed

Fig. 6. Outputs of different methods for Lena (gray scale) image for noise model 1 (50% impulse noise density, m=10)



(1) Original (2) Noisy (3) BDND (4) Proposed

Fig. 7. Outputs for Lena (color) image for noise model 2 (40% impulse noise density {high density=30, low density=10}, m=30)

## 5. CONCLUSION

We have presented a multi-step fuzzy approach to a boundary discriminative noise detection algorithm for both gray scale and color images. The pixels are classified by forming the fuzzy rules based on the membership functions of decision boundaries obtained using a boundary discriminative approach. The major modification in the technique is incorporated by considering the interactions between the color components. Our approach considers both kinds of interactions i.e. between the same components (e.g. R-R, G-G in the detection step) and between different components (e.g. R-G, G-B and R-B in the color correction step). Thus an additional stage for color images to consider the interaction between the color components removes any residual noise present in the color components after filtering. The quantitative, qualitative and graphical comparison in terms of PSNR and average value of $\Delta E$ proves that the proposed algorithm performs well as compared to other existing algorithms. From extensive experimentation of the proposed algorithm, it was found that it tends to give a blurred image over 50% noise density, but the quality of the output images were under considerable limits as represented by $\Delta E$ value. As shown, using two kinds of noise model, our approach is more suitable for random impulse noise. The proposed work can be extended to include other types of noise like Gaussian or mixed noise.

## REFERENCES


[1] I. Pitas and A. N. Venetsanopoulos, *"Order statistics in digital image processing,"* Proc. *IEEE*, Vol.80, No.12, December, 1992, pp.1893-1921.

[2] D. R. K. Brownrigg, *"The weighted median filter,"* Commun. *ACM*, Vol.27, No.8, August, 1984, pp.807-818.

[3] S.-J. Ko and Y. H. Lee, "*Center weighted median filters and their applications to image enhancement,"* IEEE Trans. Circuits Syst., Vol.38, No.9, September, 1991, pp.984-993.

[4] H. Hwang and R. A. Haddad, "*Adaptive median filters: New algorithms and results," IEEE* Transaction Image Processing, Vol.4, No.4, April, 1995, pp.499-502.

[5] Z. Wang and D. Zhang, *"Progressive switching median filter for the removal of impulse noise from highly corrupted images,"* IEEE Trans. Circuits Syst. II, Vol.46, No.1, January, 1999, pp.78-80.

[6] S. Zhang and M. A. Karim, "*A new impulse detector for switching median filters,"* IEEE Signal Processing Letter, Vol.9, No.4, November, 2002, pp.360-363.

[7] H.-L. Eng and K.-K. Ma, *"Noise adaptive soft-switching median filter,"* IEEE Transaction Image Processing, Vol.10, No.2, February, 2001, pp.242-251.

[8] P.-E. Ng and K.-K. Ma, *"A switching median filter with boundary discriminative noise detection for extremely corrupted images,"* IEEE Transaction Image Processing., Vol.15, No.6, Junuary, 2006, pp.1506-1516.

[9] K.S. Srinivasan, D.Ebenezer, *"A New Fast and Efficient Decision Based Algorithm for removal of High Density Impulse Noises",* IEEE Signal Processing Letters, Volume 14, Issue 3, 2007, pp.189-192.

[10] Tripathi, Ghanekar, Mukhopadhyay, *"Switching Median Filter: advanced boundary discriminative noise detection algorithm"*, Image Processing, IET, Volume 5 Issue 7, 2011, pp.598-610.

[11] Smail Akkoul, Roger Ledee, Remy Leconge and Rachid Harba, *"A New Adaptive Switching Median Filter"*, IEEE Signal Processing Letters, 2010, pp 587-590.

[12] Fei Duan and Yu-Jin Zhang, *"A Highly Effective Impulse Noise Detection Algorithm for Switching Median Filter"*, IEEE Signal Processing Letters, Vol.17, No.7, July, 2010, pp.647-650.

[13] K. Arakawa, *"Median Filter based on Fuzzy Rules and its Application to Image Restoration"*, Fuzzy Sets and Systems, Vol.77, Issue 1, January, 1996, pp.3-13.

[14] K. Arakawa, *"Fuzzy-Rule Based Image Processing with Optimization"*, in *"Fuzzy Techniques in Image Processing"*, E.E. Kerre, M. Nachtegael (Editors), Physica-Verlag, Heidelberg, New York, 2000, pp.222-247.

[15] F. Russo, "*FIRE operators for Image Processing*", Fuzzy Sets and Systems, Vol.103, 1997, pp.265-275.

[16] F. Russo, *"Noise Cancellation using Non-linear Fuzzy Filters",* Proc. of IEEE Instrumentation and Measurement Technology Conference, May, 1997, pp.772-777.

[17] F. Russo and G. Ramponi, *"A noise smoother using cascaded FIRE filters"*, Proc. of 4th Intl. Conf. on Fuzzy Systems, Vol.1, 1995, pp.351-358.

[18] D. Androutsos, K.N. Plataniotis, and A.N. Venetsanopoulos, *"Color image processing using vector rank filters",* International Conf. on Digital Signal Processing, Vol.2, 1995, pp.614-619.

[19] S. Schulte, V. De Witte, M. Nachtegael, D. Van der Weken, and E.E. Kerre, *"Fuzzy Two-Step Filter for Impulse Noise Reduction From Color Images",* IEEE Trans. on Image Processing, Vol.15, No.11, November, 2006, pp.3567-3578.

[20] Om Prakash Verma, Madasu Handmandlu, Anil Singh Parihar, Vamsi Krishna Madasu, *"Fuzzy Filters for Noise Reduction in Color Images"*, ICGST-GVIP Journal, Volume 9, Issue 5, ISSN: 1687-398X., September, 2009, pp.29-43.

[21] K.K.V. Toh, N.A.M. Isa, *"Cluster-based adaptive fuzzy switching median filter for universal impulse noise reduction"*, IEEE Transactions on Consumer Electronics, Volume 56, Issue 4, 2010, pp.2560-2568.

[22] Madhu S. Nair, G. Raju, *"A new fuzzy-based decision algorithm for high-density impulse noise removal"*, Signal, Image and Video Processing, 2010 Springer-Verlag. DOI:0.1007/s11760-010-0186-4.

[23] Aborisade, D.O., *"A Novel Fuzzy Logic Based Impulse Noise Filtering Technique",* International Journal of Advanced Science and Technology, Vol.32, July, 2011, pp.79-88.

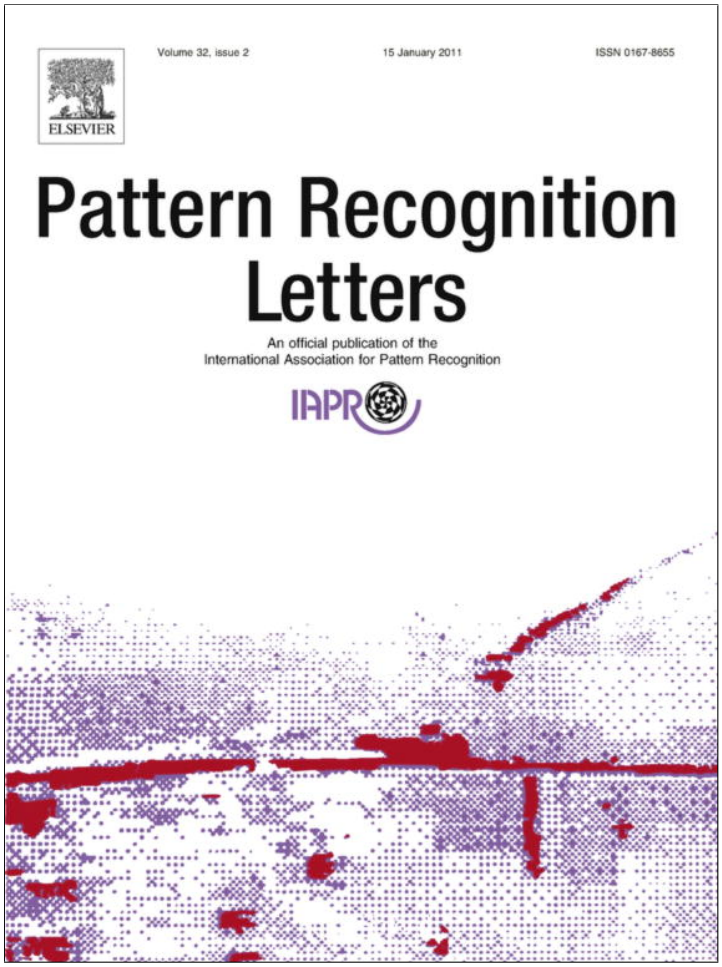
[24] Werner Backhaus, Reinhold Kliegl, *"Color Vision: Perspective from Different Disciplines"*, John Simon Werner, 1998.
[25] M. Tkalcic and J. F. Tasic, *"Color spaces: Perceptual, historical and application background",* in Proceeding s of IEEE EUROCON, Vol.1, September, 2003, pp.304-308.
[26] Recep Demirci, *"Rule-based automatic segmentation of color images",* AEU-International Journal of Electronics and Communications, Volume 60, Issue 6, 2006, pp 435-442.
[27] R. Demirci , Ugur Güvenç, "*Fuzzy filter for color images",* 1st international fuzzy systems Symposium, Ankara, October, 2009, pp.365-370.

**Om Prakash Verma**

Om Prakash Verma received his B.E. degree in Electronics and Communication Engineering from Malaviya National Institute of Technology, Jaipur, India, M. Tech. degree in Communication and Radar Engineering from Indian Institute of Technology (IIT), Delhi, India and Ph.D. degree from Delhi University. From 1992 to 1998, he worked at Department of Electronics & Communication Engineering at Malaviya National Institute of Technology, Jaipur, India as an Assistant Professor. He joined Department of Electronics & Communication Engineering, Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India, as an Associate Professor in 1998. Presently, he is Professor and Head, Department of Information Technology at Delhi Technological University. He has authored a book on Digital Signal Processing in 2003. His research interests include Computer Vision and Image Processing, Application of Soft Computing techniques in Image Processing, Artificial Intelligence, Optimization techniques, and Digital Signal Processing. He has published 35 research papers in International Journal and conference proceedings. He has guided 30 M. Tech. students, and presently 5 Ph.D. scholars are working under his supervision. He is Principal investigator of the "Information Security Education Awareness" Project, sponsored by the Department of Information Technology, Ministry of MHRD, Government of India.

**Shweta Singh**

Shweta Singh received her B.Tech. degree in Computer Science and Engineering from Indira Gandhi Institute of Technology, Guru Gobind Singh Indrapastha University, Delhi, India in 2009 and M. Tech. degree in Computer Technology and Applications from Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India in 2012. She joined the Department of Information Technology, Delhi Technological University in 2009 as an Assistant Professor and worked there for 2 years. Her areas of research include Computer Vision, Image Processing and Game theoretic optimization, and watermarking techniques.




(This is a sample cover image for this issue. The actual cover is not yet available at this time.)

Pattern Recognition Letters 32 (2011) 1187–1196

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

# A novel bacterial foraging technique for edge detection

Om Prakash Verma [a,*], Madasu Hanmandlu [b], Puneet Kumar [c], Sidharth Chhabra [a], Akhil Jindal [a]

[a] Delhi Technological University (Formerly Delhi College of Engineering), Delhi, India
[b] Department of Electrical Engineering, IIT Delhi, Delhi, India
[c] Advanced Systems Laboratory, Hyderabad, India

ARTICLE INFO

*Article history:*
Received 11 February 2010
Available online 12 March 2011
Communicated by H.H.S. Ip

*Keywords:*
Ant Colony System
Bacterial foraging
Derivative
Direction probability matrix
Edge detection

ABSTRACT

A new approach for edge detection using a combination of bacterial foraging algorithm (BFA) and probabilistic derivative technique derived from Ant Colony Systems, is presented in this paper. The foraging behavior of some species of bacteria like *Escherichia coli* can be hypothetically modeled as an optimization process. A group of bacteria search for nutrients in a way that maximizes the energy obtained per unit time spent during the foraging. The proposed approach aims at driving the bacteria through the edge pixels. The direction of movement of the bacteria is found using a direction probability matrix, computed using derivatives along the possible directions. Rules defining the derivatives are devised to ensure that the variation of intensity due to noise is discarded. Quantitative analysis of the feasibility of the proposed approach and its comparison with other standard edge detection operators in terms of kappa and entropy are given. The effect of initial values of parameters of BFA on the edge detection is discussed.



## 1. Introduction

Detection of edges in an image is a very important step for the image understanding. Indeed, high-level processing tasks such as image segmentation and object recognition, etc. directly depend on the quality of the edges detected. Moreover, the generation of an accurate edge map becomes a very critical issue when the images are corrupted by noise.

Edges in an image are marked with discontinuity or significant variation in intensity or gray levels. The methods of identifying the intensity discontinuity associated with edges in an image are normally based on the calculation of the intensity gradient in the whole image. The underlying idea of most edge detection techniques is the computation of a local first or second derivative operator, followed by some regularization technique to reduce the effects of noise.

Canny' method (Canny, 1986) for the edge detection counters the noise problems, wherein an image is convolved with the first-order derivatives of Gaussian filter for smoothing in the local gradient direction followed by the edge detection using thresholding. Marr and Hildreth (1980) propose an algorithm that finds edges at the zero-crossings of the image Laplacian. Non-linear filtering techniques for edge detection also have witnessed much advancement through the SUSAN method (Smith and Brady, 1997), which works by associating a small area of neighboring pixels with similar brightness to each center pixel. Existing edge detectors like Gradient operator and the Laplacian Operator are based on the assumption that edges in the images are step functions in intensity. Prewitt detector (Raman and Sobel, 2006) uses the local gradient operators which only detect edges having certain orientations and perform poorly on blurred or noisy images. Different algorithms for the fuzzy based edge detection are proposed in Cheung and Chan (1995), Kuo et al. (1997), Russo (1998), El-Khamy and S. (2000). Abdallah and Ayman (2009) introduce a fuzzy logic reasoning strategy for the edge detection in the digital images without determining a threshold.

Most of the existing operators are confronted with a huge search space for the detection of edges. Considering an image of size 1024 by 1024 pixels, the required solution space is of the order of $2^{1024*1024}$. Therefore the task of edge detection is time consuming and memory exhausting without the optimization.

A novel bacterial-derivative based algorithm that exploits the foraging behavior of bacteria to collectively detect edges in an image is developed in Passino (2002), Liu and Passino (2002). Bacterial foraging optimization algorithm (BFAO) has already been applied in the optimal control engineering, harmonic estimation (Mishra, 2005), transmission loss reduction (Tripathy et al., 2006), machine learning (Kim and Cho, 2005), active power filter design (Mishra and Bhende, 2007), color image enhancement (Hanmandlu et al., 2009), etc.

Bacteria foraging is an optimization process where bacteria seek to maximize the energy by eating up as many nutrients as they can. Nutrients in our case correspond to tracing the edge pixels. Bacteria move by either tumbling or swimming. In the classical approach, the direction of movement is decided randomly while

* Corresponding author. Tel.: +91 1127294672; fax: +91 1127871023.
*E-mail addresses:* opverma@dce.ac.in, opverma.dce@gmail.com (O.P. Verma).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.03.008

tumbling and every direction is equally preferred. In our method, a probabilistic approach, inherited from Ant Colony System (Dorigo and Gambardella, 1997; Verma et al., 2009), is used. This causes the bacterium to move along the direction, where the probability of finding a nutrient (edge) is highest. The proposed algorithm also distinguishes between the local variations due to noise and image structures, using a derivative. Another important characteristic of bacteria's life cycle is swarming. In our approach, swarming is not only dependent upon other bacteria's positions but also on the clique of its current position.

The proposed approach is well suited to address the uncertainty introduced while extracting the edge information from the image data. The innovative aspect of this approach lies in the development of a noise-protected operator that combines the rules framed for the noise cancellation and edge detection in the same structure. The application of modified bacterial foraging in combination with a derivative approach leads to a minimal set of input data to be processed thus making the process faster and memory-efficient. As a result, the proposed approach outperforms the existing state-of-the-art techniques.

The paper is organized as follows: In Section 2 a brief introduction to bacterial foraging technique is provided to present the basic idea. A modification of this technique is discussed in Section 3. Section 4 presents the probabilistic derivative approach to find the direction of movement of a bacterium. An algorithm for the edge detection along with the pseudo code is developed in Section 5. Section 6 gives the experimental results followed by the conclusions in Section 7.

## 2. Bacterial foraging technique

A new evolutionary technique, called bacterial foraging scheme appeared in Passino (2002), Liu and Passino (2002). Foraging can be modeled as an optimization process where bacteria seek to maximize the energy obtained per unit time spent during foraging. In this scheme, an objective function is posed as the effort or a cost incurred by the bacteria in search of food. A set of bacteria tries to reach an optimum cost by following four stages: Chemo taxis, swarming, reproduction, and elimination and dispersal. To start with, there will be as many solutions as the number of bacteria. So, each bacterium produces a solution iteratively for a set of optimal values of parameters. Gradually all the bacteria converge to the global optimum.

In the chemo taxis stage, the bacteria either resort to a tumble followed by a tumble or make a tumble followed by a run or swim. This is the movement stage of bacteria accomplished through swimming and tumbling. On the other hand, in swarming, each *Escherichia* coli bacterium signals another bacterium via attractants to swarm together. This is basically the cell to cell signaling stage. Furthermore, in the reproduction the least healthy bacteria die and of the healthiest, each bacterium splits into two bacteria, which are placed at the same location. While in the elimination and dispersal stage, any bacterium from the total set can be either eliminated or dispersed to a random location during the optimization. This stage prevents the bacteria from attaining the local optimum.

Let $\theta$ be the position of a bacterium and $J(\theta)$ be the value of the objective function, then the conditions $J(\theta) < 0$, $J(\theta) = 0$, and $J(\theta) > 0$ indicate whether the bacterium at location $\theta$ is in nutrient-rich, neutral, and noxious environments, respectively. Basically, chemo taxis is a foraging behavior that implements a type of optimization where bacteria try to climb up the nutrient concentration (find the lower values of $J(\theta)$), avoid noxious substances, and search for ways out of neutral media (avoid being at positions $\theta$ where $J(\theta) \geqslant 0$). This is just like a type of biased random walk.

## 3. The modified bacterial foraging technique for edge detection

The original bacterial foraging (BF) technique Liu and Passino (2002) is now modified to make it suitable for the edge detection. The nutrient concentration at each position, i.e. the cost function is calculated using a derivative approach. The implications in lieu of modifications of the technique for the edge detection are furnished here.

### 3.1. Search space

The 2-dimensional search space for bacteria consists of the $x$ and $y$-coordinates (i.e. discrete values) of a pixel in an image. Being limited by the image dimensions, i.e. horizontal and vertical pixels of the image, the search space is finite.

### 3.2. Chemotaxis

This is a very important stage of BF. It decides the direction in which the bacterium should move. Depending upon the rotation of the flagella, each bacterium decides whether it should swim (move in a predefined direction) or tumble (move in an altogether different direction). Our goal is to let the bacterium search for the edge pixels in an image. Another important goal is to keep the bacterium away from the noisy pixels.

As we are dealing with 2D discrete values of coordinates in an image with 8-connectivity of the neighborhood pixels, the probable directions to move for a bacterium from a particular pixel are: E, W, N, S, NE, SE, NW, SW. Out of these eight directions the bacterium of interest has to decide one direction that lead to an edge pixel but not a noisy pixel. A probabilistic derivative approach is used to find out the direction (one out of the eight possible directions) most suitable to hit upon an edge and to cut off the directions leading to noise. This is a major deviation in the chemotaxis step of BFA where the bacteria either tumble in a random direction or swim in the same direction as in the previous step. This is elaborated here.

Let $\theta^i(j,k,l)$ represent the position of the $i$th bacterium at $j$th chemotactic, kth reproductive and $l$th elimination-dispersal step. A pixel position in an image can be represented by the $x$ and $y$ -co-ordinates of the bacterium. So let $\theta^i(j,k,l)$ be the position of a bacterium in an array $\varphi[m,n,i,j,k,l]$ where $m$, $n$ represent the $x$–$y$ coordinates of the bacterium.

The initial positions of the bacteria are selected randomly. They move on to the edge pixels during the run of the algorithm. The path is recorded only after a certain number of steps made by each bacterium.

The movement of the bacterium may be represented by

$$\theta^i(j+1,k,l) = \phi[m',n',i,j+1,k,l], \tag{1}$$

where $m'$, $n'$ are the coordinates to which the bacterium should move in order to reach the edge pixel by avoiding the noisy pixels. During the tumble, the direction is determined from

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + \frac{C(i)\cdot\Delta(i)}{\sqrt{\Delta(i)^T\Delta(i)}} \tag{2}$$

where $\Delta(i)$ indicates a random number in $\mathscr{R}^2$ and $C(i)$ is the length of a step size.

The above approach is modified to include a probabilistic derivative as explained in detail in Section 4.

### 3.3. Swarming

It is assumed that a bacterium relies on other bacteria. This property of bacterium is exploited here. In this step, the bacterium that has searched an optimum path, signals other bacteria so that

they can together reach the desired optimum path swiftly. As each bacterium moves, it releases an attractant to signal other bacteria to swarm towards it. Also, each bacterium releases a repellent to warn other bacteria by keeping a safe distance from them. Because of this, bacteria congregate into groups and move in a concentric pattern. This is achieved by using a cell to cell signaling function which combines both the attraction and repelling effects. Thus we have

$$\begin{aligned} J_{cc}(\theta^i(j,k,l),\theta(j,k,l)) &= \sum_{t=1}^{s} j_{cc}^{t}(\theta^i,\theta^t) \\ &= \sum_{t=1}^{s}\left[-d_{att}\exp\left(-w_{att}\sum_{m=1}^{P}\left(\theta_m^i-\theta_m^t\right)^2\right)\right] \\ &\quad+\sum_{t=1}^{s}\left[h_{rep}\exp\left(-w_{rep}\sum_{m=1}^{P}\left(\theta_m^i-\theta_m^t\right)^2\right)\right] \end{aligned} \tag{3}$$

where $\theta^i$ is the location of $i$th bacterium, $P$ is the number of dimensions of the optimization domain (here, $P = 2$), $\theta = \{\theta^i | i = 1,\ldots,S\}$ represents the positions of the $i$th bacteria, $\theta_m^i$ is the $m$th component of $\theta^i$, $d_{att}$ is the measure of how much attractant is released, $w_{att}$ is the diffusion rate, $h_{rep}$ and $w_{rep}$ are the magnitude and width of the repelling effect. Empirically, $d_{att} = 0.1$, $w_{att} = 0.2$, $h_{rep} = 0.1$, $w_{rep} = 10$ are found to be optimal.

It may be noted from Eq. (3) that this is basically a function of distance of the bacterium under consideration from all other bacteria. This function is now modified to make it dependent on the clique of the position (pixel) of the bacterium and its distance from other bacteria.

$$J_{cc}(\theta^i(j,k,l),\theta(j,k,l)) = f(distance) + f(clique) \tag{4}$$

The clique is defined as the local group of pixels around the pixel of interest. It is represented as

$$f(clique) = \mu\Delta_{\theta^i} \tag{5}$$

where $\mu$ is a constant (i.e. 1). This is introduced to make the value of cell to cell signaling function negative as the bacterium reaches an edge pixel, and $\Delta_{\theta^i}$ is the function that uses the local intensity difference statistic at the pixel location given by $\theta^i \cdot \Delta_{\theta^i}$:

$$\Delta_{\theta^i} = \frac{V_c(I_{\theta^i})}{Z} \tag{6}$$

where, $I_{\theta^i}$ represents the pixel of interest in the image. If the pixel location is at $(m, n)$ then

$$\begin{aligned} V_c(I_{\theta^i}) = V_c(I_{m,n})&|I_{m-2,n-1} - I_{m+2,n+1}| + |I_{m-2,n+1} - I_{m+2,n-1}| \\ &+ |I_{m-1,n-2} - I_{m+1,n+2}| + |I_{m-1,n-1} - I_{m+1,n+1}| \\ &+ |I_{m-1,n} - I_{m+1,n}| + |I_{m-1,n+1} - I_{m-1,n-1}| \\ &+ |I_{m-1,n+1} - I_{m-1,n-2}| + |I_{m,n-1} - I_{m,n+1}| \end{aligned} \tag{7}$$

and $Z$ is the normalization factor defined as

$$Z = \sum_{m=1:M}\sum_{n=1:N} V_c(Im,n) \tag{8}$$

The modified form of cell to cell signaling for edge detection may be represented as

$$\begin{aligned} J_{cc}(\theta^i(j,k,l),\theta(j,k,l)) &= \sum_{t=1}^{s} j_{cc}^{t}(\theta^i,\theta^t) \\ &= \sum_{t=1}^{s}\left[-d_{att}\exp\left(-w_{att}\sum_{m=1}^{P}\left(\theta_m^i-\theta_m^t\right)\right)\right] \\ &\quad+\sum_{t=1}^{s}\left[h_{rep}\exp\left(-w_{rep}\sum_{m=1}^{P}\left(\theta_m^i-\theta_m^t\right)^2\right)\right] \\ &\quad-\mu\frac{V_c(I_{\theta^i})}{Z} \end{aligned} \tag{9}$$

$J_{cc}(\theta^i(j,k,l), \theta(j,k,l))$ has its initial value the derivative at the pixel given by $\theta^i(j,k,l)$. The calculation of the derivative at a pixel $(x, y)$ is explained in Section 4.1.

### 3.4. Reproduction step

After $N_c$ chemotactic steps, a reproduction step begins its loop. Let $N_{re}$ be the number of reproduction steps. For convenience, we assume that the number of bacteria $S$ is a positive even integer. We choose

$$S_r = \frac{S}{2} \tag{10}$$

as the number of population members having sufficient nutrients so that they will reproduce (split up into two) with no mutations. For reproduction, the population is sorted in the ascending order of the accumulated cost (the higher cost indicates that a bacterium has not got as many nutrients during its lifetime of foraging and hence is not "healthy" and thus unlikely to reproduce); then the least healthy bacteria $S_r$ die and each of the other healthiest bacteria $S_r$ split up into two, which are placed at the same location. This method rewards bacteria that have encountered a lot of nutrients and allows us maintain a constant population size.

### 3.5. E. Elimination step

Let $N_{ed}$ be the number of elimination-dispersal events and each bacterium in the population is subjected to elimination-dispersal with a probability $P_{ed}$. We assume that the frequency of chemotactic steps is greater than the frequency of reproduction steps, which is in turn greater than the frequency of elimination-dispersal events (e.g. a bacterium will take many chemotactic steps before reproduction, and several generations may take place before an elimination-dispersal event).

## 4. Finding the direction of movement using probablistic derivative approach

### 4.1. Computing derivative value for a pixel

The procedure to find the nutrient concentration as well as the suitable direction in which the bacterium should move in order to find the edge pixels is now explained. The common feature between an erroneous pixel and a real edge pixel is that the intensity difference around both the pixels is high. It is very important to differentiate them properly in case the image is noisy; hence a derivative based approach is used. This restrains the bacterium from moving towards the noisy pixels.

Consider the neighborhood of a pixel $(x, y)$ in Fig. 1(a). We consider positive $X$-axis as vertical downward and positive $Y$-axis as horizontal right side. A derivative at the central pixel position $(x, y)$ in the direction $D$ (D = {NW, W, SW, S, SE, E, NE, N}) is defined as the difference between the pixel of interest and its neighbor in the corresponding direction.

For example, the derivative at a pixel $(x, y)$ in the north-west direction is given by

$$\partial_{(x,y)}^{NW} = I(x-1,y-1) - I(x,y) \tag{11}$$

where $I(x, y)$ is the intensity at pixel $(x, y)$.

The choice of the derivative is made from the intensities of pixels. Consider an edge passing through the pixel $(x, y)$ in the NE-SW direction as in Fig. 1(b). Since it's an edge it should contain pixels at NE and SW positions as its constituents. Therefore, the derivative

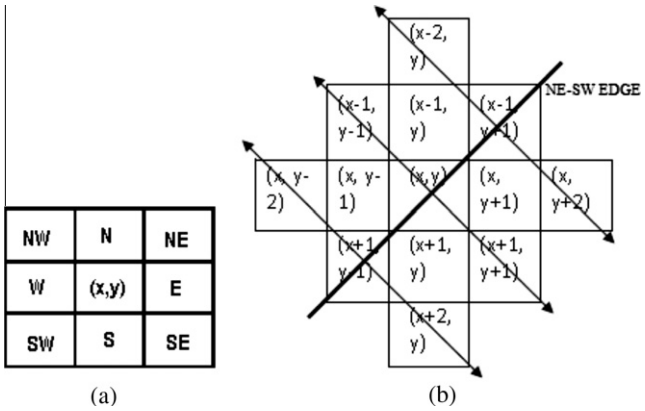



**Fig. 1.** (a) Pixel $(x, y)$ with its 8-connectivity (b) The pixels to be considered for the edge in NE-SW direction.

values in the direction NW and SE (perpendicular to the edge) for the pixels at positions $(x, y)$, $(x + 1, y - 1)$ and $(x - 1, y + 1)$ should also be high (Verma et al., 2009). This greatly reduces the chances of selecting a noisy pixel. The number of noisy pixels is negligible as compared to the number of edge pixels. Thus, for the pixel $(x, y)$ let $\partial_1$ be given by

$$\partial_1 = \partial^{NW}_{(x,y)}. \tag{12}$$

Similarly, we have at the neighboring pixels as

$$\partial_2 = \partial^{NW}_{(x+1,y-1)} = I(x+1, y-1) - I(x, y-2) \tag{13}$$

$$\partial_3 = \partial^{NW}_{(x-1,y+1)} = I(x-1, y+1) - I(x-2, y) \tag{14}$$

The average of the above three values can be taken as the net derivative value in NW direction, given by

$$\partial_{avg1} = \frac{(\partial_1 + \partial_2 + \partial_3)}{3} \tag{15}$$

Similarly, we calculate $\partial_{avg2}$ as the average of $\partial^{SE}_{(x,y)}$, $\partial^{SE}_{(x+1,y-1)}$ and $\partial^{SE}_{(x-1,y+1)}$. Having these two net derivatives we are looking for an edge having the intensity difference in one region only. The maximum of these two is the derivative for the edge pixel at $(x, y)$ in the direction NE–SW, denoted by

$$\partial_{NE-SW} = \max(\partial_{avg1}, \partial_{avg2}) \tag{16}$$

Similarly the derivatives along the remaining three directions are: $\partial_{NW-SE}$, $\partial_{N-S}$ and $\partial_{W-E}$.

The final derivative at pixel $(x, y)$ which is the maximum of the net derivatives obtained in all the four possible edge directions is:

$$\partial_{x,y} = \max(\partial_{W-E}, \partial_{NW-SE}, \partial_{NE-SW}, \partial_{N-S}) \tag{17}$$

Since our aim is to find the edge pixels, use is made of the observation that larger the value of $\partial_{x,y}$ more is the chance for a bacterium to reach an edge pixel. Therefore, the nutrient concentration at any position, $(x, y)$ should be the function of $\partial_{x,y}$. So we have

$$[\eta_i] = \partial_{(x,y)} \tag{18}$$

Hence, higher the nutrient concentration along an edge more is the movement of bacteria along it.

### 4.2. Direction probability matrix

The concern here is to locate an edge pixel to which a bacterium should move from a given pixel. To do this we find the values of the derivatives for all the 8 neighborhood pixels around the pixel under consideration in terms of their nutrient concentration. Out of the eight possible directions the next direction is found out using the transition matrix derived from the Ant Colony System (Dorigo and Gambardella, 1997; Verma et al., 2009).

In an artificial Ant Colony System, developed by imitating the real ant's behavior, ant chooses its next step for its movement depending upon the transition probability matrix which is a function of amount of pheromone discharged by ants on the path and the heuristic factor. The transition probability matrix at position $i$ is given by:

$$\rho_{ij} = \left\{ \frac{([\tau_j(t)]^{\alpha})([N_j]^{\beta})}{\sum_{j \in allowed_j}(([\tau_j(t)]^{\alpha})([N_j]^{\beta}))} \right\} \tag{19}$$

where, $N_j$ is the heuristic factor, $\tau_j(t)$ is the pheromone concentration that gives the permissible directions in which an ant movement (in our case ant is replaced by bacterium) has to move; $\alpha$ and $\beta$ are two parameters that control the relative importance of the two main factors: $\tau_j(t)$ and $N_j$.

In the proposed approach, pheromone concentration and $\alpha$ and $\beta$ are taken to be unity. The heuristic factor $N_j$ is taken as the nutrient concentration value $\eta_j$. Also, from Eq. (18), nutrient concentration is high along the direction of an edge. Thus the probability of movement along the direction $i \to j$ is calculated from:

$$\rho_{ij} = \left\{ \frac{[\eta_j]}{\sum_{j \in allowed_j}([\eta_j])} \right\} \tag{20}$$

A random direction is selected for the movement, with $\rho_{ij}$ as the probability of selection of direction $i \to j$

$$\Delta(k) = rand(\rho_{ij} \forall j \in allowed_j) \tag{21}$$

where, $\Delta(k)$ is the direction vector for the $k$th bacterium at position $i$ and $allowed_j$ is the set of possible moves for the bacterium. This random direction gives the direction of movement out of the eight possible directions. Suppose the current location is $[x, y]$, step size is unity and the random direction selected is NW then the next location of bacteria would be $[x - 1, y - 1]$.

## 5. The algorithm and pseudo code

### 5.1. Algorithm

[Step 1] Initialize the parameters $n$, $S$, $N_C$, $N_{re}$, $N_{ed}$, $P_{ed}$, $C(i)$ $(i = 1, 2, \ldots, , S)$, $F(i)$, $\theta^i(1, 1, 1)$, $V_c(I)$, $J(i, 1, 1, 1)$
where
$n$: dimension of the search space(2),
$S$: the number of bacteria in the population,
$S_r$: bacteria split ratio,
$N_C$: chemotactic steps,
$N_{re}$: the number of reproduction steps,
$N_{ed}$: the number of elimination-dispersal events,
$N_S$: swim length,
$P_{ed}$: elimination-dispersal with probability,
$C(i)$: the size of the step taken in the direction specified by the tumble(unit),
$F(i)$: flag bit for each pixel indicating whether it has already been traversed or not,
$V_c(I)$: is the clique matrix for the image I.
$\theta^i(1, 1, 1)$: initial positions of the bacterium selected randomly.
$J(i, 1, 1, 1)$: Initialized derivative value at the pixel given by $\theta^i(1, 1, 1)$.

[Step 2] Elimination-dispersal loop: $l = l + 1$
[Step 3] Chemotaxis loop: $j = j + 1$

[a] For $i = 1, 2, \ldots, S$, take a chemotactic step for bacterium i as follows.
[b] Compute the fitness function, $J(i, j, k, l)$.

Let, $J(i, j, k, l) = J(i, j, k, l) + J_{cc}(\theta^i(j, k, l), \theta(j, k, l))$ (i.e. add on the cell-to cell attractant–repellant profile to simulate the swarming behavior)

[c] Tumble: Find the directions of possible movement from the derivative value and compute the direction probability matrix using:

$$\rho_{ij} = \left\{ \frac{[\eta_j]}{\sum_{j \in allowed_j}([\eta_j])} \right\}$$

Now select a random direction using Eq. (21) and find the next direction of movement.
[d] Reproduction loop:

For each possible direction, $k = k + 1$.

[e] Move: Let

$$\theta^i(j+1, k, l) = \theta^i(j, k, l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\Delta(i)}}$$

This results in a step of size $C(i)$ in the direction of the tumble for bacterium $i$. In our case Eq. (1) is used where $(m', n')$ is found out using the step size movement along the direction of tumble.
[f] Compute $J(i, j+1, k, l)$ and let $J(i, j, k, l) = J(i, j, k, l) + J_{cc}(\theta^i(j, k, l), \theta(j, k, l))$
[g] Swim
(i) Let $m = 0$ (counter for swim length).
(ii) While $m < Ns$ (if have not climbed down too long).
- Let $m = m + 1$.
- If $J(i, j+1, k, l) > J_{last}$

then update $\theta^i(j+1k, l)$ as done in step 3(e). Use this $\theta^i(j+1k, l)$ to compute the new $J(i, j+1, k, l)$ as in step 3[f]

- Else, let $m = m + 1$

[h] Go to next bacterium $(i+1)$ if $i \neq S$ (i.e., go to [b] to process the next bacterium).
[Step 4] If $j < N_c$, go to Step 3[e]. In this case, continue chemotaxis, since the life of bacteria is not over.
[Step 5] Reproduction:
[a] For each $i = 1, 2, \ldots, S$, let

$$J^i_{health} = \sum_{j=1}^{N_{c+1}} J(i, j, k, l)$$

be the health of the bacterium $i$ (a measure of how many nutrients it got over its lifetime and how successful it was at avoiding the noxious substances).

Sort bacteria and chemotactic parameters $C(i)$ in the ascending order of the cost $J_{health}$ (higher cost means lower health).[b] The $S_r$ bacteria with the highest $J_{health}$ values die and the remaining $S_r$ bacteria with the best values split (this process is performed by the copies that are placed at the same location as that of their parents).
[Step 6] If $k < N_{re}$, go to Step 3[e]. This means that the number of specified reproduction steps is not reached, so the next generation of the chemotactic loop is started.
[Step 7] Elimination-dispersal: For $i = 1, 2, \ldots, S$, eliminate and disperse each bacterium with probability $P_{ed}$, This results in the number of bacteria a constant. To do this, if a bacterium is eliminated, simply disperse one to a random location in the optimization domain. If $l < N$, then go to Step 2; otherwise end.

### 5.2. Pseudo code

Bacteria_Edge (Image I)
FOR each pixel in I

$$\begin{aligned} V_c(I_{m,n}) = {} & |I_{m-2,n-1} - I_{m+2,n+1}| + |I_{m-2,n+1} - I_{m+2,n-1}| \\ & + |I_{m-1,n-2} - I_{m+1,n+2}| + |I_{m-1,n-1} - I_{m+1,n+1}| \\ & + |I_{m-1,n} - I_{m+1,n}| + |I_{m-1,n+1} - I_{m-1,n-1}| \\ & + |I_{m-1,n+1} - I_{m-1,n-2}| + |I_{m,n-1} - I_{m,n+1}| \end{aligned}$$

```
FOR (each bacterium i = 1:S)
  θ^i(1, 1, 1) = rand_post( )
  J(i, 1, 1, 1) = derivative_value(θ^i(1, 1, 1))
  END FOR
FOR (elimination-dispersal loop l = 1:N_ed)
    FOR (reproduction-loop k = 1:N_re)
    FOR (chemotactic-loop j = 1:N_c)
    FOR (each bacterium i = 1:S)
      Calculate
      J(i, j, k, l) = J(i, j, k, l) + J_cc(θ^i(j, k, l), θ(j, k, l))
      Set J_last = J(i, j, k, l)
        Tumble:
      Find the direction of possible movement from the
  direction probability matrix.
        Move:
      θ^i(j + 1, k, l) = φ[m', n', i, j + 1, k, l]
      Compute J(i, j + 1, k, l)
      m = 0
      WHILE (m < N_s)
        m = m + 1
        IF (J(i, j + 1, k, l) < J_last)
          J_last = J(i, j + 1, k, l)
          Update θ(i, j + 1, k, l)
          Recalculate J(i, j + 1, k, l)
        ELSE
          m = N_s
        ENDIF
      END WHILE
    END FOR (Bacterium)
  END FOR (Chemotaxis)
Reproduction:
    For given k and l, and each bacterium i = 1, 2, ..., S
    Sum:
      J^i_health = Σ_{j=1}^{N_c+1} J(i, j, k, l)
    Sort:
      Sort bacteria and chemotactic parameters C(i) in order
  of ascending cost J_health.
    Split and Eliminate:
      The S_r bacteria with the highest J_health values die and the
  remaining S_r bacteria with the best values split.
    END FOR (Reproduction)
    Disperse:
      For i = 1, 2, ... , S, with probability P_ed, randomize a
  bacterium's position
    END FOR (Elimination and Dispersal)
END
```

The pixels visited by the bacterium are considered to be the desired edge pixels. Thus, the path traced out by bacteria gives the edges.

## 6. Results

An edge detector can be evaluated based on two parameters. First, its accuracy in determining the edge pixels, and second, it should provide useful information in the form of meaningful edges.

The accuracy is ascertained using Relative Grading technique (Bryant and Bouldin, 1979). In this, a majority image is found using the results of other five edge detectors: Canny, Edison, Rothwell, Sobel, and SUSAN. Then a pixel-by-pixel comparison of the output of the proposed method is made with the true image.

A pixel in the majority image is an edge pixel, if the majority of the methods claim to have an edge pixel in its neighborhood, with at least one centered on it. For example, Fig. 5(h), shows the majority image obtained from Fig. 5(b)–(f).

A majority image is obtained from methods $1, 2, \ldots, n$ as $M(method1, method2, \ldots, methodn)$.

The kappa (a measure of accuracy) (Cohen, 1960) for the pixel-to-pixel comparison between two images $I_1$ and $I_2$ is denoted by $k(I_1, I_2)$.

The information content of the output image is measured by using Shannon's entropy function (Shannon, 1948). It gives the indefiniteness in an image and is calculated from

$$H(I) = \sum_{-i=o}^{L} p_i log p_i \tag{22}$$

where, $I$ stand for the Image. $p_i$ is the frequency of pixels with intensity $i$. As we have binary levels a window of $3 \times 3$ centered at the pixel of concentration is considered as the intensity value.

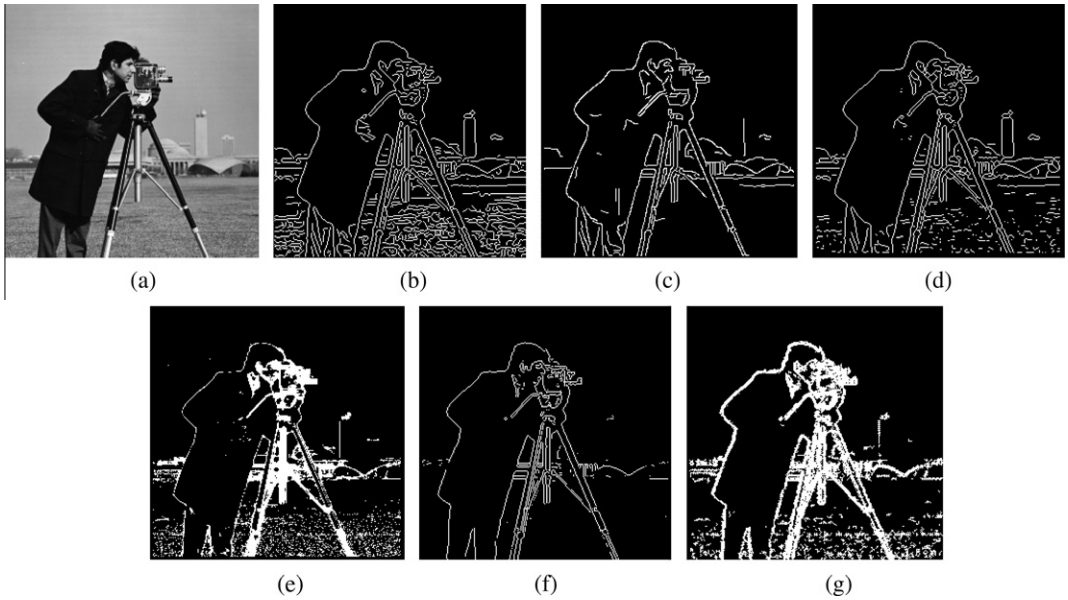



**Fig. 2.** (a) Original Cameraman image (b) Canny Edge Detector (c) Edison Edge Detector (d) Rothwell Edge Detector (e) SUSAN Edge Detector (f) Sobel Edge Detector and (g) The proposed approach.

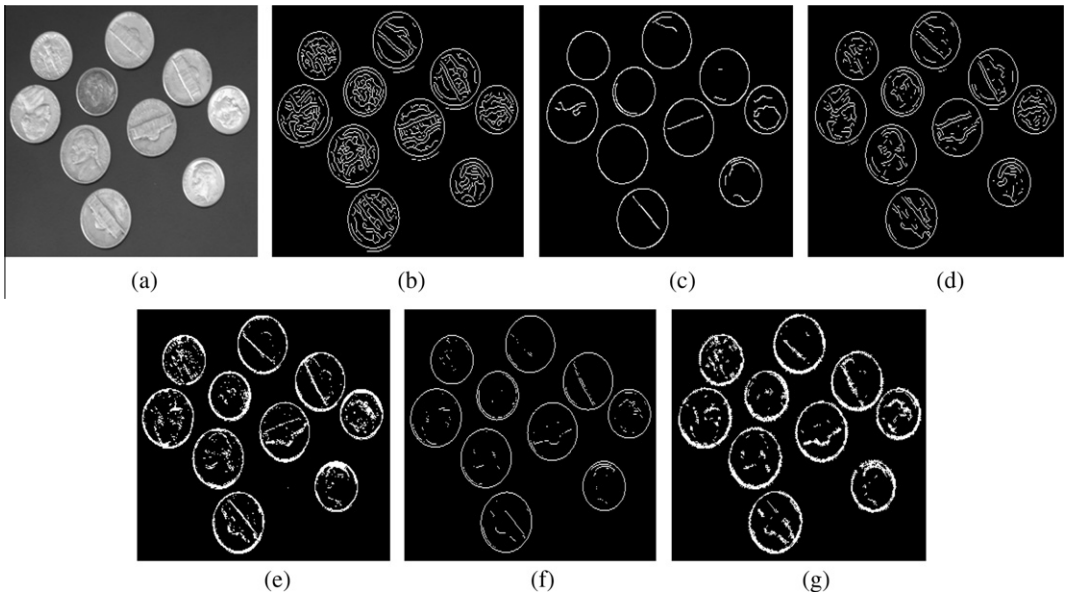



**Fig. 3.** (a) Original Coins image (b) Canny Edge Detector (c) Edison Edge Detector (d) Rothwell Edge Detector (e) SUSAN Edge Detector (f) Sobel Edge Detector and (g) The proposed approach.

### 6.1. Comparison with other techniques

The performance of the proposed technique is compared against that of the traditional edge detectors such as Canny, Edison, Rothwell, Sobel and SUSAN. The, traditional edge detectors are implemented using the MATLAB toolbox. The proposed method is also coded in the MATLAB. The parameters are taken as: $S = 100$, $S_r = 0.05S$, $N_s = 10$, $P_{ed} = 0.95$, $N_{ed} = 15$, $N_{re} = 1$, $N_c = 100$. The path traversed by a bacterium represents the edge pixels and is colored white on the black background.

For images in Figs. 2–6, the captions are as follows: (a) the original image, (b) the result of Canny Edge Detector, (c)the result of Edison Edge Detector, (d) the result of Rothwell Edge Detector, (e) the result of SUSAN Edge Detector, (f) the result of Sobel Edge Detector, and (g) the result of the proposed approach.

It is observed that the edges are accurately detected. But our method lacks in presenting the complete edges. This is because of restrictions imposed on the maximum swim length for a bacterium. Also, thick edges can be seen due to bacteria moving parallel to an edge.

Table 1 shows the kappa values in a comparison of several edge detectors. The column 2 of Table 1 shows $k(M(C, E, R, So, Su), P)$, where, $C$ stands for Canny, $E$ for Edison, $R$ for Rothwell, $So$ for Sobel, $Su$ for SUSAN, and $P$ for the proposed method. It may be noted that the values of kappa around 0.5 indicate the poor performance.

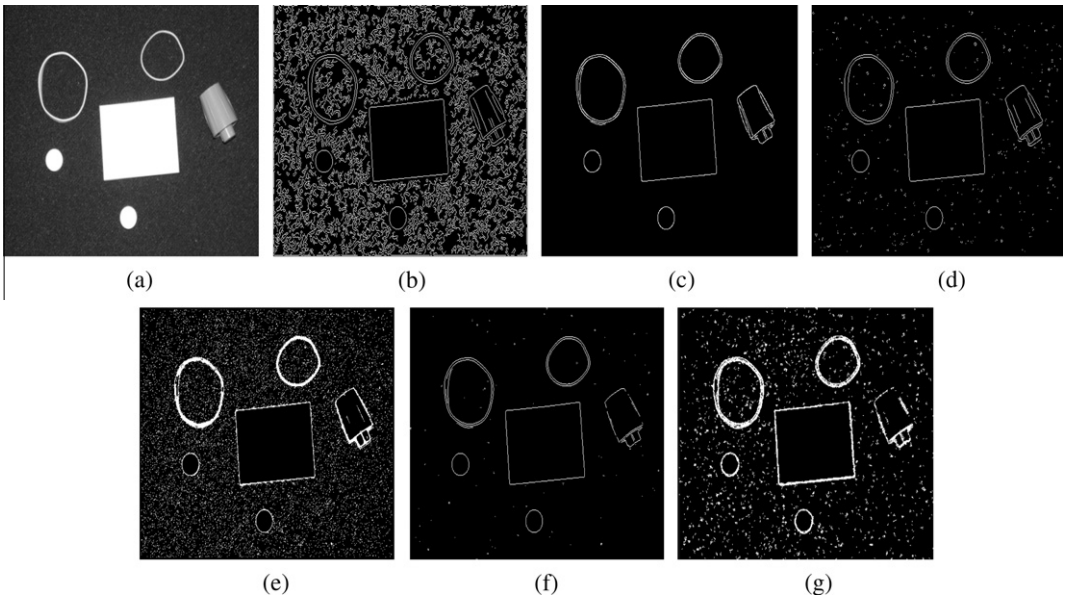



**Fig. 4.** (a) Original Pillsetc image (b) Canny Edge Detector (c) Edison Edge Detector (d) Rothwell Edge Detector (e) SUSAN Edge Detector (f) Sobel Edge Detector and (g) The proposed approach.

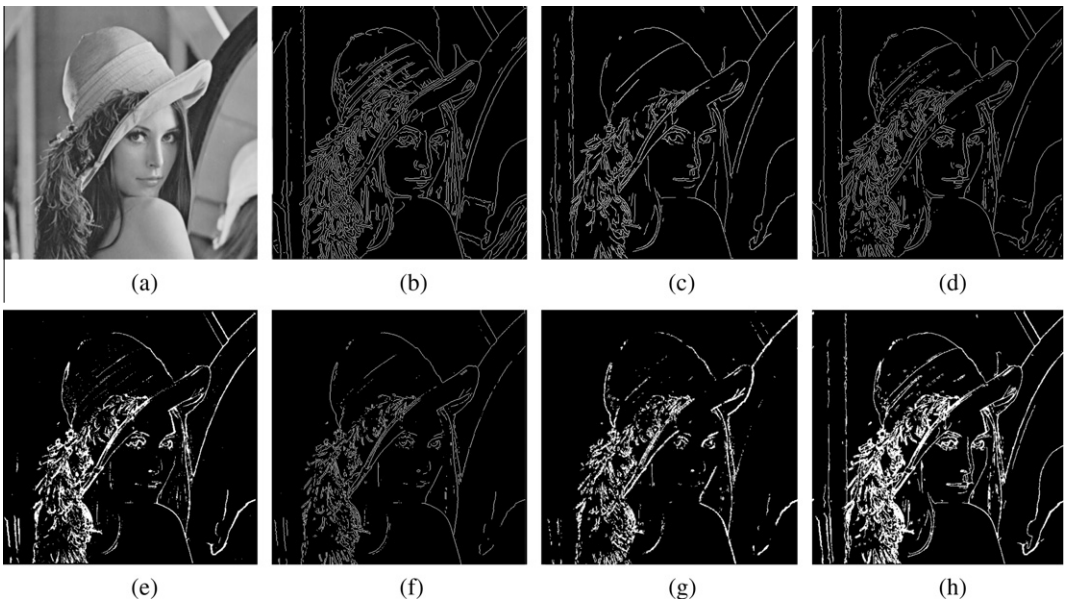



**Fig. 5.** (a) Original Lena image (b) Canny Edge Detector (c) Edison Edge Detector (d) Rothwell Edge Detector (e) SUSAN Edge Detector (f) Sobel Edge Detector (g) The proposed approach and (h) Majority image obtained using (b) to (f).

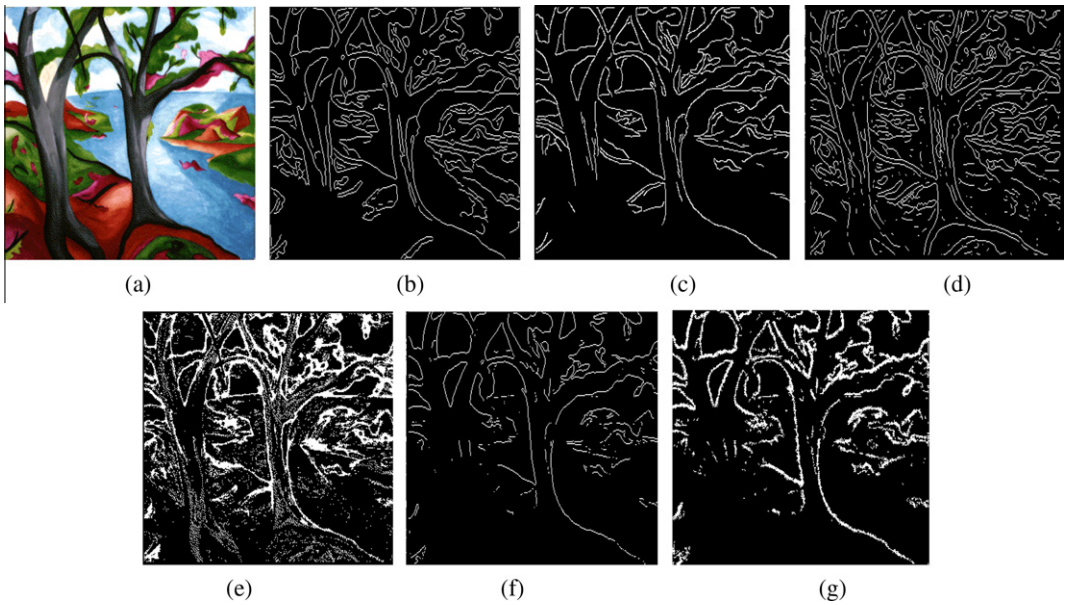



**Fig. 6.** (a) Original Trees image (b) Canny Edge Detector (c) Edison Edge Detector (d) Rothwell Edge Detector (e) SUSAN Edge Detector (f) Sobel Edge Detector and (g) The proposed approach.

**Table 1**
Kappa values for comparison with majority image. Column 2 Kappa for proposed edge detector's output comparison with $M(C, E, R, So, Su)$, Columns 3–7: A comparison of Kappa for the ratio, standard edge detector/the proposed edge detector with majority image from other detectors.

| Image | Majority | Sobel/Prop. | Canny/Prop. | Edison/Prop. | Rothwell/Prop. | SUSAN/Prop. |
|---|---|---|---|---|---|---|
| Trees | 0.4381 | 0.2802/0.4357 | 0.3751/0.5103 | 0.4243/0.4546 | 0.3349/0.5234 | 0.3771/0.4915 |
| Lena | 0.4594 | 0.3994/0.4567 | 0.3831/0.5163 | 0.4658/0.4996 | 0.4491/0.5172 | 0.4622/0.3884 |
| Pillsetc | 0.4792 | 0.3749/0.4744 | 0.0785/0.4967 | 0.4133/0.4755 | 0.4727/0.5016 | 0.2541/0.3505 |
| Coins | 0.5433 | 0.3878/0.5406 | 0.3577/0.6098 | 0.43/0.549 | 0.4504/0.6108 | 0.4835/0.4806 |
| Cameraman | 0.5953 | 0.3581/0.5895 | 0.3474/0.6272 | 0.3855/0.61 | 0.433/0.6309 | 0.4252/0.4198 |

The column 3 of Table 1 shows the ratio between M of *So* and M of *P*.

$$k(M(C,E,R,So,Su),So)/k(M(C,E,R,So,Su),P)$$

Similarly, the ratios due to other standard edge detectors are given in other columns.

It may be observed from Table 1 that the proposed method outperforms 4 out of 5 methods in all images. The results of the proposed method are poor with respect to Susan edge detector in three images viz. Lena, Coin and Camera man.

Table 2 shows the entropy values for the outputs of different methods on various images. A high entropy value signifies more randomness and less information. Sobel edge detector has the least value of entropy for all images and can be seen to have most appropriate edges. Edison edge detector also performs well. The edge detectors of Rothwell and Canny have a comparable performance over the proposed method but SUSAN edge detector performs poorer than all. It is evident from the results that our method finds meaningful edges in most images but is partially successful in curbing noise as shown in Fig. 4(g).

**Table 2**
Entropy values of different edge detector for multiple images.

| Image | Edison | SUSAN | Rothwell | Canny | Sobel | Proposed |
|---|---|---|---|---|---|---|
| Trees | 0.8682 | 1.7299 | 1.0936 | 0.9109 | 0.5791 | 0.8682 |
| Lena | 0.6777 | 0.7928 | 0.7438 | 0.8848 | 0.5303 | 0.7146 |
| Pillsetc | 0.2369 | 1.1692 | 0.3332 | 1.421 | 0.2265 | 0.7903 |
| Coins | 0.4992 | 0.8759 | 0.7201 | 0.9201 | 0.4821 | 0.86 |
| Cameraman | 0.6852 | 1.1499 | 0.8015 | 0.9931 | 0.5633 | 1.2212 |

### 6.2. Effects of parameter variation

We now discuss the effects of varying different parameters of the proposed method namely: $S$, $Sr$, $N_s$, $P_{ed}$, $N_{ed}$, $N_{re}$ and $N_c$ on the resultant image of cameraman.

The variation of parameters is judged by two measures: entropy and kappa. To calculate kappa, the output image is compared with the majority image obtained from other 5 methods as explained above. Moreover, it is well known that an optimum value of parameter is the one with less entropy and high kappa.

The results are of edge detection as shown in Fig. 7(a)–(g). We find that $N_{re}$ and $N_c$ have no considerable effect on the results. Both kappa and entropy remain constant. $N_s$, the swim length causes both entropy and kappa to decrease as it is increased. Decrease in kappa is gradual than the decrease in entropy. Change in bacteria split ratio ($S_r$) has no effect on kappa but entropy is significantly varied as can be seen in Figs. 7(b) and 8. In Fig. 7(b) we have taken different values of $S_r$ from $0.1S$ to $0.5S$, where $S$ is the number of becteria. Change in $N_{ed}$ causes entropy to drop after $N_{ed} = 10$. Though, kappa also drops but not significantly whereas in Fig. 7(a), we find that there is no change in entropy with change in number of bacteria ($S$) but kappa increases (Fig. 9(a) and (b)). It may also be noted that though increase in $S$ is favorable here but it adds more burden on computing resources. Thus, an optimal

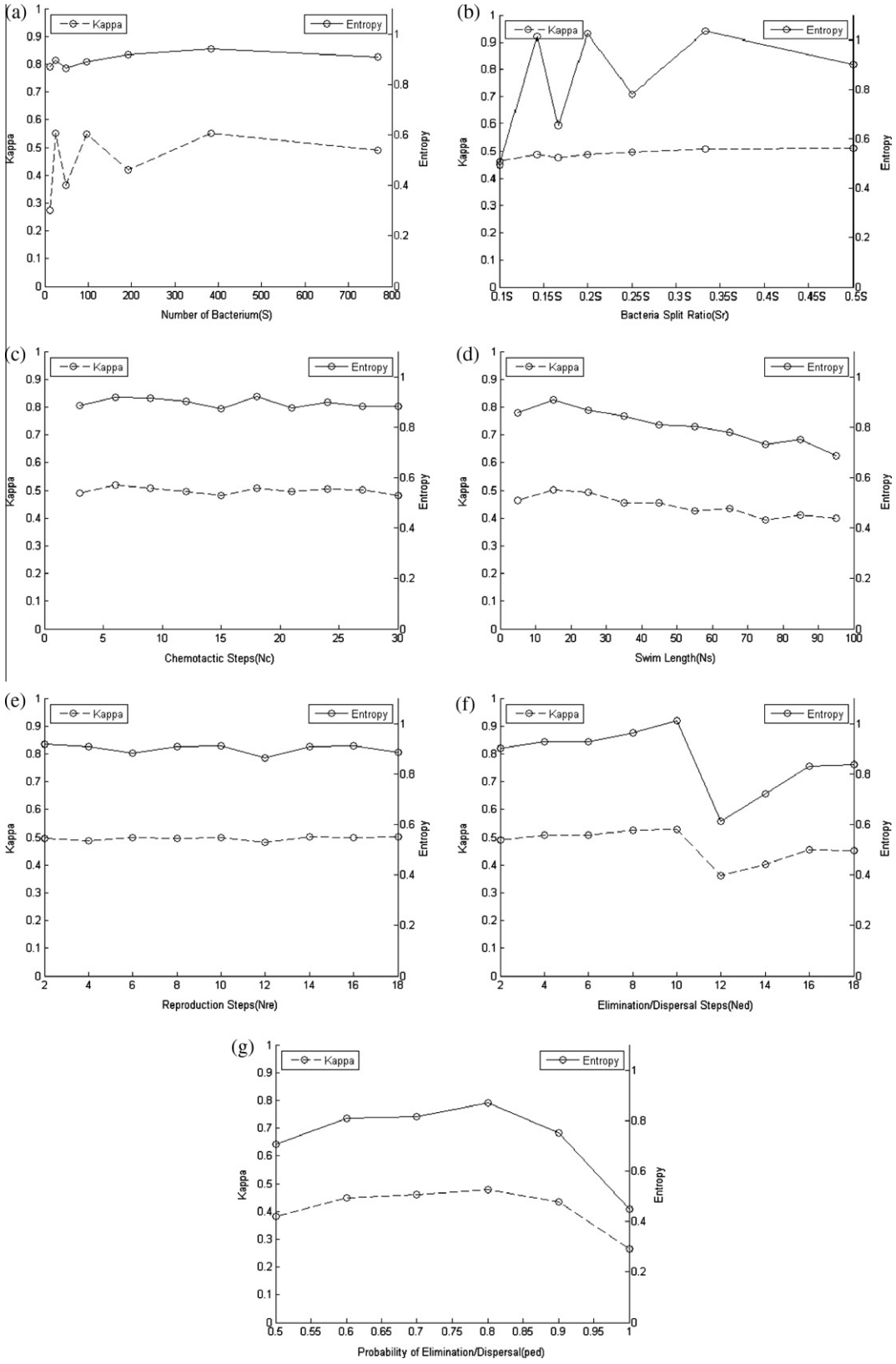



**Fig. 7.** Plot between Kappa and Entropy v/s initial values parameters (a) $S$ (b) $S_r$ (c) $N_c$ (d) $N_s$ (e) $N_{re}$ (f) $N_{ed}$ (g) $P_{ed}$.

trade-off has to be found between performance and resources. In Fig. 7(g), we observe that the shape of plots of both kappa and entropy is a parabola facing downwards and centered on 0.8. It also validates that a value of $P_{ed}$ around 0.9 would be optimum.

## 7. Conclusion and future work

Edge detection is essential in many tasks of image processing. This study proposes a novel BF based approach for edge detection.

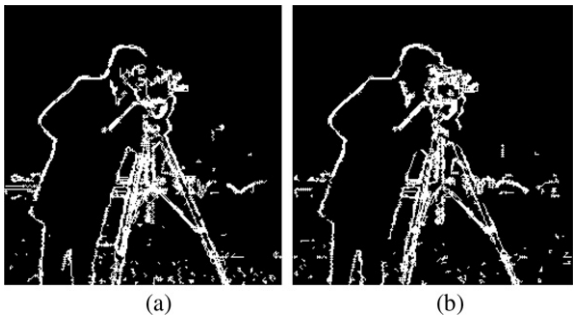



**Fig. 8.** Output images for the split ratio (a) $S_r = 0.25S$ (b) $S_r = 0.33S$.

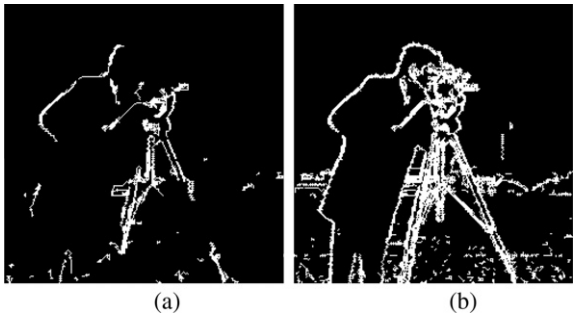



**Fig. 9.** Output images for the split ratio (a) $S = 12$ and (b) $S = 768$.

The proposed method finds robust edges even in the complex and noisy images. This work opens a new domain of research in the field of edge detection using bio-inspired algorithms.

The results of the proposed method are compared with those of other standard methods using kappa and entropy. Our method performs better than many other standard methods. The variation of several initial parameters on the output of the proposed edge detector is discussed. Their values are derived empirically. These values may also be found specifically for each image to gain maximum benefit. So a way to calculate the optimum values of all the parameters in less time needs to be investigated.

It is noted that our results show some disconnected edges. Since BFOA has been devised with the aim of finding global extremes, this error is expected. If a form of preferential treatment such that pixels connected to edge pixels get an advantage is introduced then this problem can be mitigated. Also, some form of repellant need to be added to the path already traced by bacteria so that parallel/double edges are not formed.

## References

Abdallah, A.A., Ayman, A.A., 2009. Edge detection in digital images using fuuzy logic techniques. World Academy of Sci. Eng. Technol. 51, 178–186.

Bryant, D.J., Bouldin, D.W., 1979. Evaluation of edge operators using Relative and absolute grading. In: Proc. IEEE Comput. Soc. Conf. Pattern Recognition and Image Processing, pp. 138–145.

Canny, J.F., 1986. A computational approach to edge detection. IEEE Trans. Pattern Anal. Machine Intell. 8 (6), 679–698.

Cheung, K., Chan, W., 1995. Fuzzy one –Mean algorithm for edge detection. IEEE Internat. Conf. Fuzzy Systems, 2039–2044.

Cohen, J., 1960. A coefficient of agreement for nominal scales. Educ. Psychol. Measure. 20 (1), 37–46.

Dorigo, M., Gambardella, L.M., 1997. Ant colony system: A cooperative learning approach to the traveling salesman problem. IEEE Trans. Evol. Comput., 73–81.

El-Khamy, S., El-Yamany, N., Lotfy, M., 2000. A modified fuzzy sobel edge detector. In: Seventeenth National Radio Science Conf. (NRSC'2000), February 22–24, Minufia, Egypt, 2000.

Hanmandlu, M., Verma, O.P., Kumar, N.K., Kulkarni, M., 2009. A novel optimal fuzzy system for color image enhancement using bacterial foraging. IEEE Trans. Instrum. Measure. 58 (8), 2867–2879.

Kim, D.H., Cho, C.H., 2005. Bacterial foraging based neural network fuzzy learning. IICAI, 2030–2036.

Kuo, Y., Lee, C., Liu, C., 1997. A new fuzzy edge detection method for image enhancement. IEEE Internat. Conf. Fuzzy Systems, 1069–1074.

Liu, Y., Passino, K.M., 2002. Biomimicry of social foraging bacteria for distributed optimization models, principles and emergent behaviors. J. Optim. Theory Appl. 115 (3), 603–628.

Marr, D., Hildreth, E.C., 1980. Theory of edge detection. Proc. Roy. Soc. London b207, 187–217.

Mishra, S., 2005. A hybrid least square-fuzzy bacterial foraging strategy for harmonic estimation. IEEE Trans. Evol. Comput. 9 (1), 61–73.

Mishra, S., Bhende, C.N., 2007. Bacterial foraging technique-based optimized active power filter for load compensation. IEEE Trans. Power Delivery 22 (1), 457–465.

Passino, K.M., 2002. Biomimicry of bacterial foraging for distributed optimization and control. Control Systems Magazine, IEEE 22 (3), 52–67.

Raman, Maini, Sobel, J.S., 2006. Performance evaluation of prewitt edge detector for noisy images. GVIP J. 6 (3).

Russo, F., 1998. Edge detection in noisy images using fuzzy reasoning. IEEE Trans. Instrum. Measure. 47 (5), 1102–1105.

Shannon, C.E., 1948. A mathematical theory of communication. Bell Syst. Tech. J. 27, 379–423. 623-656.

Smith, S.M., Brady, J.M., 1997. SUSAN – A new approach to low level image processing. Internat. J. Comput. Vision 23 (1), 45–78.

Tripathy, M., Mishra, S., Lai, L.L., Zhang, Q.P., 2006. Transmission loss reduction based on FACTS and bacteria foraging algorithm. PPSN, 222–231.

Verma, O.P., Hanmandlu, M., Kumar, P., Srivastava, S., 2009. A novel approach for edge detection using ant colony optimization and fuzzy derivative technique. Proc. IEEE, IACC, 1206–1212.

# A Wireless Sensor Network for Greenhouse Climate Control

*Using a wireless sensor network, the authors developed an online microclimate monitoring and control system for greenhouses. They field-tested the system in a greenhouse in Punjab, India, evaluating its measurement capabilities and network performance in real time.*

**Roop Pahuja**
*National Institute of Technology Jalandhar*

**H.K. Verma**
*Sharda University, India*

**Moin Uddin**
*Delhi Technological University*

Wireless sensor networks are an important pervasive computing technology invading our environment. Over the years, research into WSN technology has matured to the extent that the ZigBee, based on IEEE 802.15.4, has emerged as a communication and networking standard to cater to the unique needs of WSN. It's a low-power (a few μW), low-data-rate (250 kbps), fault-tolerant, easily scalable, short-range (100 m) wireless protocol for embedded electronic devices called *sensor nodes*.[1] ZigBee and ZigBee-like standard-based WSN products and systems are now available to suit a variety of applications, including environment monitoring, precision agriculture, home and building automation, healthcare, traffic management, and so on.[2–4] WSNs are also gaining importance in *controlled environmental agriculture* technology, especially in greenhouse horticulture, because they offer wireless and flexible installation and reliable operation. (See the "Related WSN Work in Greenhouse Horticulture" sidebar for more information.)

Motivated by the idea of integrating a WSN into a high-level programming language to develop a custom application of public interest, we developed and field-tested a novel system for greenhouse climate control. Our system provides microclimate monitoring of the greenhouse temperature and relative humidity and analyzes the greenhouse crops' vapor pressure deficit (VPD), an important climate parameter related to plant growth, health, and yield conditions. To enhance the monitoring network, we integrated into it a VPD-based MIMO (multiple input and multiple output) fuzzy climate controller and an RS-485 actuator network to automate the greenhouse climate-control operations. Here, we discuss the technical issues we faced in designing this greenhouse-specific WSN and in implementing its intelligent application software.

## Material and Methods

First, we had to gather greenhouse domain knowledge to gain a better understanding of greenhouse crops, work activities, operations, and research trends.[1,2]

In particular, we needed to understand the microenvironment. A greenhouse is a complex, multivariable interactive system. Because of local weather fluctuations, the plant-growing process and its interaction with internal climatic conditions, and the use of different climate control equipment, the greenhouse environment is highly dynamic and varies spatially, thus creating

## Related WSN Work in Greenhouse Horticulture

Carlos Serodio and his colleagues—part of a multidisciplinary research group in Portugal—designed and implemented a networked platform using a wireless sensor network (WSN), controller area network (CAN), and Internet and email tools. Their platform offers computerized agriculture management systems in a greenhouse to support distributed data acquisition and control, helping growers make better decisions in carrying out agricultural practices in a greenhouse.[1]

Zhou Yiming and his colleagues discussed the hardware and software design of a ZigBee WSN node with temperature, relative humidity, and moisture sensors and proposed a star or mesh network architecture for a greenhouse WSN system.[2] Hui Liu and his colleagues designed the Crossbow WSN for measuring temperature, light intensity, and soil moisture using a terminal interface for logging and displaying data, along with experimental testing of antenna heights that effect radio range.[3]

Teemu Ahonen and colleagues developed a node using the Sensinode sensor platform fitted with temperature, relative-humidity, and light-intensity sensors based on the 6LoWPAN protocol. They tested its feasibility by deploying a simple sensor network into a greenhouse in Western Finland.[4] One-day experiment data was collected to evaluate the network reliability and its ability to detect the microclimate layers.

Dae Heon Park and his colleagues developed a WSN-based greenhouse environmental monitoring and dew point control system that prevents dew condensation phenomena on the crop's surface, helping to prevent diseases and infections.[5]

### REFERENCES


1. C. Serodio et al., "A Network Platform for Agriculture Management System," *Computer and Electronics in Agriculture*, vol. 31, no.1, 2001, pp. 75–90.
2. Z. Yiming et al., "A Design of Greenhouse Monitoring and Control System Based on ZigBee WSN," *Proc. Int'l Conf. Wireless Comm. Networking and Mobile Computing*, WiCOM, 2007, pp. 2563–2567.
3. H. Liu, Z. Meng, and S. Cui, "A Wireless Sensor Network Prototype for Environmental Monitoring in Greenhouse," *Proc. Int'l Conf. Wireless Comm. Networking and Mobile Computing*, WiCOM, 2007, pp. 2344–2347.
4. T. Abhonen, R. Virrankoski, and M. Elmusrati, "Greenhouse Monitoring with Wireless Sensor Network," *Proc. IEEE/ASME Int'l Conf.: Mechtronics and Embedded Systems and Applications*, MESA, 2008, pp. 403–408.
5. D. Heon Park and J.-W. Park, "Wireless Sensor Network-Based Greenhouse Environment Monitoring and Automatic Control System for Dew Condensation Prevention," *Sensors (Basel)*, vol. 11, no. 4, 2011, pp. 3640–3651.


microclimate zones. Climatic conditions vary from the greenhouse center to its side walls, and from the canopy to the aerial levels.[5]

We also needed to automate various operations and decisions. We had to gather and analyze plant-related sensory data using proactive computing methodology[2] to obtain the actionable output decisions for the different plant-related operations driving the in-house actuating equipment (operations such as climate control, fertigation control, irrigation control, and integrated pest management).

Finally, we needed to provide online information. Sophisticated software must provide online connectivity to the greenhouse control system with GUIs to display microclimate data statistics, decision support, and control operations.[6]

The need for greenhouse microenvironment monitoring to get reliable climate measurements, with better spatial and temporal resolution, has caused a paradigm shift from single-point fixed sensing to multipoint, multivariable flexible sensing. This has given WSNs an edge over wired networks. Moreover, to help automate greenhouse operations, a WSN can be integrated with an actuator network and a high-level programming platform to implement an online information system that growers can easily access.

To address these issues, we implemented the WSN-based online microclimate monitoring and control system shown in Figure 1.

### The Wireless Sensor Network

Our greenhouse WSN is a deterministic network based on IEEE 802.15.4 and XMesh.[7] Like ZigBee, it's a robust, full-featured multihop, ad-hoc, mesh networking protocol for embedded sensor devices for Crossbow motes. Network hardware comprises numerous battery-operated sensor nodes with embedded temperature and relative humidity sensors and a gateway node.[8,9] Each node is preprogrammed in TinyOS,[10] the component-based, event-driven embedded operating system that defines each node's sensing, computation, communication, and routing capabilities. To design a versatile, flexible, and robust WSN for greenhouse custom applications, we had to address the following technical issues.

***Climate-control variables.*** The important climate-control variables we considered to evaluate the greenhouse-crop VDP were the temperature (ranging from –10 to 50° C) and relative humidity (from 0 to 100 percent) within the greenhouse canopy and aerial height levels. The crop VDP is defined as the difference between the saturated vapor pressure at the canopy level at a given temperature and the actual vapor pressure present in the air at that temperature and relative humidity. Saturated vapor

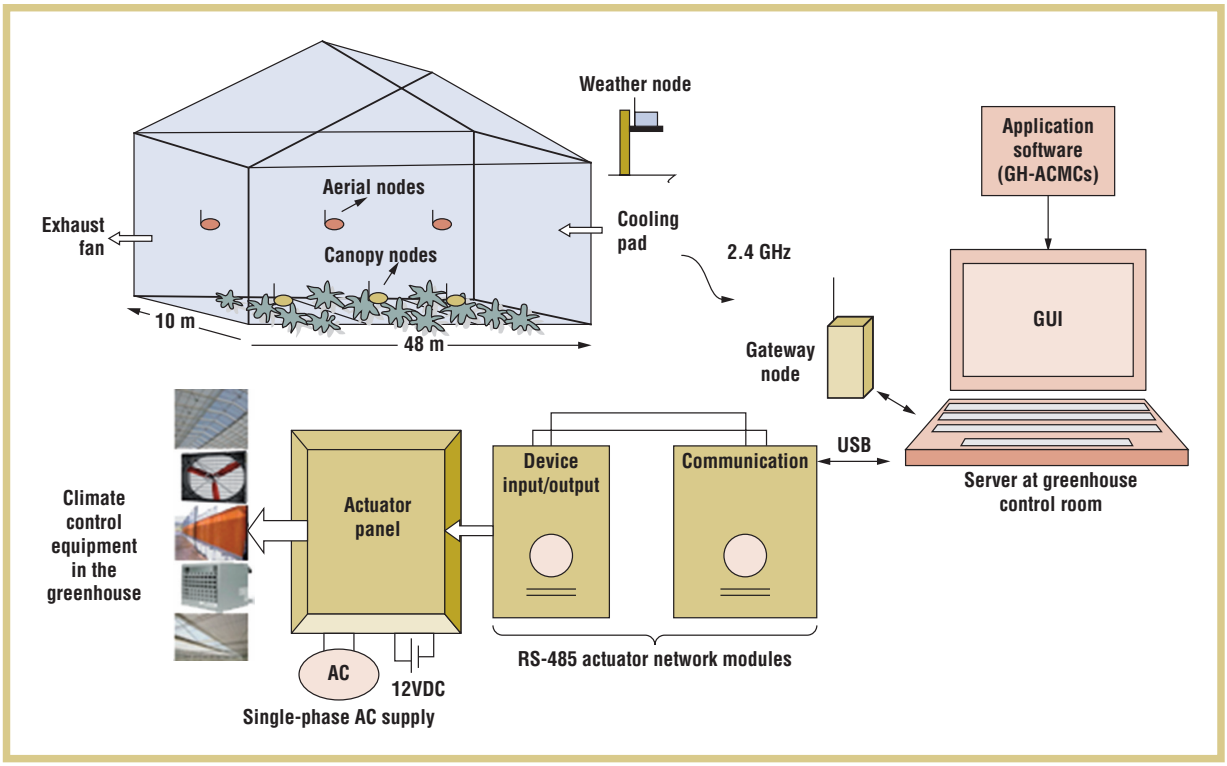



**Figure 1. A typical architecture of an online microclimate monitoring and control system for greenhouses, based on a wireless sensor network. Wireless sensor nodes are deployed in the greenhouse at different grid and height locations, and one weather node resides outside the greenhouse.**

pressure is a function of temperature and is calculated using the Arrhenius equation.[11]

The effect of having different temperatures at the canopy and aerial levels and relative humidity in a greenhouse is quantized as the greenhouse-crop VPD, which controls the transpiration rate and thus plant growth, health, and yield.[12] These climate variables fluctuate slowly and spatially within the greenhouse. A reporting-time interval of 15 minutes or so is recommended to extract dynamic statistics.[6]

***Network deployment scheme.*** The greenhouse WSN comprises a few sensor nodes deployed inside the greenhouse at specific locations and a single node deployed outside to sense weather parameters. A gateway node is connected to the server in the control room to facilitate data communication. To calculate the crop VPD and measure spatial variability of parameters within the greenhouse, we used a grid- and height-based deterministic, site-specific, and uniform deployment scheme (see Figure 1).

In this scheme, the greenhouse floor area is virtually partitioned into grids of uniform size, with two height levels—one within the canopy and the other at aerial levels (1.5 m above the canopy). Each location has an ID label, $g_i h_j$, where $i$ denotes the grid number (1 to $n$, where $n$ is the maximum number of grids in the greenhouse), and $j$ denotes the height number (1, 2). We placed at least one sensor node (preprogrammed with a relative location node ID) at each grid and height zone to measure the temperature and relative humidity. The grid size (200 to 500 $m^2$) characterizes the spatial variation of these parameters, keeping in mind that nodes are within radio range of each other (50 m).[8] There's a trade-off between the grid size, number of nodes (cost), and spatial resolution.

***Node addressing scheme.*** We used an addressing scheme based on the relative location of the node (RLN) for location identification. The RLN ID tags have a four-digit number represented as GHNN, where G denotes the grid number (1 to 9), H (1, 2) denotes the height number, and NN (00–99) denotes the node ID (which uniquely identifies the node in the greenhouse). This scheme allows site-specific deployment of nodes at the identified locations in the greenhouse.

Based on this ID tag, we used location mapping to extract data from each node location at the server. In addition to providing location-aware sensing of greenhouse parameters, this mapping

was more power efficient, simple, and straightforward to implement than a GPS-enabled, dynamic location-aware sensing scheme.[1]

***Data acquisition and transmission.*** To continuously monitor the greenhouse signal dynamics for better precision control, we used a periodic sampling with averaged delayed transmission (PSDT) data-acquisition and transmission algorithm. Instead of periodically sampling the sensor signals and transmitting the acquired data at the same instant, the PSDT algorithm lets the node periodically sample the sensor signals at a faster sampling rate, collect a few samples, and periodically transmit the averaged sampled values at a higher transmission than sampling rate. Both temperature and relative humidity are sampled at two-minute time intervals, and the latest averaged sampled values of each within a 15-minute time frame are packetized and transmitted to the gateway node.

This technique improves the reliability with which signals are sampled and is power efficient. It supports high latency operation and averaging to compress data samples without sacrificing measurement accuracy and data integrity. The transmission rate is fixed and is appropriate for any crop.

***Network topology and routing.*** Thick plantation and canopy coverage in a greenhouse can diminish the signal range of nodes forming the network, so packet losses can occur. To have a reliable and scalable network under such situations, a true mesh multihop network topology is preferred over a star- or hybrid-cluster-based topology.

All greenhouse nodes are full functional devices with both sensing and routing capabilities and one coordinator (gateway) node. Based on the XMesh routing algorithm, each node transmits its data packet to the coordinator node via the other parent node using multihops and the most efficient energy path. The node periodically updates its routing information (every six minutes) and dynamically creates new routes. As soon as a new node is added, it joins the network and starts transmitting data. The node remains in a low-power mode (with the transmitter in sleep mode and the receiver in a low-power listening mode) when there's no data or other message to receive or transmit.[7]

This topology extends the radio range of the devices with low-power consumption (a few µA), providing a field life of more than a month on a pair of AA standard cells (3 volts). Furthermore, it provides better coverage with self-healing and self-configuration capabilities, forming a fault-tolerant network well-suited for greenhouses.

## The Actuator Network

To automate the greenhouse climate-control process, we used an actuator network from Advantech, based on the RS-485 industrial serial-networking standard. We connected greenhouse end devices (single-phase 220-volt AC motors, pumps, and starters) associated with climate-control equipment to the network module's relay ports (12 V DC 220 ohm). In response to the climate controller output, which decides the operating load and presents the status of climate-control equipment, the actuator network drive-layer program issues commands to RS-485 network devices to actuate particular relays, controlling the equipment by varying the load.

## Application Software

We programmed the application software—the *greenhouse advance microclimate monitoring and control software* (GH-ACMCs)—for discrete packet acquisition and time-series analysis of network multivariate data. The GH-ACMCs lets us execute microclimate monitoring, climate control, and decision support functions, and it updates all vital greenhouse information on the GUI. Figure 2 shows the software's functional design model.

***Packet acquisition, scanning, and deciphering.*** Packet acquisition is fundamental in providing connectivity to the greenhouse WSN. Using a point-by-point packet-acquisition and -collection method, network data packets interfacing with a PC port at the gateway node at any instant of time $t$ are acquired and logged with a timestamp. Based on a variable timeframe-scanning algorithm, the logged data packets are scanned within the latest timeframe window $t_w$ (with a typical value of 20 minutes). The timeframe window denotes the time during which the latest data from the node is available—in other words, during this timeframe, the node has (at least once) communicated its data to the central server system. The scanning starts with the current time. Then, by decrementing the time by one minute until the timeframe limit is met, the algorithm collects the packets transmitted from nodes within this timeframe.

From the scanned packets, each node ID is location mapped to separate the node packets for different locations. Corresponding to each location node ID, the latest packet available in the timeframe window is considered as the node packet from that location at that time instant. The raw data packets are deciphered and 16-bit digital data for node voltage, temperature, and humidity are converted to corresponding values in engineering units using sensor conversions.[13] An amplitude rejection-filtering technique filters out data packets that are undesirable due to a sensor malfunctioning. This method synchronizes the display of data from each node in the timeframe at each time instant and updates the information on the GUI as soon as a new packet from any node is detected. Without this method, it's difficult to coordinate all nodes and provide stable, reliable, and timely network information from an asynchronous, high-latency (15 minutes) network with an intermediate low-power mode of operation.

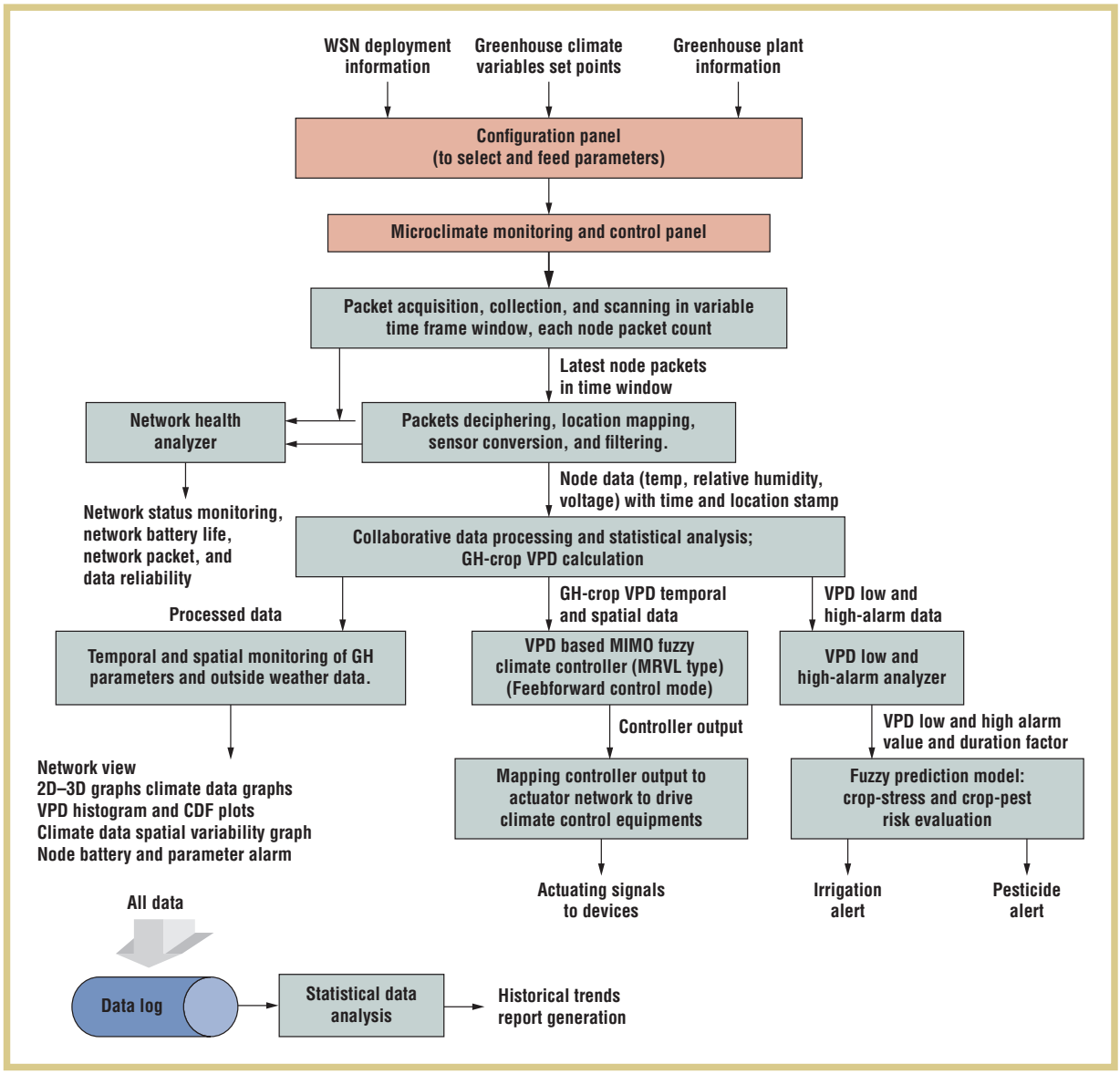



**Figure 2. A functional flow chart of the application software—greenhouse advance microclimate monitoring and control software (GH-ACMCs).**

***Collaborative data processing and microclimate monitoring.*** The preprocessed time-instant values of temperature and relative humidity, as measured by nodes at different locations, are collaboratively processed to evaluate the crop VPD at each grid and provide reliable average values of climate variables.[11,12] Data from different nodes is statistically analyzed to project the spatial variation of climatic variables in the grids and at various height levels (3D plots). The spatial variability (standard deviation) value for each climate variable is calculated to estimate the data spread. VPD-histogram and cumulative frequency-distribution graphs are obtained to analyze the frequency of VPD variation in different ranges and the amount of time during which the VPD is less than a certain value.

***Climate control.*** To regulate the greenhouse climatic conditions automatically, we implemented a VPD-based fuzzy climate controller,[14,15] operating under a feed-forward mode using a variable-load design method. With respect to the variations in the

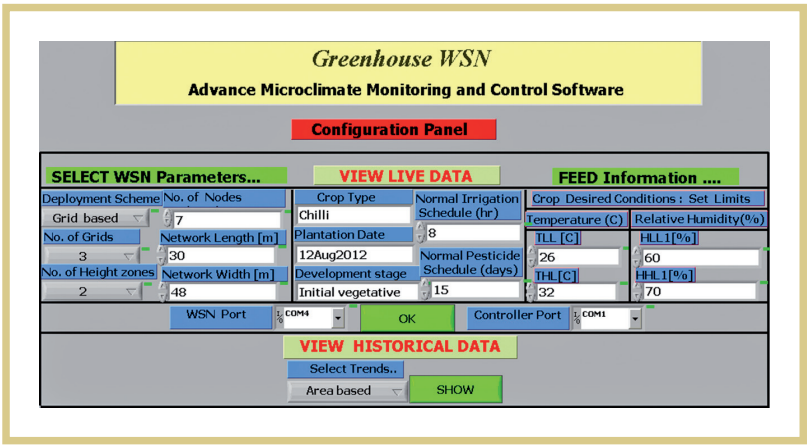



**Figure 3. The configuration panel GUI. This screenshot shows user-selected network deployment and crop-specific climatic parameters when the network system was deployed in a greenhouse for growing chilies.**

outside-weather VPD, we calculated the VPD error and its rate of change. Corresponding to each error range (positive and negative) are different rule-based fuzzy controllers that issue signals to drive different climate-control equipment and regulate inside conditions. Controller inputs, the VPD error, and the VDP error's rate of change are normalized into seven linguistic terms, and 49 control rules for each actuating system (roof vents, exhaust fan, cooling pad, and heaters) are designed, taking into account how varying the operating load of each affects the inside climatic condition.

When the VPD error is positive, the greenhouse is humidified by controlling the operating load of ventilation and cooling equipment. When the VPD error is negative, the greenhouse is dehumidified by controlling the operating load of ventilation and heating equipment. We operate a shade screen to increase the efficiency of the cooling and heating system in extreme summer and winter conditions. When the outside temperature exceeds 35°C during the day and falls below 5°C at night, the shade screen is covered. The RS-485 actuator network driver-layer program maps the controller output to issue commands to drive the devices. Based on the user-specified temperature and relative humidity limits, suited for the crop during its growing phase, VPD set limits are calculated to drive controller inputs.

***Crop-stress and -disease risk prediction.*** The quantitative analysis of alarming VPD conditions helps predict VPD's harmful effect on crops—that is, when the crop is under stress due to continuous high VPD conditions (high temperature and low humidity), or when there are chances of disease outbreak due to continuously low VPD (low temperature and high humidity), which can cause condensation on leaves. This functional module uses heuristic rule-based fuzzy controllers and a cumulative moving-average method to analyze VPD high and low alarm conditions (in terms of the grid and area), based on the VPD's alarm duration (during the normal irrigation and pesticide schedule) and the extent of the high and low VDP values to estimate a crop-stress and crop-disease risk index (0 to 1). After comparing these two indices to the threshold limit, irrigation and pesticide warning messages are issued for each grid and greenhouse area.[12,16]

***Network performance monitoring.*** This functional module evaluates some of the important parameters related to a WSN's field performance in a greenhouse.[1] It tracks the number of nodes connected at each timeframe limit and displays the network connectivity status ("okay," "network initializing," or "detection or connection problem"). It has a low-battery alarm (2.3 V) and indicates the *network mean-battery life* and *network mean packet reliability* (as percentages). It also tracks data connectivity based on the display of vital information—the average crop VPD or outside VPD—and calculates *network data reliability* (as a percentage).

***Implementation.*** All of the functional elements of the GH-ACMCs are implemented on the graphical dataflow programming platform, LabVIEW (8.5), from National Instruments.[17] It's a multipanel, modular, hierarchically designed software tool with an intuitive GUI that supports different levels of functionalities with intelligent, salient features. The configuration panel lets the user feed and select user-defined parameters related to greenhouse network deployment and plant-specific requirements, such as the temperature and relative humidity (see Figure 3). Based on these inputs, the microclimate monitoring and control program executes all the functional modules to perform online monitoring, control operations, and display vital information (see Figure 4).

## Field Results

To test the system's feasibility and judge its measurement and network performance in a real-world field situation, we deployed it in a commercial chili-growing glasshouse, situated in a tropical region of northern India in the state of Punjab, near the city of Ludhiana. Greenhouse cultivation is a daunting task for the region's local growers during the summer season (May through August), when the outside VPD exceeds 70 millibars (mB) owing to high temperatures (above 40°C) and moderate relative humidity (20 to 60 percent). To evaluate the greenhouse climate dynamics and statistics under this harsh summer

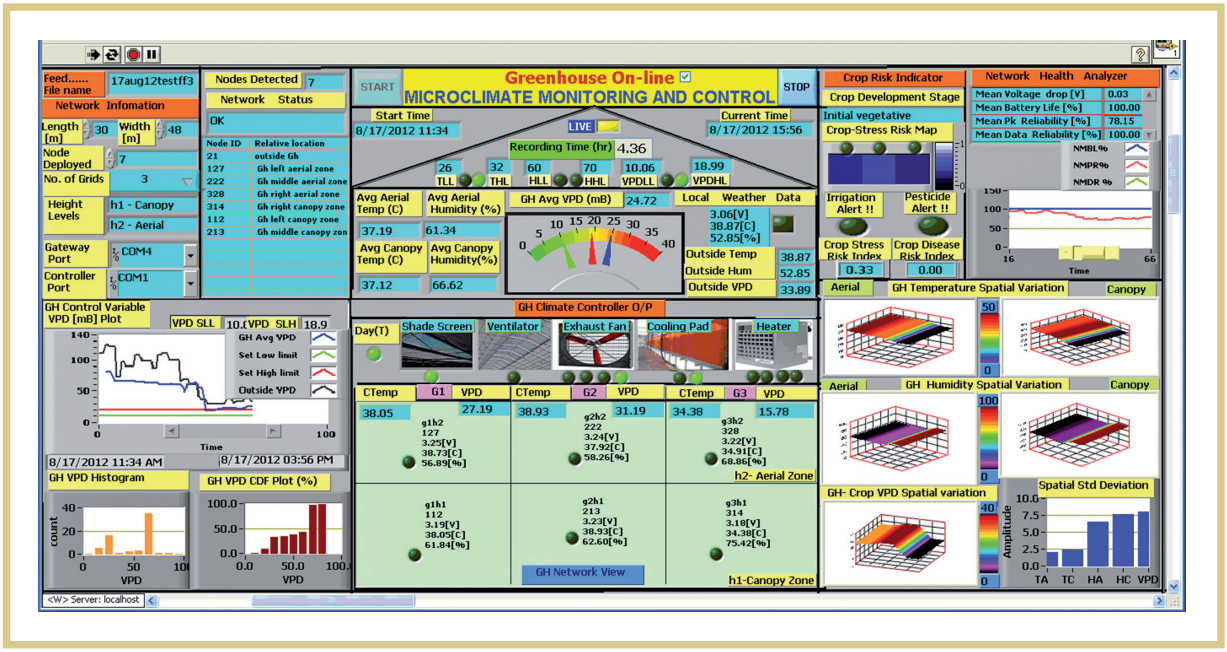



**Figure 4. The microclimate monitoring and control panel GUI. This screenshot indicates the state of the greenhouse at the current time instant, showing online network connectivity and displaying greenhouse climate-control dynamics and vital information.**

weather, we conducted a short-term experiment during the peak day hours.

Preprogrammed wireless sensor nodes packed in PVC housings, equipped with a standard pair of AA cell batteries (3.0–3.2 V), were powered "on" and deployed at specific locations in the greenhouse (using the grid-height based deployment scheme explained earlier). As shown in Figure 1, the greenhouse section area ($30 \times 48$ m$^2$, East to West direction) was divided into three grids—left (G1), middle (G2), and right (G3)—of size $30 \times 16$ m$^2$, and at the center of each grid, one node was placed in the canopy (the average crop size was initially 12 cm) to measure the canopy temperature and relative humidity. Another node was fixed at the aerial height level (1.5 m above the canopy) to measure the aerial temperature and relative humidity. The weather node was mounted outside the greenhouse, and the gateway node was interfaced to the server housed at the greenhouse control room (40 m from greenhouse site).

To automate the climate-control process, actuating device terminals were interfaced to an RS-485 actuator panel at the greenhouse site, and an RS-485 communication module was serially connected to the host PC. Depending on the availability of the devices at the greenhouse site, the climate controller was tuned to drive the cooling pad (0, 50, and 100 percent), exhaust fan (0, 25, 50, 75, and 100 percent), shade screen (0 to 100 percent), and roof ventilation (0 to 100 percent) with variable loads.

Before starting the monitoring and control operations, we configured the GH-ACMCs. We opened the configuration panel to feed input parameters (Figure 3). We entered the crop comfort zone ("set limits") of 26° C to 32° C, with a relative humidity of 60 to 70 percent, along with the crop normal-irrigation schedule (typically eight hours) and pesticide schedule (typically 15 days), as decided by the grower during the crop development cycle (initial vegetative growth period of chili). We also entered the network deployment information. As the configuration panel was executed, it linked the information to the microclimate monitoring and control panel GUI (Figure 4).

After the data-log file name was fed into the system, clicking the "start" button executed the program continuously with a start time of 11:34 a.m. on 17 August 2012. The information on the panel was updated as soon as a new network packet arrived within the 20-minute timeframe, with the scanning-time increment of one minute. At the start, there was an initial display lag (6 to 15 minutes) because of the network's high latency and low power mode operation.[7]

Results depicted the system state at the current time instant. At 3:56 p.m. on 17 August 2012, the network status was "OK." All seven nodes deployed were detected, and their node IDs and location information were displayed. None of the nodes were blinking (low battery alarm) or indicating a "malfunctioning." The weather node indicated a high outside temperature of 38.87° C and relative humidity of 52.85 percent, resulting in a high VPD (33.89 mB). Based on the user's set limits, the crop VPD set limits were calculated and displayed (10.06 mB to 18.99 mB).

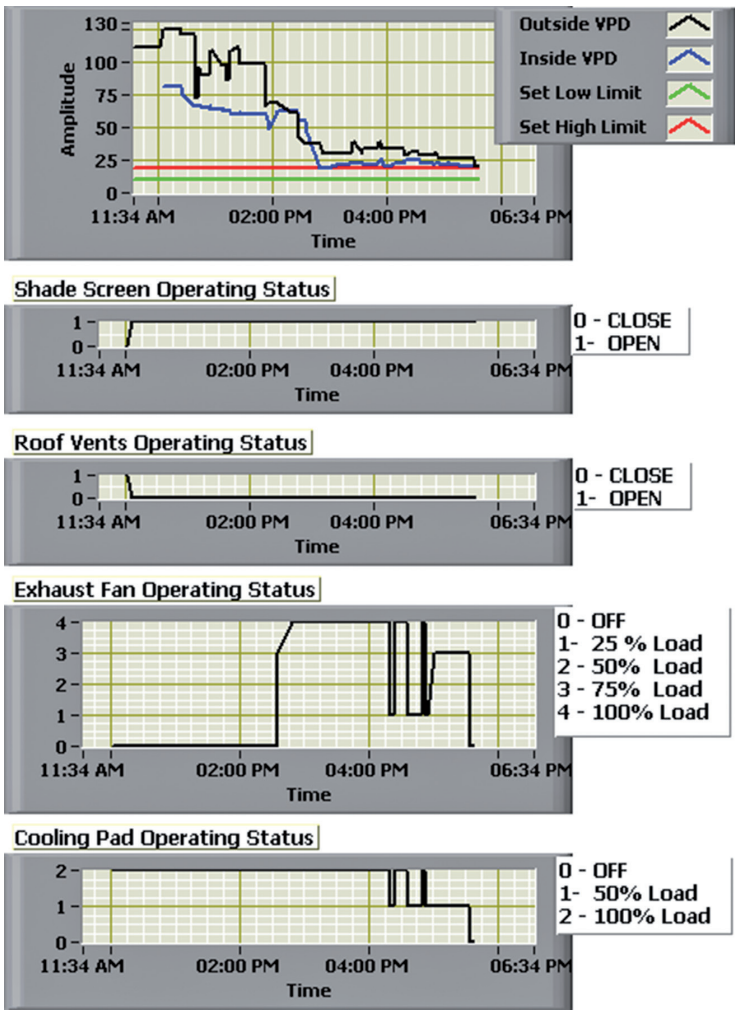



**Figure 5. Time variation of average greenhouse-crop vapor pressure deficit (VPD) (inside VPD) under climate-control operation, along with outside VPD and set limits and corresponding variations in the operating status of climate-control equipment over the recorded period of six hours.**

With respect to the current VPD error (14.9 mB), the climate controller output actuated the RS-485 network to drive the equipment with the appropriate load/operating status to humidify the greenhouse to minimize the VPD error. As indicated, the shade screen was "open" (covered), the roof vents were "closed," and the cooling pad and exhaust fan were operating under a "full" load. The network view graphically showed the nodes at the respective grid locations, indicating the local measured value of voltage, temperature, and relative humidity. The VPD value at each grid was calculated and displayed, along with the display of average values of the climate variables at the respective indicators. The greenhouse's average canopy temperature (37.12° C) and aerial temperature (37.19° C) remained high, with moderate aerial humidity (61.34 percent), resulting in the high average value of crop VPD (24.72 mB). The VPD time plot indicated that because of the climate-control operation, the greenhouse VPD decreased with time and was lower than outside, but it remained higher than the set limit.

Owing to the high average VPD, the crop was under stress, and the estimated value of the crop-stress index (area) increased 0.33 over time (after 4.36 hours) but was lower than the threshold (0.5) to initiate the irrigation alert alarm for the area. Also, the crop-stress risk-intensity map showed the spatial variation of the crop-stress risk index at the respective grids, based on grid VPD alarm analysis using prediction model and the corresponding grid irrigation alarm status. 3D plots indicated instantaneous spatial variation of the microclimate variables at different grid locations, and the bar graph plotted the spatial standard deviation of each, indicating the extent of variability of each from its average value. The greenhouse-crop grid VPD varied in the range (15.78 mB to 31.19 mB), with a spatial variability value of 7.5mB. Histogram and cumulative distribution frequency (CDF) plots indicated the crop VPD distribution pattern. For more than half of the time, the VPD was very high (greater than 50 mB) and then remained within 20 to 30 mB.

As indicated by the network health analyzer, the network's mean battery life at the current time instant was highest (100 percent), because all nodes had a voltage greater than 3.0 V, with a network mean voltage drop of 0.03 V. Due to packet losses, the network mean packet reliability remained between 75 and 100 percent, but the data reliability was very high (100 percent). Data was simultaneously recorded and logged in the respective files for historical display and trend analysis.

Figure 5 shows variations in the average value of the greenhouse-crop VPD under climate-control operation, as well as variations in the outside VPD

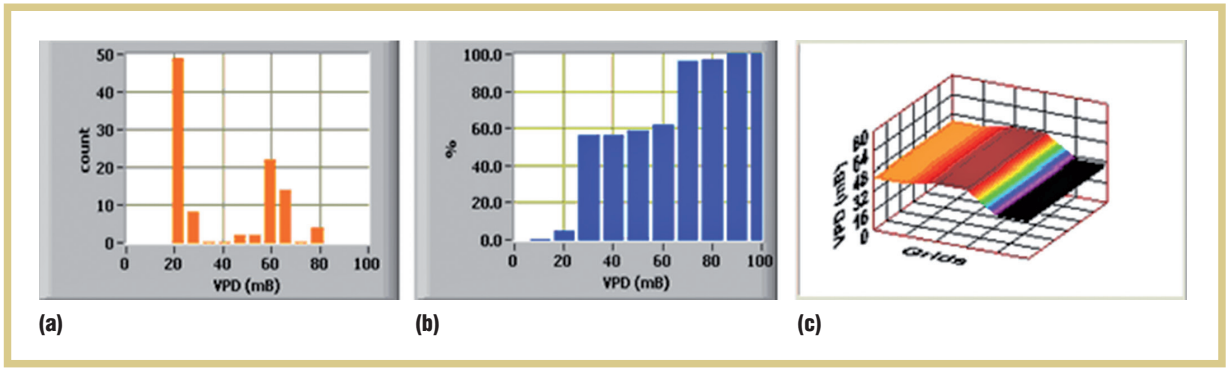



Figure 6. Greenhouse-crop VPD distribution patterns: (a) histogram, (b) cumulative distribution frequency (CDF) plot, and (c) mean VPD spatial plot over the recorded period of six hours.

and set limits and the corresponding variations in the actuating status and load of the climate-control equipment over the recorded period of six hours. Because the outside temperature was higher than 35°C, the shade screen was opened (covered) and the roof vents remained off. Under the extreme weather conditions, during the initial hours—when the outside VPD was extremely high (VPD error above 50 mB)—the greenhouse was humidified with the maximum cooling rate with exhaust fan ventilation turned off, resulting in a decrease of the average value of greenhouse-crop VPD, from 81.6 to 55 mB.

When the outside VPD dropped below 60 mB, the exhaust fan operating load increased from 75 percent to 100 percent. Increasing the ventilation rate with cooling further reduced the VPD (from 25 mB to 20 mB). When the outside VPD error dropped below 10 mB, the cooling pad and ventilation were switched to lower rates, which decreased the inside VPD to a minimum value of 18.18 mB. When the outside error was less than 1 mB and its rate of change was also negative, all the actuating devices were switched off (to the low-power mode).

Figure 6 shows the frequency (Figure 6a) and CDF distribution pattern (Figure 6b) of the crop VPD, indicating that for more than half the time, during the initial hours, the crop VPD remained higher than 50 mB. Also during that time, 30 percent of the values fell between 60 and 70 mB. During the later time, the crop VPD remained lower than 40 mB, and 50 percent of the values fell between 20 and 30 mB. Figure 6c shows the mean spatial variation pattern of the VPD at grids, which ranged from 47.0 mB to 30.14 mB, with a mean standard deviation of 8.78 mB. Crop VPD was lowest at the right grid (which was near the cooling pad), followed by left grid (near the exhaust fans), and it was highest at the middle grid (at the center of the greenhouse).

For six hours, our WSN monitored and controlled a greenhouse climate with high data and packet reliability (85 to 100 percent) and low battery drop (0.03 V). The climate controller tracked the initially high inside VPD and lowered the value to meet optimal conditions by operating the devices as needed. Furthermore, the real-time display of greenhouse climate-control statistics helped the grower make better decisions in executing greenhouse operations, eventually leading to healthy crop growth and better yields. Improving greenhouse climate control under such extreme weather conditions will require more efficient cooling systems (with variable loads) or grid-based humidifiers that could quickly decrease the initial rise of greenhouse-crop VPD during peak hours to obtain more uniform and optimum VPD control.

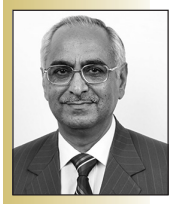
We plan to perform long-term system testing under different weather conditions with further improvement in control strategies and hope to implement irrigation, light-intensity, and carbon-dioxide monitoring modules to enhance the capabilities and performance of our automatic greenhouse climate controller. P

## REFERENCES

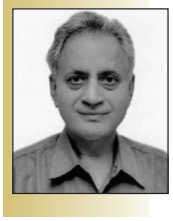

1. K . Sohraby, D. Minoli, and T. Znati, *Wireless Sensor Networks—Technology, Protocols and Applications*, Wiley Interscience, 2007.
2. J. Burel, T. Brooke, and R. Beckwith, "Vineyard Computing: Sensor Networks in Agricultural Production," *IEEE Pervasive Computing*, vol. 3, no. 1, 2004, pp. 38–45.
3. Accenture, "Remote Sensor Network Accenture Prototype Helps Pick Berry Vineyard Improve Crop Management," tech. report, 2005; www.accenture.com/SiteCollectionDocuments/PDF/pickberry.pdf.
4. S. Shanmuganthan, A. Ghobakhlou, and P. Sallis, "Sensor Data Acquisition for Climate Change Modeling," *WSEAS Trans. Circuit and Systems*, vol. 7, no. 11, 2008, pp. 942–952.
5. H. Liu, Z. Meng, and S. Cui, "A Wireless Sensor Network Prototype for Environmental Monitoring in Greenhouse," *Proc. Int'l Conf. Wireless Comm. Networking and Mobile Computing*, WiCOM, 2007, pp. 2344–2347.
6. G.J. Timmerman and P.G.H Kamp, *Computerized Environmental Control in Greenhouses*, Ede, the Netherland PTC, 2003.

## the AUTHORS

**Roop Pahuja** is an associate professor and a PhD student in the Department of Instrumentation and Control Engineering at the National Institute of Technology Jalandhar, India. Her research interests include sensors, sensor networks, virtual instrumentation and graphical system design, the development of intelligent PC-based measurement and control systems, and WSN applications. Pahuja received her M.Tech in measurement and instrumentation from the Indian Institute of Technology, Roorkee, India. Contact her at pahujar@nitj.ac.in.



**H.K. Verma** is a distinguished professor in the Department of Electrical and Electronics Engineering at Sharda University, India. His research interests include smart sensors and sensor networks, embedded systems, digital relays and power system protection, hydraulic measurement and hydro-electric power development, and intelligent and energy-efficient buildings. Verma received his PhD in power system instrumentation and protection from the University of Roorkee, India. He is a member of the Institution of Engineers (India), Institution of Electronics and Telecommunication Engineers, and International Society of Automation. Contact him at hkvfee@gmail.com.



**Moin Uddin** is the pro vice–chancellor at Delhi Technological University, India. His research interests include computer networking, AI and soft computing, wireless communication and networks, and robotics. Uddin received his PhD in electronics and computer engineering from the University of Roorkee, India. He is a member of All India Council for Technical Education, the Ministry of Human Resource Development, India, and the India Society for Technical Education. Contact him at prof_moin@yahoo.comemail.

7. *XMesh User Manual*, Crossbow Technology, 2007.
8. *MTS/MDA Sensor Board User Manual*, Crossbow Technology, 2007.
9. *MPR/MIB User Manual*, Crossbow Technology, 2007.
10. P. Levis and D. Gay, *TinyOS Programming*, Cambridge Univ. Press, 2009.
11. J.J. Prenger and P.P. Ling, "Greenhouse Condensation Control," Fact Sheet (Series) AEX-800, Ohio State Univ. Extension, 2000.
12. "Understanding Humidity Control in Greenhouse," Fact Sheet no. 400-5, BC Ministry of Agriculture, Fisheries and Food, 1994.
13. *Xserve User Manual*, Crossbow Technology, 2007.
14. A. Errahmani, M. Benyakhlef, and I. Noumhidi, "Decentralized Fuzzy Control Applied in a Greenhouse," *ICGST-ACSE J.*, vol. 9, no. II, 2008, pp. 35–40.
15. K. Gottschalk, L. Nagy, and I. Farkas, "Improved Climate Control for Potato Stores by Fuzzy Controllers," *Computers and Electronics in Agriculture*, vol. 40, nos. 1–3, 2003, pp. 127–140.
16. E. Turban, J.E. Aronson, and T.-P. Liang, *Decision Support Systems and Intelligent Systems*, Prentice Hall, 2006.
17. S. Sumathi and P. Surekha, *LabVIEW Based Advanced Instrumentation Systems*, Springer, 2007.

Microelectronics Journal ■ (■■■■) ■■■–■■■

Contents lists available at SciVerse ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

# Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells

Kirti Gupta [a], Neeta Pandey [a,*], Maneesha Gupta [b]

[a] Delhi Technological University (Formerly Delhi College of Engineering), Electronics and Communication, Bawana Road, Delhi 110042, India
[b] Netaji Subhas Institute of Technology, Electronics and Communication, Dwarka, Delhi-110078, India

ARTICLE INFO

*Article history:*
Received 26 December 2011
Received in revised form
24 October 2012
Accepted 11 March 2013

*Keywords:*
MCML
Exclusive-OR gate
Low-voltage
Triple-tail cell

ABSTRACT

A new low-voltage MOS current mode logic (MCML) topology for an exclusive-OR (XOR) gate using triple-tail cell concept is proposed. The design of the proposed MCML XOR gate is carried out through analytical modeling of its static parameters. The delay is expressed in terms of the bias current and the voltage swing so that it can be traded off with the power consumption. The proposed low-voltage MCML XOR gate is analyzed for three design cases namely high-speed, power-efficient, and low-power and the performance is compared with the traditional MCML XOR gate for each design case. The theoretical propositions are validated through extensive SPICE simulations using TSMC 0.18 μm CMOS technology parameters.



## 1. Introduction

With the emergence of the high-resolution mixed-signal applications, there is a demand for the integration of high performance digital and analog circuits on the same chip [1,2]. The traditional CMOS logic style does not provide an analog friendly environment due to the large switching noise [3,4]. Hence, it is necessary to explore alternate logic styles. Among the different logic styles suggested in literature [5–12], MOS current mode logic (MCML) style is the most promising one due to the lower switching noise in comparison to traditional CMOS logic style [13]. Also, MCML style exhibits better power-delay than the traditional CMOS logic style at high frequencies [14,15]. Therefore, MCML style is appropriate for designing high performance digital circuits wherein an exclusive-OR (XOR) gate is widely used as a building block in different applications such as comparators, adders, multipliers, test pattern generators etc [14–20].

The traditional MCML XOR gate uses stacked source-coupled transistor pairs which puts a limit on the minimum power supply [21]. The XOR gate presented in [22–24] enables low voltage operation by avoiding the stacking of transistor pairs. The circuit of [22] is symmetric with respect to the inputs and provides single-ended output. A folded circuit presented in [23] provides differential output. The circuits suggested in [22,23] use additional voltage and current sources for correct operation as compared to the traditional MCML XOR gate. The PFSCL style proposed in [24] is single-ended approach. It uses NOR based implementation and would therefore require multiple stages.

In this paper, a new low-voltage MCML XOR gate is proposed. It uses triple-tail cell concept of bipolar transistors [25–27]. The proposed circuit is differential, does not require any additional voltage sources [22,23] or multiple stages [24] as compared to similar available low-voltage topologies [22–24]. The static parameters for the proposed XOR gate are analytically modeled and applied to develop a design approach. The delay is expressed in terms of the bias current and the voltage swing so that it can be traded off with the power consumption. The proposed low-voltage MCML XOR gate is analyzed for the three design cases namely high-speed, power-efficient, and low-power. A comparison in performance of the proposed XOR gate with the traditional one is carried out for all the cases.

The paper first briefs the operation of the traditional MCML XOR gate in Section 2. Thereafter, a new low-voltage MCML topology for the XOR gate is proposed and analytical formulations for different static parameters and delay are put forward in Section 3. The analysis of the proposed XOR gate for the three design cases namely high-speed, power-efficient, and low-power and its performance comparison with the traditional one is discussed in Section 4. Extensive SPICE simulations are carried out to validate the proposed theory. Section 5 concludes the paper.

* Corresponding author. Tel.: +91 112 6868090.
*E-mail address:* n66pandey@rediffmail.com (N. Pandey).

0026-2692/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.mejo.2013.03.007

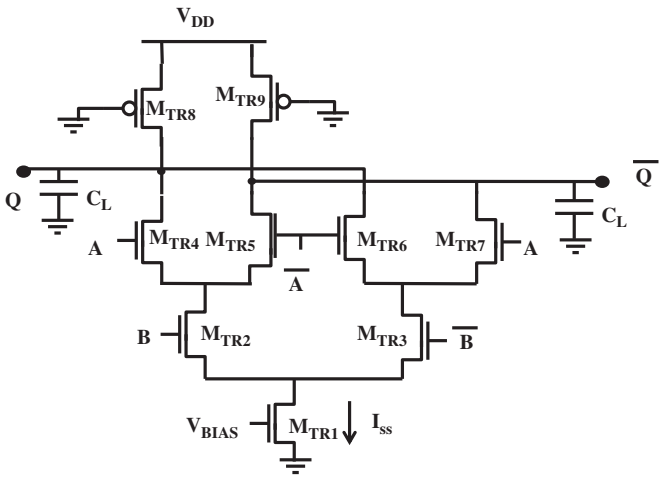



**Fig. 1.** Traditional MCML XOR gate.

## 2. Traditional MCML XOR gate

A traditional MCML XOR gate with differential inputs A and B is shown in Fig. 1. It consists of two levels of source-coupled transistor pairs to implement the logic function and a constant current source $M_{TR1}$ to generate bias current $I_{SS}$. The differential input B drives the lower level transistor pair $M_{TR2}$–$M_{TR3}$ that alternatively activates the upper level transistor pairs $M_{TR4}$–$M_{TR5}$ and $M_{TR6}$–$M_{TR7}$. When differential input B is high, $M_{TR3}$ is off, the bias current $I_{SS}$ flows through $M_{TR2}$ and is steered either to $M_{TR4}$ or $M_{TR5}$ according to the differential input A. Conversely, when differential input B is low, the bias current $I_{SS}$ flows through $M_{TR3}$ and is steered to one of the two transistors i.e. either $M_{TR6}$ or $M_{TR7}$ depending on the input A. The bias current $I_{SS}$ is converted to the differential output voltage ($V_Q - \overline{V_Q}$) through the PMOS transistors $M_{TR8}$ and $M_{TR9}$ [28]. The load capacitance $C_L$ includes the effect of fanout, and the interconnect capacitances.

The minimum supply voltage, $V_{DD\_MIN\_TR}$ for the traditional XOR gate is defined as the lowest voltage at which all the transistors in the two levels and the current source operate in the saturation region [29] and is computed as

$$V_{DD\_MIN\_TR} = 3V_{BIAS} - 3V_{T\_TR1} + V_{T\_TR} \quad (1)$$

where $V_{T\_TR}$ is the threshold voltage of the transistors $M_{TR4,5,6,7}$, $V_{T\_TR1}$ is the threshold voltage of $M_{TR1}$, and $V_{BIAS}$ is the biasing voltage of $M_{TR1}$.

## 3. Proposed low-voltage MCML XOR gate

The proposed low-voltage XOR gate with differential inputs A and B is shown in Fig. 2. It consists of two triple-tail cells ($M_{LV3}$, $M_{LV4}$, $M_{LV7}$) and ($M_{LV5}$, $M_{LV6}$, $M_{LV8}$) biased by separate current sources of $I_{SS}/2$ value. The transistors $M_{LV7}$ and $M_{LV8}$ are driven by the differential B input and are connected between the supply terminal and the common source terminal of transistor pairs $M_{LV3}$–$M_{LV4}$ and $M_{LV5}$–$M_{LV6}$ respectively. A high differential B voltage turns on the transistor $M_{LV7}$, and deactivates the transistor pair $M_{LV3}$–$M_{LV4}$. At the same time, the transistor $M_{LV8}$ turns off so that the transistor pair $M_{LV5}$–$M_{LV6}$ generates the output according to the differential input A. Similarly, the transistor pair $M_{LV3}$–$M_{LV4}$ gets activated for low differential B voltage and produces the corresponding output.

The minimum supply voltage, $V_{DD\_MIN\_LV}$ for the proposed XOR gate is computed by the method outlined in [29] as

$$V_{DD\_MIN\_LV} = 2V_{BIAS} - 2V_{T\_LV1} + V_{T\_LV} \quad (2)$$

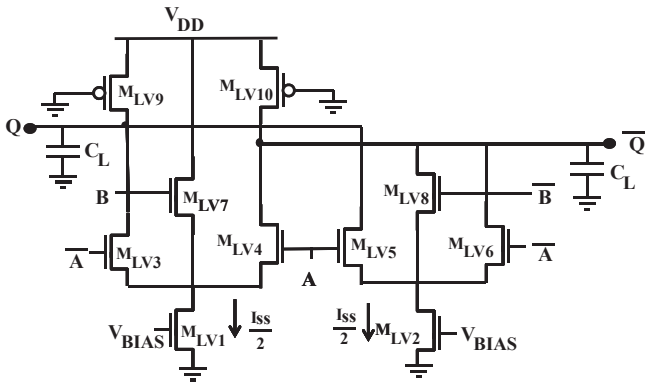



**Fig. 2.** Proposed low-voltage XOR gate.

where $V_{T\_LV}$ is the threshold voltage of transistor $M_{LV3,4,5,6}$, $V_{T\_LV1}|$ is the threshold voltage of $M_{LV1}$, and $V_{BIAS}$ is the biasing voltage of $M_{LV1}$.

### 3.1. Static model

The static model is derived by modeling the load transistors $M_{LV9}$, $M_{LV10}$ by an equivalent linear resistance, $R_P$ [30]. Using the standard BSIM3v3 model, the linear resistance, $R_P$ is computed as

$$R_P = \frac{R_{int}}{1 - \frac{(R_{DSW}10^{-6})/W_P}{R_{int}}} \quad (3)$$

where $R_{DSW}$ is the empirical model parameter and the parameter $R_{int}$ is the intrinsic resistance of the PMOS transistor in the linear region and is given as

$$R_{int} = \left[\mu_{eff,p} C_{ox} \frac{W_P}{L_P}(V_{DD} - |V_{T,p}|)\right]^{-1} \quad (4)$$

where $C_{ox}$ is the oxide capacitance per unit area. The parameters $\mu_{eff,p}$, $V_{T,p}$, $W_P$ and $L_P$ are the effective hole mobility, the threshold voltage, the effective channel width and effective channel length of the load transistor respectively.

It may be noted that if equal aspect ratio of all transistors in the triple-tail cells is considered, then the transistors $M_{LV7}$ and $M_{LV8}$ will not be able to completely switch off the transistor pair $M_{LV3}$–$M_{LV4}$ and $M_{LV5}$–$M_{LV6}$. Hence, for proper operation, the aspect ratio of transistors $M_{LV7}$ and $M_{LV8}$ is made greater than other transistors' aspect ratio by a factor $N$. As an example if the value of differential input A is chosen such that the transistors $M_{LV3}$ and $M_{LV6}$ are on while the transistors $M_{LV4}$ and $M_{LV5}$ are off. Then, a high differential B voltage turns on the transistor $M_{LV7}$. But since the transistors $M_{LV3}$ and $M_{LV7}$ have the same gate-source voltages, the currents flowing through $M_{LV3}$ and $M_{LV7}$ can be written as

$$i_{D,3} = \frac{I_{SS}}{2}\frac{1}{1+N} \quad (5a)$$

$$i_{D,7} = \frac{I_{SS}}{2}\frac{N}{1+N} \quad (5b)$$

The current through $M_{LV3}$ can be minimized by increasing factor $N$. This input condition produces $V_{OH}$ as

$$V_{OH} = V_Q - \overline{V_Q} = R_P[(i_{D,4} + i_{D,6}) - (i_{D,3} + i_{D,5})] = \frac{R_P I_{SS}}{2}\left(\frac{N}{1+N}\right) \quad (6)$$

where $i_{D,3}, i_{D,4}, i_{D,5}$, and $i_{D,6}$ are the currents through transistors $M_{LV3}$, $M_{LV4}$, $M_{LV5}$, and $M_{LV6}$ respectively. The differential output voltages for various input combinations are enlisted in Table 1.

Please cite this article as: K. Gupta, et al., Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.03.007

**Table 1**
Differential output voltages of the proposed XOR gate for various input combinations.

| Differential inputs | | Currents through the transistors | | | | | | Differential output ($V_Q - \overline{V_Q}$) | |
|---|---|---|---|---|---|---|---|---|---|
| ***A*** | ***B*** | $M_{LV3}$ | $M_{LV4}$ | $M_{LV5}$ | $M_{LV6}$ | $M_{LV7}$ | $M_{LV8}$ | **Level** | $R_P[(i_{D,4}+i_{D,6})-(i_{D,3}+i_{D,5})]$ |
| L | L | $I_1$ | 0 | 0 | $I_3$ | 0 | $I_2$ | $V_{OL}$ | $-R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| L | H | $I_3$ | 0 | 0 | $I_1$ | $I_2$ | 0 | $V_{OH}$ | $R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| H | L | 0 | $I_1$ | $I_3$ | 0 | 0 | $I_2$ | $V_{OH}$ | $R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| H | H | 0 | $I_3$ | $I_1$ | 0 | $I_2$ | 0 | $V_{OL}$ | $-R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |

Where L/H=low/high differential input voltage, $I_1=I_{SS}/2$, $I_2=I_{SS}/2(N/(1+N))$ and $I_3=I_{SS}/2(1/1+N)$.

Hence, from Table 1, the voltage swing of the circuit can be expressed as

$$V_{SWING}=V_{OH}-V_{OL}=R_PI_{SS}\left(\frac{N}{1+N}\right) \tag{7}$$

Eq. (7) has been further approximated as

$$V_{SWING}=R_PI_{SS} \quad \text{for large values of } N \tag{8}$$

The small-signal voltage gain ($A_v$) and noise margin (NM) for the proposed XOR gate are computed by the method outlined in [30] as

$$A_v=g_{m,n}R_P=\frac{1+N}{N}\frac{V_{SWING}}{2}\sqrt{2\mu_{eff,n}C_{OX}\frac{W_N}{L_N}\frac{1}{I_{SS}}} \tag{9}$$

$$NM=\frac{V_{SWING}}{2}\left[1-\frac{\sqrt{2}}{A_v}\right] \tag{10}$$

where $\mu_{eff,n}$, $g_{m,n}$, $W_N$ and $L_N$ are the effective electron mobility, the transconductance, the effective channel width and length of transistors $M_{LV3,4,5,6}$ respectively.

### 3.2. Transistor sizing

In this section, an approach to size the transistors of the proposed low-voltage XOR gate on the basis of static model is developed.

For a specified value of NM, factor $N$ and $A_v$ (≥1.4 for MCML [16]), the voltage swing of the proposed XOR gate is calculated using Eq. (10) as

$$V_{SWING}=\frac{2NM}{1-\frac{\sqrt{2}}{A_v}}\left(\frac{N}{1+N}\right) \tag{11}$$

It may be noted that $V_{SWING}$ should be lower than the maximum value of $2V_T$ so as to ensure that transistors $M_{LV3,4,5,6}$ operates in saturation region. The voltage swing obtained from Eq. (11) requires sizing of the load transistor with equivalent resistance $R_P\left(=\frac{1+N}{N}\frac{V_{SWING}}{I_{SS}}\right)$. To this end, the equivalent resistance, $R_{P\_MIN}$, for the minimum sized PMOS transistor is first determined and then the bias current $I_{HIGH}$ for the required voltage swing is determined as

$$I_{HIGH}=\frac{V_{SWING}}{R_{P\_MIN}} \tag{12}$$

If the bias current is higher than $I_{HIGH}$, then $R_P$ should be less than $R_{P\_MIN}$ and this is achieved by setting $L_P$ to its minimum value i.e. $L_{MIN}$ and $W_P$ which is calculated by solving Eqs. (3) and (4) as

$$W_P=\frac{N}{1+N}\frac{I_{SS}}{V_{SWING}}\frac{L_{MIN}}{\mu_{eff,p}C_{ox}(V_{DD}-|V_{T,p}|)1-(R_{DSW}10^{-6}/L_{MIN})[\mu_{eff,p}C_{ox}(V_{DD}-|V_{T,p}|)]} \tag{13}$$

Similarly, if the bias current is lower than $I_{HIGH}$, then $R_P$ should be greater than $R_{P\_MIN}$ and this is achieved by setting $W_P$ to its minimum value i.e. $W_{MIN}$ and $L_P$ which is calculated by solving Eqs. (3) and (4) as

$$L_P=W_{MIN}\,\mu_{eff,p}C_{ox}(V_{DD}-|V_{T,p}|)\left(\frac{1+N}{N}\frac{V_{SWING}}{I_{SS}}-\frac{R_{DSW}10^{-6}}{W_{MIN}}\right) \tag{14}$$

The small-signal voltage gain ($A_v$) (Eq. (9)) has been used to size transistors $M_{LV3,4,5,6}$. Assuming minimum channel length for the said transistors, the width is computed as

$$W_N=\frac{2}{\mu_{eff,n}C_{ox}}\left(\frac{N}{1+N}\right)^2\left(\frac{A_v}{V_{SWING}}\right)^2I_{SS}L_{MIN} \tag{15}$$

Sometimes Eq. (15) results in a value of $W_N$ smaller than the minimum channel width. This happens when the bias current is lower than the current of the minimum sized NMOS transistor, $I_{LOW}$ given as from Eq. (9)

$$I_{LOW}=\frac{1}{2}\left(\frac{1+N}{N}\right)^2\frac{W_{MIN}}{L_{MIN}}\,\mu_{eff,n}C_{ox}\left(\frac{V_{SWING}}{A_v}\right)^2 \tag{16}$$

therefore, in such cases, $W_N$ is also set to $W_{MIN}$. For proper switching, the width of transistors $M_{LV7,8}$ is made $N$ times the width of transistors $M_{LV3,4,5,6}$.

The accuracy of the static model for the proposed XOR gate is validated through SPICE simulations by using TSMC 0.18 μm CMOS process parameters and with a power supply of 1.1 V. The proposed XOR gate was designed and simulated for wide range of operating conditions: voltage swing of 300 mV and 400 mV, small-signal voltage gain of 2 and 4, $N=5$, and the bias current ranging from 10 μA to 100 μA. It is found that there is a close agreement between the simulated and the predicted values of static parameters for all the operating conditions. The error plots in the simulated values of static parameters with respect to the predicted values in particular for small-signal voltage gain of 4 and voltage swing of 300 mV and 400 mV are shown in Fig. 3a–c. It may be noted that maximum error in voltage swing, small-signal voltage gain and noise margin are 11%, 13% and 11% respectively. An error plot of noise margin for small-signal voltage gain of 2 and 4, $N=5$ and the voltage swing ranging from 0.2 V to 0.7 V is plotted in Fig. 3d. It may be observed that the maximum error in noise margin is 14%. Using the $3\sigma$ variations [16], the Monte Carlo simulations for noise margins is carried out for 300 sample runs and the results for different cases are reported in Table 2. It is found that the mean value of noise margin is always larger than the $(3\sigma)$ variations [13].

The impact of parameter variation on proposed low-voltage and traditional MCML XOR gates performance is studied at different design corners. The findings for various operating conditions are given in Table 3. It is found that the voltage swing, small-signal voltage gain, and noise margin of the proposed low-voltage XOR gate varies by a factor of 1.87, 2.94, and 2.28 respectively between the best and the worst cases. For the traditional MCML XOR gate, the voltage swing, small-signal voltage

Please cite this article as: K. Gupta, et al., Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.03.007

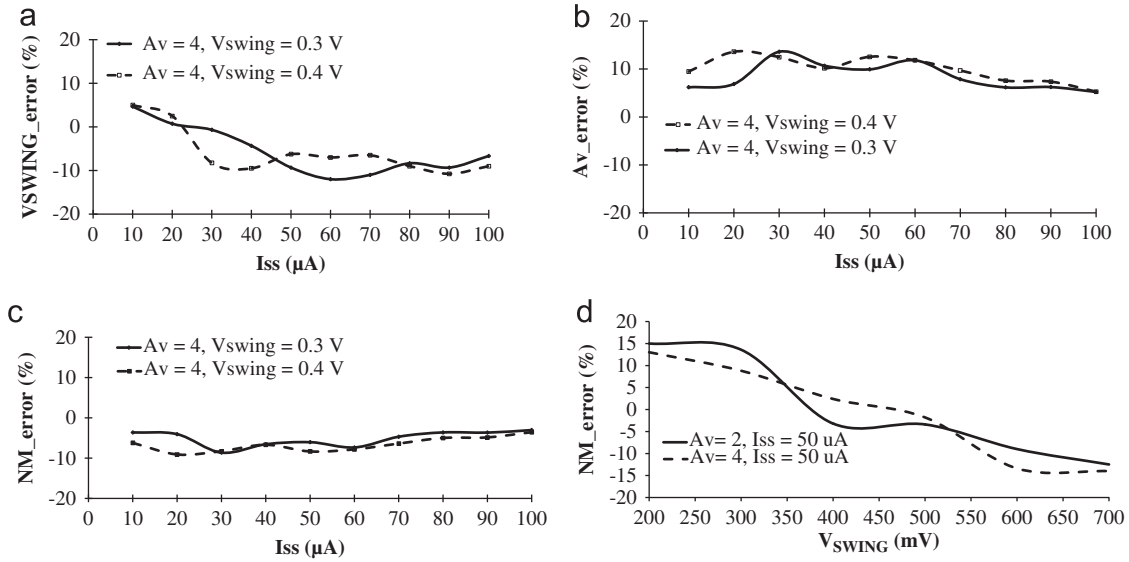



**Fig. 3.** Error in the static parameters.

**Table 2**
Results of Monte Carlo simulations (300 sample runs) on noise margin for different operating conditions.

| Operating conditions | Mean, $\mu$ (mV) | Sigma, $3\sigma$ (mV) |
|---|---|---|
| $A_v$=4, $V_{SWING}$=0.2 V | 61.8 | 39.6 |
| $A_v$=4, $V_{SWING}$=0.4 V | 128 | 46 |
| $A_v$=4, $V_{SWING}$=0.6 V | 191 | 66 |

**Table 3**
Effect of process variation on static parameters of the proposed and the traditional XOR gates.

| Parameter | NMOS / PMOS | T / T | F / F | S / S | F / S | S / F |
|---|---|---|---|---|---|---|
| **Simulation Condition:** $A_v$=4, $V_{SWING}$=0.4 V, $C_L$=50 fF, $I_{SS}$=100 μA | | | | | | |
| $V_{SWING}$ (mV) | Proposed | 344 | 481 | 260 | 430 | 350 |
| | Traditional | 366 | 465 | 267 | 378 | 370 |
| $A_V$ | Proposed | 3.1 | 2.1 | 5.2 | 3.1 | 3.1 |
| | Traditional | 3.2 | 2.1 | 4.3 | 3.1 | 3.1 |
| NM (mV) | Proposed | 94.2 | 78.5 | 94.6 | 116.6 | 95.4 |
| | Traditional | 100.6 | 76.7 | 90 | 103.1 | 101.1 |
| **Simulation Condition:** $A_v$=4, $V_{SWING}$=0.4 V, $C_L$=50 fF, $I_{SS}$=10 μA | | | | | | |
| $V_{SWING}$ (mV) | Proposed | 410 | 498 | 265 | 420 | 415 |
| | Traditional | 342 | 519 | 294 | 443 | 407 |
| $A_V$ | Proposed | 3.8 | 1.9 | 5.5 | 2.9 | 3.7 |
| | Traditional | 2.98 | 1.81 | 4.39 | 2.67 | 2.81 |
| NM (mV) | Proposed | 130.2 | 63.6 | 98.9 | 110.6 | 129.4 |
| | Traditional | 89.8 | 56.7 | 99.6 | 104.2 | 101.1 |

Where different design corners are denoted by T=Typical, F=Fast, S=Slow.

gain, and noise margin varies by a factor of 1.76, 2.42, and 1.8 respectively between the best and the worst cases. Thus, the proposed low-voltage XOR gate show slightly higher variations than the traditional MCML XOR gate for different design corners which can be attributed to the smaller aspect ratio of transistors in the proposed low-voltage XOR gate [16]. To demonstrate the effect of mismatch, the width of NMOS and PMOS transistors of the proposed low-voltage and the traditional MCML XOR gates are varied by 10%. The variation in voltage swing, small-signal voltage gain, noise margin are within 3% for both the gates.

### 3.3. Delay model

In this section, a delay model of the proposed low-voltage XOR gate is formulated in terms of bias current and the voltage swing. In case of a high-to-low transition on B input that causes output to switch by activating (deactivating) the transistor pair $M_{LV3}$–$M_{LV4}$ ($M_{LV5}$–$M_{LV6}$), the circuit reduces to a simple MCML inverter. The equivalent linear half circuit is shown in Fig. 4 where $C_{gdi}$ and $C_{dbi}$represents the gate–drain capacitance and the drain–bulk junction capacitance of the $i$th transistor. For NMOS transistors operating in saturation region, $C_{gd}$ is equal to the overlap capacitance $C_{gdo}W_n$ between the gate and the drain [30]. For the PMOS transistor operating in linear region, $C_{gd}$ is evaluated as the sum of the overlap capacitance and the intrinsic contribution associated with its channel charge [30]. The junction capacitance $C_{db}$ for the transistors are computed as explained in [20].

The delay of the proposed XOR gate can be expressed as

$$t_{PD} = 0.69\, R_P(C_{db3} + C_{gd3} + C_{gd9} + C_{db9} + C_{db5} + C_{gd5} + C_L) \quad (17)$$

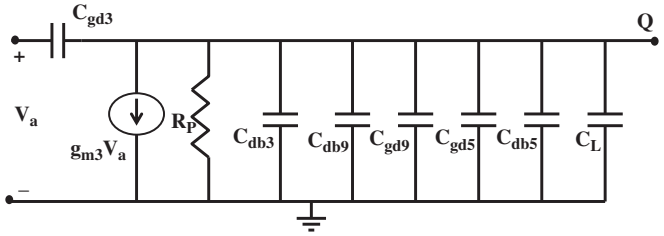



**Fig. 4.** Linear half-circuit (with low value of differential input A).

with $C_{db3} = C_{db5}$, $C_{gd3} = C_{gd5}$ and, $R_P = \frac{1+N}{N}\frac{V_{SWING}}{I_{SS}}$, Eq. (17) can be rewritten as

$$t_{PD} = 0.69\,\frac{1+N}{N}\frac{V_{SWING}}{I_{SS}}(2C_{db3} + 2C_{gd3} + C_{gd9} + C_{db9} + C_L) \quad (18)$$

The capacitances may be expressed in terms of the bias current and the voltage swing as

$$C_{xy} = \frac{a_{xy}}{(V_{SWING})^2} I_{SS} + b_{xy}\frac{V_{SWING}}{I_{SS}} + c_{xy} \quad (19)$$

Please cite this article as: K. Gupta, et al., Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.03.007

**Table 4**
Coefficients of the capacitances for the proposed XOR gate.

| | |
|---|---|
| **NMOS coefficients** | |
| $a_{db3}$ | $\frac{2A_v^2 L_{MIN}}{\mu_{eff,n}C_{ox}}\left(\frac{N}{1+N}\right)^2(K_{jn}C_{jn}L_{dn}+2K_{jswn}C_{jswn})$ |
| $a_{gd3}$ | $2A_v^2 C_{gdo}\left(\frac{N}{1+N}\right)^2\frac{L_{MIN}}{\mu_{eff,n}C_{ox}}$ |
| $c_{db3}$ | $2K_{jswn}C_{jswn}L_{dn}$ |
| $b_{db3}$, $b_{gd3}$, and $c_{gd3}$ | 0 |
| **PMOS coefficients** | |
| $b_{gd9}$ | $\frac{3}{4}\left(\frac{1+N}{N}\right)A_{bulk,max}\,\mu_{eff,p}C_{ox}^2W_{MIN}^2(V_{DD}-\lvert V_{T,p}\rvert)$ |
| $c_{gd9}$ | $C_{gdo}W_{MIN}-\frac{3}{4}A_{bulk,max}\,\mu_{eff,p}C_{ox}^2W_{MIN}(V_{DD}-\lvert V_{T,p}\rvert)R_{DSW}10^{-6}$ |
| $c_{db9}$ | $K_{jp}C_{jp}L_{dp}W_{MIN}+2K_{jswp}C_{jswp}(L_{dp}+W_{MIN})$ |
| $a_{gd9}$, $a_{db9}$, and $b_{db9}$ | 0 |

where the symbols have their usual meaning.

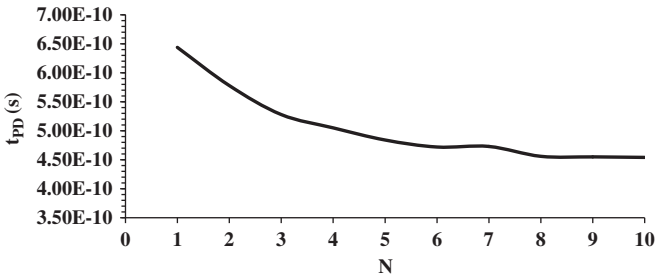



**Fig. 5.** Propagation delay vs. *N*.

where $C_{xy}$ is the capacitance between the terminals $x$ and $y$ and $a_{xy}$, $b_{xy}$, $c_{xy}$ are the associated coefficients. Using Eqs. (14) and (15), various capacitances in Eq. (18) for $I_{SS}$ ranging from $I_{LOW}$ to $I_{HIGH}$ may be expressed as

$$C_{gd3}=C_{gdo}W_3=2A_v^2C_{gdo}\left(\frac{N}{1+N}\right)^2\frac{L_{MIN}}{\mu_{eff,n}C_{OX}}\frac{I_{SS}}{(V_{SWING})^2} \tag{20}$$

where $C_{gdo}$ is the drain–gate overlap capacitance per unit transistor width.

$$C_{db3}=W_3(K_{jn}C_{jn}L_{dn}+2K_{jswn}C_{jswn})+2K_{jswn}C_{jswn}L_{dn} \tag{21}$$

$$=2A_v^2\frac{L_{MIN}}{\mu_{eff,n}C_{OX}}\left(\frac{N}{1+N}\right)^2(K_{jn}C_{jn}L_{dn}+2K_{jswn}C_{jswn})\frac{I_{SS}}{(V_{SWING})^2}+2K_{jswn}C_{jswn}L_{dn} \tag{22}$$

where $C_{jn}$, and $C_{jswn}$ are the zero-bias junction capacitance per unit area and zero-bias sidewall capacitance per unit parameter respectively. The coefficients $K_{jn}$, and $K_{jswn}$ are the voltage equivalence factor for the junction and the sidewall capacitances [20].

$$C_{gd9}=C_{gdo}W_{MIN}+\tfrac{3}{4}A_{bulk,max}\,W_{MIN}L_PC_{ox} \tag{23}$$

$$=C_{gdo}W_{MIN}+\tfrac{3}{4}A_{bulk,max}\,W_{MIN}C_{ox}\times\left\{\mu_{eff,p}C_{ox}W_{MIN}(V_{DD}-\left|V_{T,\,p}\right|)\left[\frac{1+N}{N}\frac{V_{SWING}}{I_{SS}}-\frac{R_{DSW}10^{-6}}{W_{MIN}}\right]\right\} \tag{24}$$

where $A_{bulk,max}$ is a parameter defined in the BSIM3v3 model [28].

$$C_{db9}=W_{MIN}(K_{jp}C_{jp}L_{dp}+2K_{jswp}C_{jswp})+2K_{jswp}C_{jswp}L_{dp} \tag{25}$$

The coefficients $a_{xy}$, $b_{xy}$ and $c_{xy}$ of all the capacitances in Eq. (18) are summarized in Table 4. Using Eqs. (20)–(25), (18) can be written as

$$t_{PD}=0.69\,\frac{1+N}{N}V_{SWING}\left(\frac{a}{V_{SWING}{}^2}+b\frac{V_{SWING}}{I_{SS}^2}+\frac{c+C_L}{I_{SS}}\right) \tag{26}$$

where

$$a=2a_{db3}+2a_{gd3} \tag{27a}$$

$$b=b_{gd9} \tag{27b}$$

$$c=2c_{db3}+c_{gd9}+c_{db9} \tag{27c}$$

The delay model can also be used for $I_{SS}$ value outside the range [$I_{LOW}$ and $I_{HIGH}$]. This is because for $I_{SS}>I_{HIGH}$, the capacitance coefficients of PMOS transistor in Eq. (26) differ as explained in Section 3.2. But since for high values of $I_{SS}$, the capacitive contribution of PMOS transistor is negligible therefore Eq. (26) can predict the delay. Similarly, for $I_{SS}<I_{LOW}$, the capacitance coefficients of NMOS transistor in Eq. (26) differs. But since for low values of $I_{SS}$, the delay majorly depends on the capacitances of PMOS transistor. So, the expression (Eq. (26)) can estimate the delay of the proposed XOR gate.

The propagation delay of the proposed gate is plotted in Fig. 5 as a function of $N$ for $A_v=4$, NM=130 mV, $I_{SS}=50$ μA and $C_L=50$ fF. It is found out that the delay asymptotically reaches to value of 455 ps. However, a high value of $N$ will result in larger transistor sizes of $M_{LV7,8}$ thus increasing the input capacitance seen from input B. Therefore, a good compromise between the two opposing requirements is to set $N=5$ as after which improvement in speed is not significant. Similar results are obtained for other operating conditions.

The accuracy of the delay model for the proposed XOR gate is validated through SPICE simulations by using the TSMC 0.18 μm CMOS process parameters and with a power supply of 1.1 V. The proposed XOR gate was designed for wide range of operating conditions: voltage swing of 300 mV and 400 mV, small-signal voltage gain of 2 and 4, the bias current ranging from 10 μA to 100 μA, $N=5$, and load capacitances of 0 fF, 10 fF, 100 fF, 1 pF and fan out of 4. It is found that there is a close agreement between the simulated and the predicted delay for all the operating conditions. The simulated and the predicted delay in particular for $V_{SWING}=400$ mV, $A_v=4$ and with different load capacitances are plotted in Fig. 6.

The impact of parameter variation on proposed low-voltage and traditional MCML XOR gates delay is studied at different design corners. The findings for various operating conditions are given in Table 5. It is found that the propagation delay of the proposed low-voltage XOR gate varies by a factor of 1.89 between the best and the worst cases. For the traditional MCML XOR gate, the delay varies by a factor of 1.85 between the best and the worst cases. Thus, the proposed low-voltage XOR gate show slightly higher variation than the traditional MCML XOR gate in delay for different design corners which can be attributed to the smaller aspect ratio of transistors in the proposed low-voltage XOR gate [16]. To demonstrate the effect of mismatch the width of NMOS and PMOS transistors of the proposed low-voltage and the traditional MCML XOR gates are varied by 10%. The variation in the propagation delay is within 8% for both the gates.

## 4. Design cases

In the previous section, the proposed XOR gate has been modeled and different parameters are expressed as a function of bias current and voltage swing. In practice, the voltage swing is set on the basis of the specified noise margin while the bias current is chosen according to power-delay considerations. Therefore, the

Please cite this article as: K. Gupta, et al., Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.03.007

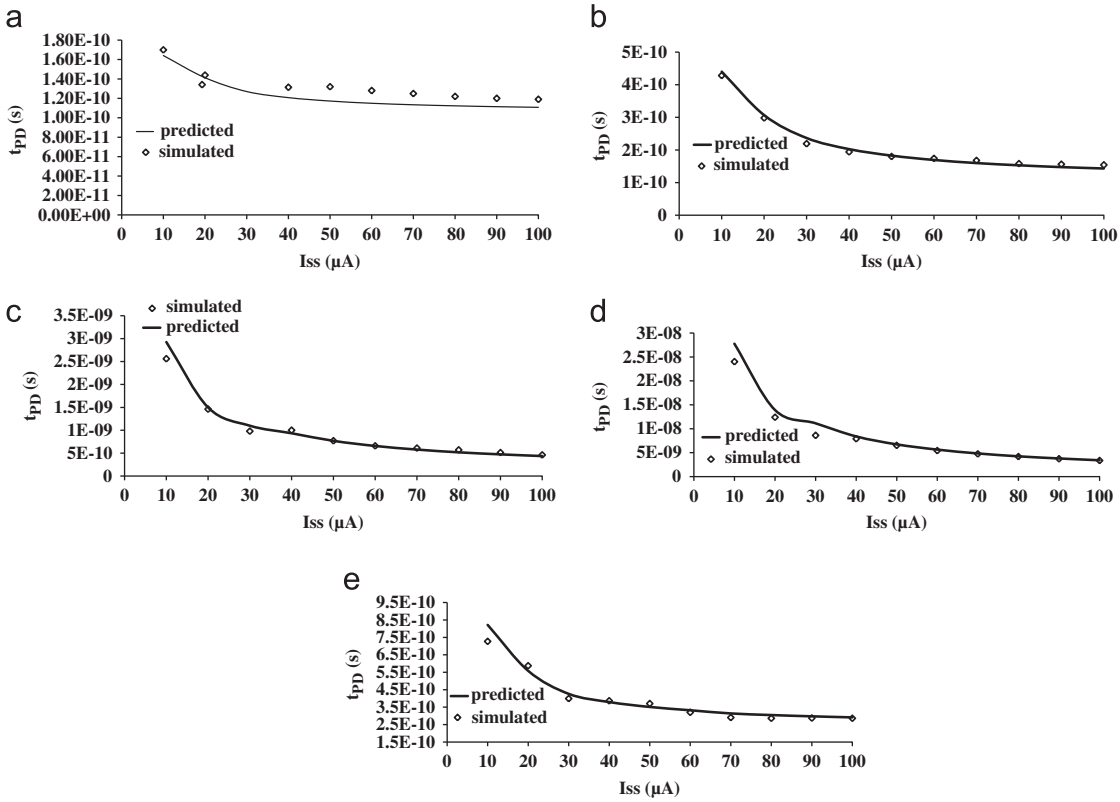



**Fig. 6.** Simulated and predicated delay of the proposed XOR gate vs. $I_{SS}$ with NM=130 mV, $A_v$=4 for different $C_L$ values (a) 0 fF, (b)10 fF, (c) 100 fF, (d) 1 pF and (e) FO=4.

**Table 5**
Effect of process variation on the delay of the proposed and the traditional XOR gates.

| Parameter | NMOS<br>PMOS | T<br>T | F<br>F | S<br>S | F<br>S | S<br>F |
|---|---|---|---|---|---|---|
| **Simulation Condition:** $A_v$=4, $V_{SWING}$=0.4 V, $C_L$=50 fF, $I_{SS}$=100 μA | | | | | | |
| $t_{PD}$ **(ps)** | Proposed | 265 | 237 | 448 | 255 | 262 |
| | Traditional | 553 | 515 | 954 | 527 | 550 |
| **Simulation Condition:** $A_v$=4, $V_{SWING}$=0.4 V, $C_L$=50 fF, $I_{SS}$=10 μA | | | | | | |
| $t_{PD}$ **(ns)** | Proposed | 2.4 | 1.7 | 3.2 | 2.1 | 2.3 |
| | Traditional | 3.7 | 3.2 | 4.6 | 3.5 | 3.6 |

proposed low-voltage XOR gate for high-speed, power-efficient, and low-power design cases is discussed.

### 4.1. High-speed design

A high-speed design requires bias current that results in minimum delay. The delay (Eq. (26)) decreases with the increasing $I_{SS}$ and tends to an asymptotic minimum value of $0.69\frac{1+N}{N}\frac{a}{V_{SWING}}$ for $I_{SS}\to\infty$. A substantial improvement in delay with increasing bias current is achieved if condition

$$\frac{a}{{V_{SWING}}^2} \geq b\frac{V_{SWING}}{I_{SS}^2} + \frac{c+C_L}{I_{SS}} \tag{28}$$

is satisfied. However, high value of bias current results in large transistor sizes. Therefore, the bias current should be set to such a value after which the improvement in speed is not significant. If equality sign in Eq. (28) is considered then the delay is close to its minimum value and the use of high bias current is avoided. Therefore, this assumption leads to a bias current ($I_{SS\_HS}$) and delay ($t_{PD\_MIN}$) as

$$I_{SS\_HS} = \frac{c+C_L}{2a}V_{SWING}^2\left(1+\sqrt{1+4\frac{ab}{(c+C_L)^2}\frac{1}{V_{SWING}}}\right) \tag{29}$$

$$t_{PD\_MIN} = 2\times 0.69\frac{1+N}{N}\frac{a}{V_{SWING}} \tag{30}$$

The proposed high-speed XOR gate designed with a power supply of 1.1 V, noise margin of 130 mV, small-signal gain of 4, $N$=5 and load capacitance of 50 fF, gives $I_{SS\_HS}$ as 160 μA. A delay of 194 ps and 211 ps is obtained from Eq. (30) and simulations respectively. On the contrary, a traditional high-speed XOR gate designed using the method outlined in [28] and with power supply of 1.4 V for the same specifications results in a delay of 528 ps. This indicates that the proposed XOR gate can achieve much higher speed than the traditional one.

### 4.2. Power-efficient design

A power-efficient design requires bias current that results in minimum power-delay product (PDP). The power is calculated as the product of $V_{DD}$ and $I_{SS}$. So, the PDP of the proposed XOR gate may be expressed as

$$\mathrm{PDP} = 0.69V_{DD}V_{SWING}\frac{1+N}{N}\left(\frac{a}{{V_{SWING}}^2}I_{SS} + b\frac{V_{SWING}}{I_{SS}} + c + C_L\right) \tag{31}$$

Therefore, the current $I_{SS\_PDP}$ for minimum PDP may be given as

$$I_{SS\_PDP} = \sqrt{\frac{b}{a}}(V_{SWING})^{3/2} \tag{32}$$

**Table 6**
Summary of the design cases for the proposed and the traditional XOR gates.

| Design case | Performance parameter | Proposed XOR gate | Traditional XOR gate |
|---|---|---|---|
| High speed | Delay | Low | High |
| Power-efficient | Power delay product | High | Low |
| Low power | Power | Low | High |

Accordingly, the minimum PDP results to

$$\mathrm{PDP} = 0.69\, V_{\mathrm{DD}} V_{\mathrm{SWING}} \frac{1+N}{N}\left(\frac{2\sqrt{ab}}{\sqrt{V_{\mathrm{SWING}}}} + c + C_{\mathrm{L}}\right) \tag{33}$$

The proposed power-efficient XOR gate designed with a power supply of 1.1 V, noise margin of 130 mV, small signal gain of 4, $N=5$ and load capacitance of 50 fF, gives $I_{\mathrm{SS\_PDP}}$ as 5.8 μA. A PDP value of 32 fJ has been obtained for the proposed XOR gate. On the other hand, a traditional power-efficient XOR gate designed using the method outlined in [28] and with power supply of 1.4 V for the same specifications results in a PDP value of 13 fJ. The result signifies that the proposed XOR gate results in higher PDP values than the traditional one.

### 4.3. Low-power design

In low-power designs, the bias current $I_{\mathrm{SS}}$ is set to low values so that the term $b(V_{\mathrm{SWING}}/I_{\mathrm{SS}}^2)$ is dominant in Eq. (26). Hence, the delay reduces to

$$t_{\mathrm{PD}} = 0.69b\left(\frac{1+N}{N}\right)\left(\frac{V_{\mathrm{SWING}}}{I_{\mathrm{SS}}}\right)^2 \tag{34}$$

The proposed low-power XOR gate designed with a power supply of 1.1 V, noise margin of 130 mV, small signal gain of 4, load capacitance of 5 fF, $N=5$ and with value of $I_{\mathrm{SS}}$ as 2 μA gives a power consumption of 2.2 μW while the traditional XOR gate designed using the method outlined in [28] and with power supply of 1.4 V for the same specifications results in power consumption of 2.8 μW with lower delay values than the proposed gate.

The comparison of the characteristics of the proposed low-voltage XOR gate with the traditional MCML XOR gate for different design cases are summarized in Table 6.

## 5. Conclusions

A new low-voltage MCML XOR gate based on the triple-tail cell concept is proposed. Its static parameters are analytically modeled and are used to develop a design approach for the proposed low-voltage MCML XOR gate. The delay is formulated as a function of the bias current and the voltage swing and is traded off with power consumption for high-speed, power-efficient, and low-power design cases. An improvement in performance is obtained for the proposed low-voltage XOR gate in comparison to traditional MCML XOR gate for high-speed and low-power design cases.

## References

[1] S. Jantzi, K. Martin, A. Sedra, Quadrature bandpass ΣΔ modulator for digital radio, IEEE J. Solid-State Circuits 32 (1997) 1935–1949.
[2] S. Luschas, R. Schreier, H.S. Lee, Radio frequency digital-to-analog converter, IEEE J. Solid-State Circuits 39 (2004) 1462–1467.
[3] B. Kup, E. Dijkmans, P. Naus, J. Sneep, A bit-stream digital-to-analog converter with 18-b resolution, IEEE J. Solid-State Circuits 26 (1991) 1757–1763.
[4] T. Takaramoto, S. Harajiri, M. Sawada, O. Kobayashi, K. Gotoh, A. bonded-SOI-wafer, CMOS 16-bit 50-KSPS delta-sigma ADC, in: Proceedings of IEEE Custom Integrated Circuit Conference, 1991, pp. 18.1.1–18.1.4.
[5] N. Weste, K. Eshraghian, Principles of CMOS VLSI Design: A system perspective Reading, Addison-Wesley, MA, 1993.
[6] D. Allstot, S. Chee, S. Kiaei, M. Shristawa, Folded source-coupled logic vs. CMOS static logic for low-noise mixed-signal ICs, IEEE Trans. Circuits Syst. I 40 (1993) 553–563.
[7] C. Choy, C. Chan, M. Ku, J. Povazanec, Design procedure of low noise high-speed adaptive output drivers, in: Proceedings of the International Symposium on Circuits and Systems, 1997, pp.1796–1799.
[8] S. Kiaei, D. Allstot, Low-noise logic for mixed-mode VLSI circuits, Microelectron. J. 23 (1992) 103–114.
[9] R. Sàez, M. Kayal, M. Declercq, M. Schneider, Digital circuit techniques for mixed analog/digital circuits applications, in: Proceedings of International Conference on Electronics, Circuits, and Systems, 1996, pp. 956–959.
[10] H. Ng, D. Allstot, CMOS current steering logic for low-voltage mixed-signal integrated circuits,, IEEE Trans. VLSI Syst. **5** (1997) 301–308.
[11] J. Kundan, S. Hasan, Enhanced folded source-coupled logic technique for low-voltage mixed-signal integrated circuits, IEEE Trans. Circuits Syst.-II 47 (2000) 810–817.
[12] M. Yamashina, H. Yamada, An MOS current mode logic (MCML) circuit for low-power sub-GHz processors, IEICE Trans. Electron. E75-C (1992) 1181–1187.
[13] S. Bruma, Impact of on-chip process variations on MCML performance, in: Proceedings of IEEE Conference on Systems-on-Chip, 2003, pp. 135–140.
[14] J.M. Musicer, J. Rabaey, MOS current mode logic for low power, low noise, CORDIC computation in mixed-signal environments, in: Proceedings of the International Symposium of Low Power Electronics and Design, 2000, pp. 102–107.
[15] K. Zhou, S. Chen, A. Rucinski, J.F. McDonald, T. Zhang, Self-timed triple-rail MOS current mode logic pipeline for power-on-demand design, in: Proceedings of IEEE Symposium on circuits and systems, 2005, pp.1394–1397.
[16] H. Hassan, M. Anis, M. Elmasry, MOS current mode circuits: analysis, design, and variability, IEEE Trans. Very Large Scale Integration (VLSI) Syst. 13 (2005) 885–898.
[17] M. Alioto, R. Mita, G. Palumbo, Design of high-speed power efficient MOS current mode logic frequency dividers, IEEE Trans. Circuits Syst.—II: Express Briefs 53 (2006) 1165–1169.
[18] V. Srinivasan, S.H. Dong, J.B. Sulistyo, Gigahertz-Range MCML multiplier architectures, in: Proceedings of the International Symposium on Circuits and Systems, 2004, pp.785–788.
[19] Y. Delcan, A. Morgul, High Performance 16-bit MCML multiplier, in: Proceedings of the IEEE European Conference on Circuit Theory and Design, 2009, pp.157–160.
[20] J. Rabaey, Digital Integrated Circuits (A Design Perspective), Prentice-Hall, Englewood Cliffs, NJ, 1996.
[21] M. Alioto, G. Palumbo, Model and Design of Bipolar and MOS Current-Mode logic (CML, ECL and SCL Digital Circuits), Springer, Neitherlands, 2005.
[22] J. Savoj, B. Razavi, A 10-Gb/s CMOS clock and data recovery circuit with a half-rate linear phase detector, IEEE J. Solid-State Circuits 36 (2001) 761–767.
[23] A. Tajalli, M. Atarodi, Linear phase detection using two-phase latch, Electronics Letter 39 (24) (2003) 1695–1696.
[24] M. Alioto, L. Pancioni, S. Rocchi, V. Vignoli, Modeling and evaluation of positive-feedback source-coupled logic, IEEE Trans. Circuits Syst.—I, Regular Pap. 51 (2004) 2345–2355.
[25] K. Kimura, Circuit design techniques for very low-voltage analog functional blocks using triple-tail cells, IEEE Trans. Circuits Syst.— I: Fundam. Theory Appl. 42 (1995) 873–885.
[26] F. Matsumoto, Y. Noguchi, Linear bipolar OTAs based on a triple-tail cell employing exponential circuits, IEEE Trans. Circuits Syst.—II: Express Briefs, 51, 670–674.
[27] M. Alioto, R. Mita, G. Palumbo, Performance evaluation of the low-voltage CML D-latch topology, Integration, VLSI J. 36 (2003) 191–209.
[28] M. Alioto, G. Palumbo, Power-delay optimization of D-latch/MUX source coupled logic gates, Int. J. Circuit Theory Appli. 33 (2005) 65–85.
[29] H. Hassan, M. Anis, M. Elmasry, Analysis and design of low-power multi-threshold MCML, in: Proceedings of the IEEE International Conference on System-on-Chip, 2004, pp. 25-29.
[30] M. Alioto, G. Palumbo, S. Pennisi, Modelling of source-coupled logic gates, Int. J. Circuit Theory Appl. 30 (2002) 459–477.

Please cite this article as: K. Gupta, et al., Analysis and design of MOS current mode logic exclusive-OR gate using triple-tail cells, Microelectron. J (2013), http://dx.doi.org/10.1016/j.mejo.2013.03.007

International Journal of Fuzzy Logic Systems (IJFLS) Vol.3, No2, April 2013

# APPLICATION OF FUZZY LOGIC TO VISUAL EXAMINATION IN THE ASSESSMENT OF SULPHATE ATTACK ON CEMENT BASED MATERIALS

Alok Verma[1], Mukesh Shukla[2] and Anil Kumar Sahu[1]

[1]Department of Civil Engineering, Delhi Technological University, Delhi, INDIA
alokverma_dce@hotmail.com

[2]Bundelkhand Institute of Engineering & Technology, Jhansi, U.P., INDIA

***ABSTRACT***

*Sulphate attack on cement based materials has been widely taken up in research studies focusing on aggressive environmental conditions. Visual observations are considered as an important first step to assess the performance of the materials under the severity of the involved conditions. Correlations of visual observations (supported by other tests, during controlled laboratory experiments) and conditions in the field are essential to assess the performance of materials in the field. Due to uncertainties in the determination of various performance related parameters in the field, a rigid reliance on the results of visual examination conducted in the laboratory is not possible. This paper considers some aspects related to visual observations concerning the performance assessment of cement based materials under sulphate attack conditions in the field. Through a theoretical study it explores the use of fuzzy logic in analysing how much reasonable confidence should be placed on visual observations as well as in determining relative importance of various parameters.*

***KEYWORDS***

*Visual observations, Visual ratings, Concrete, Sulphate attack, Deterioration, Fuzzy logic*

## 1. INTRODUCTION

Sulphate attack, constituting a major risk of chemical aggression for concrete and other building materials, has been taken up vigorously in the research studies [1]. It is reported that sulphate attack is difficult to measure because it is the result of a complex set of chemical processes and there is still a lot of controversy about the mechanism of such attack [2]. Samples of cement based building materials, such as mortar and concrete, are exposed to artificially created sulphate environments and are tested in laboratories [3]. While the results of these short-term accelerated tests help to give due indication of the likely performance of such materials in field conditions, these methods are criticized on many counts. They are perceived to change the attack mechanism and make it differ from the one which exists in the field under consideration [3]. Sometimes, laboratory tests may not simulate the conditions in the field correctly [4]. Visual examination of specimens is very important in such tests to correlate laboratory and field conditions [4-15].

Usually cubical samples of concrete or cement mortar of various specifications are kept in different types of aggressive acidic or alkaline solutions in laboratory for some time. These aggressive solutions model aggressive environmental conditions in which actual structures made of concrete or cement mortar, exist. Effects of such aggressive solutions on these laboratory samples are observed visually. These effects may be in the form of Appearance of fine cracks on

DOI : 10.5121/ijfls.2013.3204 

the surface of specimen [16], formation of subsurface cracks [17], spalling of surface material [6,18,19], crazing [20], blistering of surface [18,21], expansion in length [18], loosening of material [19], corner cracking combined with transverse surface cracks [6], swelling of corners [22], a complete breakdown of samples [16], appearance of soft pulpy mass of mortar [4,16], debonding of matrix from aggregate particles [19], extreme distress [6], onion peeling type degradation of samples [23], change in colour [19], erosion of faces [18,21], softening of external layer of mortar specimen [16], loss of cohesion of mortar [7], formation of gypsum crystals [19,21,24], formation of efflorescence on the surface [19,21] and formation of mushy layer on the surface [25]. If sufficient data is available, durability state of a material in the field can be assessed by comparing the visual effects and the results of laboratory experiments to those observed in the field. In this direction, some visual ratings have been proposed under various conditions of laboratory tests. Visual rating of samples is given in order to find the most suitable and easiest way of detecting and quantifying damages occurring due to various effects [35]. These visual ratings categorise visual effects and link these effects to relative extent of damage in an aggressive environmental condition. Many visual ratings have been provided considering different grades of cement based materials and conditions of exposure to sulphate environments [2,6,8,9,12,13,16,17,20,34,36-49]. A ten point grading of various stages of damage for limestone cement mortars proposed in a study [31], is shown in Table 1.

| **Visual rating** | **Recorded effects from visual examination** |
|---|---|
| 0 | No visible deterioration |
| 1 | Some deterioration at corners |
| 2 | Deterioration at corners |
| 3 | Deterioration at corners and some cracking along the edges |
| 4 | Deterioration at corners and cracking along the edges |
| 5 | Cracking and expansion |
| 6 | Bulging of surfaces |
| 7 | Extensive cracking and expansion |
| 8 | Extensive spalling |
| 9 | Complete damage |

Table 1. Ten Point Grading

Without doubting the importance of visual examination in the assessment of sulphate attack, it has been reported that sole visual inspection can be misleading in some situations [26]. Therefore, visual observations are sometimes supported by advanced tests also [1,3,5,8,16,17,27-31,33,34]. Results of these tests and visual examination are correlated and calibrated with effects seen in actual structures.

## 2. RESEARCH SIGNIFICANCE

Visual examination is an important tool for the assessment of the effects of an aggressive environment on cement based materials and for interpreting the visual effects seen in the field. Still, the findings of the examination may not be treated as fully conclusive and final in view of many uncertainties. Though theoretically the use of visual ratings to assess the conditions and performance of materials should be an easy task, it is not so due to variations of conditions in respect of material composition & characteristics, site, exposure and other related aspects. Time variations of these factors in these combinations would make the exact correspondence of conditions and effects very difficult to be indicated. It has been reported that there is no universally accepted criterion for measuring failure of laboratory specimens exposed to sulphate [2-4,18]. Though it is simple to make visual observations in laboratory experiments of sulphate

attack, it is difficult to rigidly correlate and apply the qualitative conclusions to field conditions [23,24]. In this connection, importance of longer term exposure is found necessary to obtain more realistic indications of durability for extended service life [44-47]. Importance of application of Fuzzy Logic system along with in interpreting results of laboratory visual examination in field conditions has been emphasized, in view of uncertainties of effects, varying conditions of tests in laboratory as well as in the field and nomenclature of visual effects [50]. Experiences from tests, already undertaken in the past, would be helpful in this respect [51-64].
Many factors are involved in the modelling of the sulphate aggressive environmental conditions and in the correlations between the laboratory and field conditions. It is important to have knowledge of relative importance of these factors as well as to determine how much confidence may be placed on the results of visual observations, made in the laboratory, in the assessment of durability of structures in the field environment. This study considers these issues with the help of fuzzy logic.

# 3. DETAILS OF STUDY

Fuzzy set theory was developed by Lotfi Zadeh in 1965 to deal with the imprecision and uncertainty often present in real-world applications [65]. It was subsequently further developed by Mamdani and several other researchers. Fuzzy system is a logical system, an extension of multi-valued logic, which is synonymous with the theory of fuzzy sets. The theory of fuzzy sets relates to the classes of objects with unsharp boundaries in which membership is a matter of degree. Similar types of conditions exist in the cases of visual examination of deterioration effects in the field and in the interpretation of field conditions such as sulphate exposure etc.

In this study, it was intended to find out the accuracy of results obtained after visual examination conducted for sulphate attack on cement based materials. The accuracy is taken to be in terms of confidence which may be placed in the output (i.e., results of visual examination) when uncertainties exist in some parameters which may affect the determination of effects using some visual ratings. Three input parameters which may have uncertainties in their evaluation are considered. These are: Type of cation in the sulphate solution present in the ground water, quality of concrete and quality of visual observation. The quality of visual observation may be dependent on the expertise of an individual at the helm of affairs for assessing the effects of aggressive exposure on the cement based material. The output called, confidence, is treated to be an indicator of the confidence which can be placed on the correctness of the results. Mamdani type inference system in Matlab was used in this study. Three stages of accuracy of prediction in all the three input values & the output value were considered. Rules for deducing conclusions were written linking various stages of all the input and output parameters. The confidence in the output was increased by one degree if any two or more than two of the three input values were upgraded to a higher grade of accuracy. These input and output parameters, their stages and rules are provided in the fuzzy inference system, given below, which was used in the study. It is to be appreciated that various inputs are considered in terms of our confidence in knowing their values accurately. For example, the term ‘severe’ may not always mean that the exposure is severe. As it is the third stage of the input parameter, it only shows that the exposure is known with certainty. Names of different stages connected with their input names were written only for the sake of easy identification. Otherwise, in a case of confusion three stages of all parameters may be considered as stage 1, 2 and 3, with 1, 2 and 3 showing less accuracy, average accuracy and good accuracy. In a way, this study tried to determine the accuracy of output (called, confidence) depending on the accuracy in the determination of three input parameters. Considering that a first idea at a time instant may influence visual observations, a triangular pattern of membership functions was chosen for it. Membership function for quality of concrete was chosen in the form of a gauss variation because of this form of variation for the strength of concrete in practice. The trapezoidal

pattern was chosen for the output. For exposure conditions, it was taken to be trapezoidal considering it to be in-between the triangular and Gaussian types of variation. For any parameter a value '0' meant complete uncertainty and '1' a very good accuracy. Figure 1 shows the interlinking of input, rules and the output.

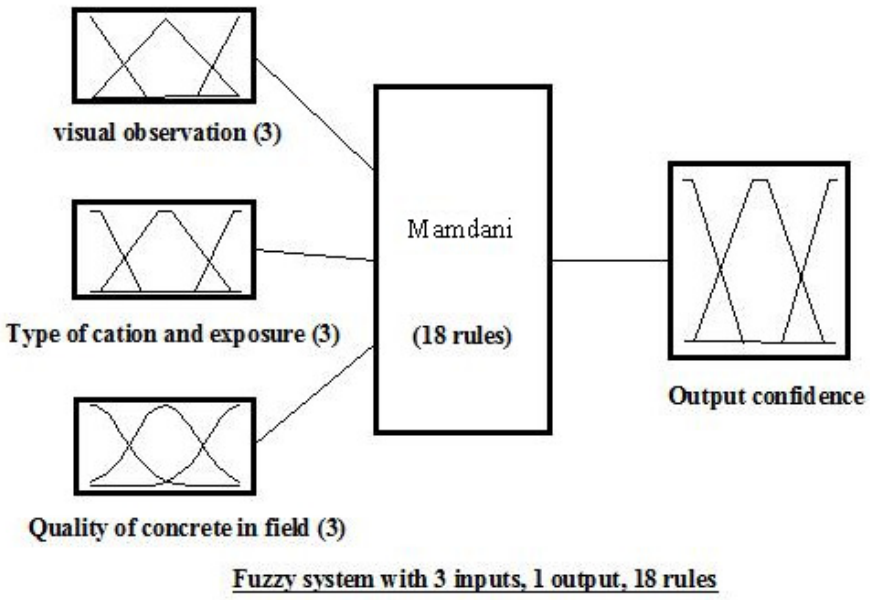


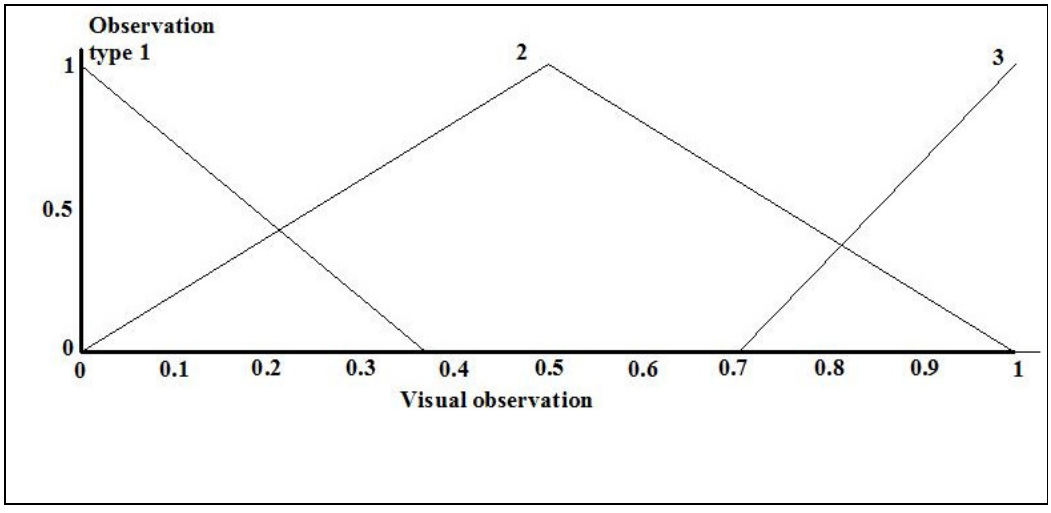

Figure 1. Interlinking of Inputs, Rules and Output in the Fuzzy System

Fuzzy Inference System used in the study is given below.

```
Name='visual5-2'
Type='mamdani'
Version=2.0
NumInputs=3
NumOutputs=1
NumRules=18
AndMethod='min'
OrMethod='max'
ImpMethod='min'
AggMethod='max'
DefuzzMethod='centroid'
[Input1]
Name='visual_observation'
Range=[0 1]
NumMFs=3
MF1='obs_type_1':'trimf',[-0.5 0 0.368]
MF2='obs_type_2':'trimf',[0 0.5 1]
MF3='obs_type_3':'trimf',[0.705 1 1.5]
[Input2]
Name='type_of_cation_and_exposure'
```

Range=[0 1]
NumMFs=3
MF1='mild':'trapmf',[-0.45 -0.05 0.05 0.34]
MF2='moderate':'trapmf',[0.05 0.45 0.55 0.95]
MF3='severe':'trapmf',[0.692 0.95 1.05 1.45]
[Input3]
Name='quality_of_concrete_in_field'
Range=[0 1]
NumMFs=3
MF1='poor':'gaussmf',[0.2123 0]
MF2='average_quality':'gaussmf',[0.2123 0.5]
MF3='good_quality':'gaussmf',[0.2123 1]
[Output1]
Name='confidence'
Range=[0 1]
NumMFs=3
MF1='poor_confidence':'trapmf',[-0.45 -0.05 0.05 0.395]
MF2='average_confidence':'trapmf',[0.05 0.45 0.55 0.95]
MF3='good_confidence':'trapmf',[0.648 0.974 1.07 1.47]
[Rules]
1 1 1, 1 (1) : 1 1 2 2, 2 (1) : 1 1 2 3, 2 (1) : 1 1 3 2, 2 (1) : 1 2 1 1, 1 (1) : 1
2 2 1, 1 (1) : 1 2 1 2, 1 (1) : 1 2 2 2, 2 (1) : 1 2 2 3, 2 (1) : 1 2 3 2, 2 (1) : 1
2 3 3, 3 (1) : 1 3 1 1, 1 (1) : 1 3 2 1, 1 (1) : 1 3 1 2, 1 (1) : 1 3 2 2, 2 (1) : 1
3 3 2, 2 (1) : 1 3 2 3, 2 (1) : 1 3 3 3, 3 (1) :

Membership functions showing different stages of the input and output parameters are shown in Figure 2.



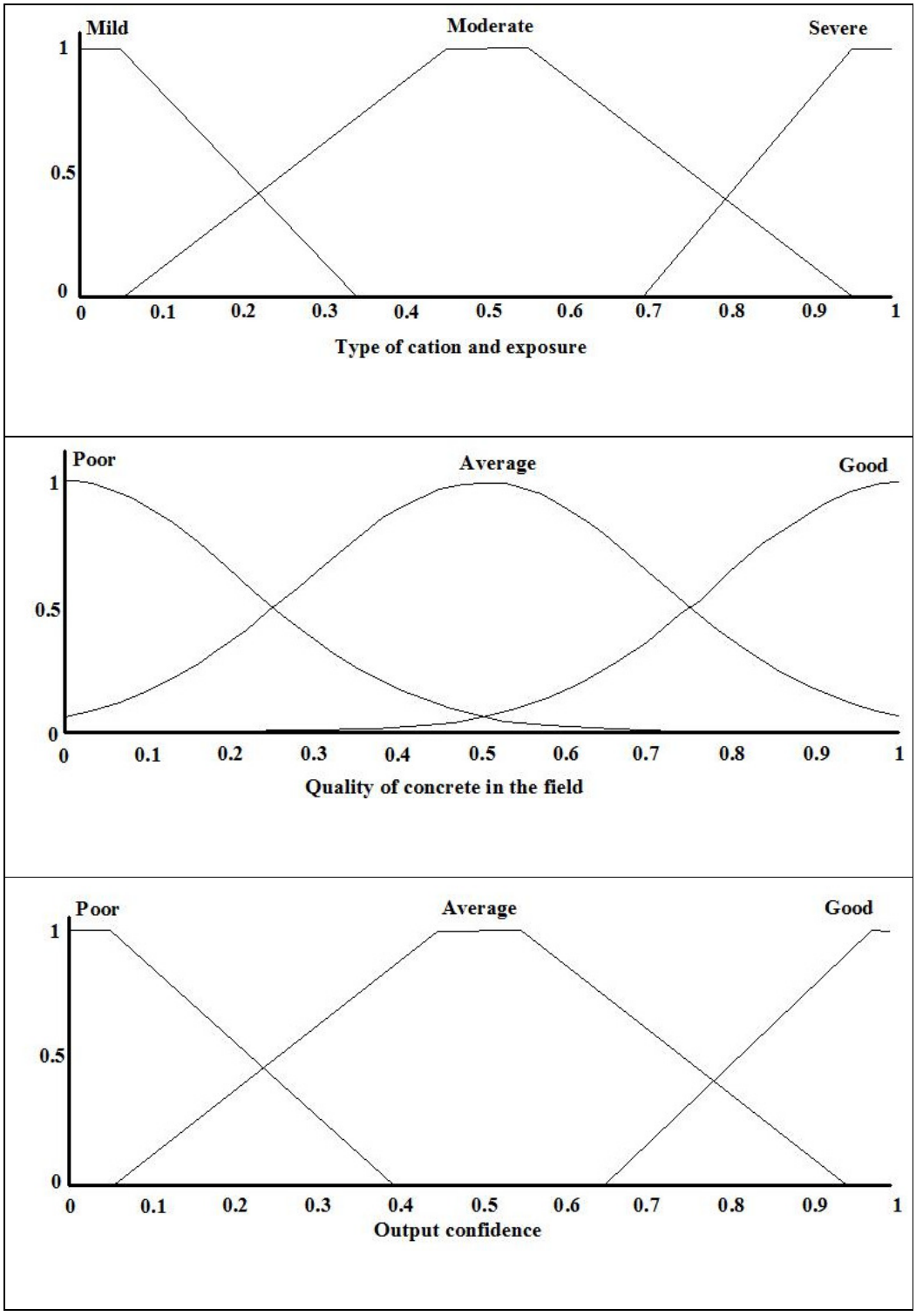



Figure 2. Membership Functions Showing Different Stages of the Input and Output Parameters

Considering the eighteen rules which were created for determining the output variable, different extent of uncertainty namely, 0, 25%, 50%, 75% and 100%, were considered for each input parameter in combination with other values of other parameters. A particular result, obtained from ruleview of Matlab, is shown in Figure 3. Confidence values for inputs and output confidence value are noted at the top of this figure.

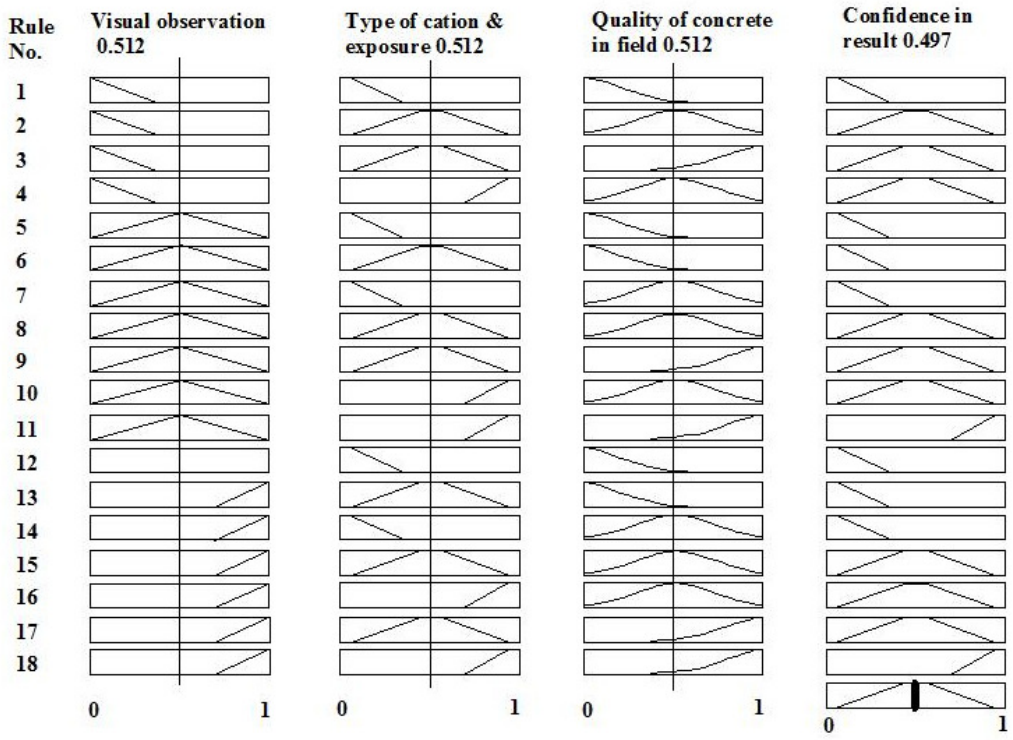



Figure 3. A Particular Result for Confidence in Result

A surface view diagram of relationship of confidence with the input parameters is shown in Figure 4. In such a diagram, only two input parameters may be considered with the output at a time.

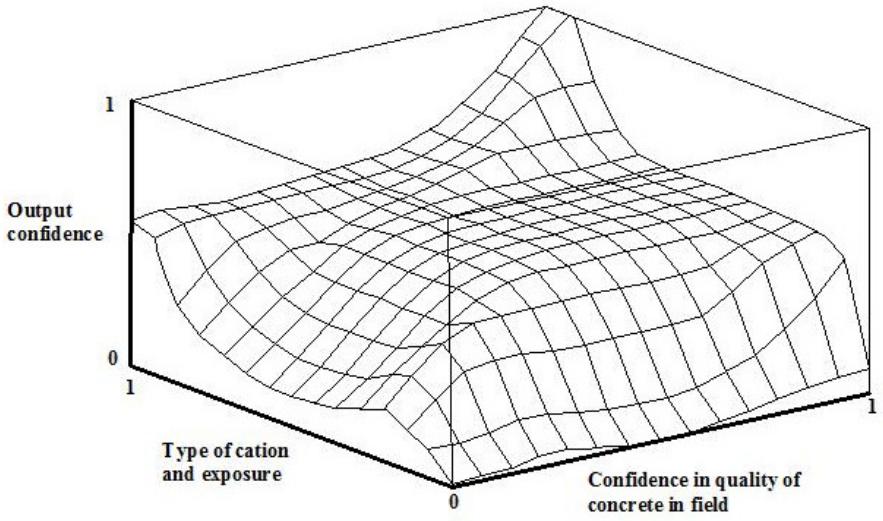



Figure 4. Relationship of Parameters

In the above relationship shown, more than one input value had a variation. Still, it can be readily seen from the surface view diagram in Figure 4 that all the input factors should have high confidence values for getting a good output confidence value. To see the effect of individual variations in input values on the output value, each input value was varied, one by one, to 75%, 50%, 25% and 0% from its maximum value of 1. Here, exact values were not taken and intentionally values close to particular percentage values were taken to emphasise the fuzziness of input parameters. Table 2 shows the input values taken. Varied values of inputs have been shown bold and underlined. Output values were determined using the fuzzy tool box of Matlab 5.3.

| Value of Input 1 | Value of Input 2 | Value of Input 3 | Value of output |
|---|---|---|---|
| **0.958** | 0.982 | 0.959 | 0.761 |
| **0.741** | 0.982 | 0.959 | 0.728 |
| **0.500** | 0.982 | 0.959 | 0.766 |
| **0.253** | 0.982 | 0.959 | 0.718 |
| **0.030** | 0.982 | 0.959 | 0.521 |
| 0.958 | **0.982** | 0.959 | 0.761 |
| 0.958 | **0.747** | 0.959 | 0.525 |
| 0.958 | **0.506** | 0.959 | 0.500 |
| 0.958 | **0.259** | 0.959 | 0.491 |
| 0.958 | **0.026** | 0.959 | 0.187 (Minimum) |
| 0.958 | 0.982 | **0.959** | 0.761 |
| 0.958 | 0.982 | **0.747** | 0.572 |
| 0.958 | 0.982 | **0.518** | 0.505 |
| 0.958 | 0.982 | **0.276** | 0.500 |
| 0.958 | 0.982 | **0.0235** | 0.500 |

Table 2. Input and Output Values

Variations in output value are shown in Figure 5. It shows the effect of type of variation chosen for an input parameter. The case of variation in sulphate exposure, which is considered in terms of a trapezoidal shape, gives the minimum values out of these variations. This happens to be the parameter which may not be 'visible' unless some information pertaining to this aspect is readily available or some tests are carried out in the determination of exposure. At the same time, as is clear from Table 8, lesser confidence about this input parameter would be more detrimental to the output confidence. This aspect of determination of relative importance of various factors, in durability studies concerning cement based materials with the help of fuzzy logic seems very useful in the field of maintenance of constructed facilities.

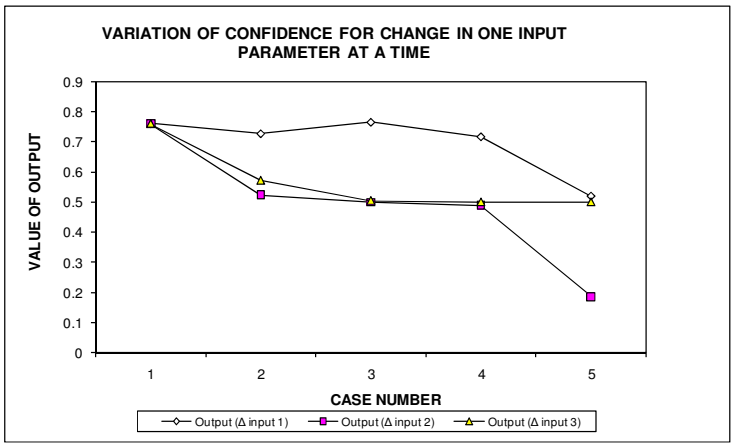



Figure 5. Variation of confidence for Change in One Input at a Time

In addition to the above, a definite variation in the accuracy of input parameters (ranging from 0 to 1, with a step size of 0.1, applied to all the three input parameters one by one) was taken and variation in the output confidence was calculated with the help of Matlab. A relationship of variations among the input and output parameters is given in Figure 6.

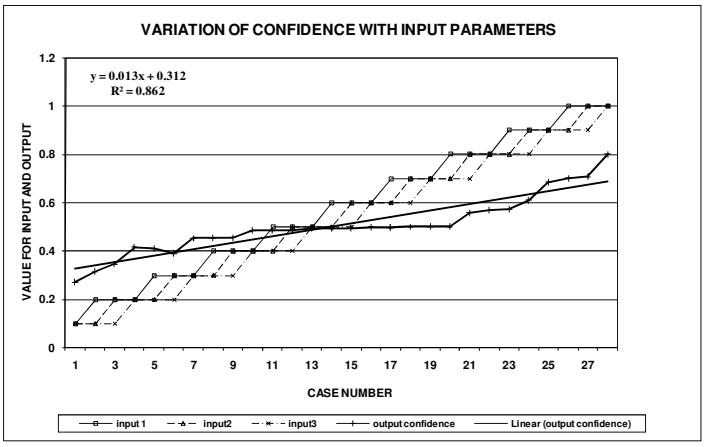



Figure 6. Relationship of Variations Among Input and Output Parameters

The value of output confidence increases with increases taking place in input values one by one. In the middle region of the output curve, there is a flat portion (from case number 10 to 18) where

the output value varied in a very narrow range, from 0.4856 to 0.501. This flat portion applied for input values set [0.4, 0.4, 0.4] to [0.7, 0.7, 0.6]. This indicated that for getting a high value of output confidence it is important to have all input confidence values crossing a threshold value. The lower threshold of this range seems to be defined by a value of 0.4 for all input values while the upper threshold of this range is defined by a value of 0.7 happening atleast for two input values. After this range, the output confidence improves and achieves the highest value of 0.8008 when all the input parameters were assigned a value of 1. Consideration of the fact that a value of 1.0 for output confidence is not achieved even when this value of 1.0 is allotted to all the input values may indicate that there may be definitely a fuzzy nature in output confidence even when definite values are taken for the input, depending on the nature of membership functions and the fuzzy rules considered for interpreting various conditions.

# 4. CONCLUSION

The following conclusions may be drawn from this study involving the use of fuzzy logic considerations to visual observations in assessing the performance of cement based materials under sulphate attack.

1. To have a high confidence in output all the input factors should have high confidence values. It means that knowledge about input factors should be good to have a good result.
2. Relative importance of various input factors may be determined with the help of fuzzy logic considerations. For example, minimum values of output confidence were achieved in the case of variation of input value for sulphate exposure, which was considered in terms of a trapezoidal shape. Shape of membership function chosen for the input parameter plays an important role in such determination. This aspect of determination of relative importance of various factors, in durability studies concerning cement based materials with the help of fuzzy logic seems very useful in the field of maintenance of constructed facilities.
3. The study reveals that for getting a high value of output confidence it is important to have all input confidence values crossing a threshold value. It may be concluded that output confidence can not be improved if even a single input parameter happens to be showing a low value of confidence attached to it.
4. Consideration of the fact that a value of 1.0 for output confidence is not achieved even when this value of 1.0 is allotted to all the input values may indicate that there may be definitely a fuzzy nature in output confidence even when definite values are taken for the input, depending on the nature of membership functions and the fuzzy rules considered for interpreting various conditions.
5. Various ratings followed in laboratory studies have some differences based on chosen parameters. A broad rating system based on extensive data, acquired in the past experiments, should be created with clearly marked effects and necessary explanations. This may be followed by the research community to help a clear interpretation and comparison of results.

# REFERENCES

[1] Friederike Weritz, Alexander Taffe, Dieter Schaurich and Gerd Wilsch, 'Detailed depth profiles of sulfate ingress into concrete measured with laser induced breakdown spectroscopy', Construction and Building Materials, 23 (2009) 275–283

[2] Nader Ghafoori, Hamidou Diawara and Shane Beasley, 'Resistance to external sodium sulfate attack for early-opening-to-traffic Portland cement concrete', Cement & Concrete Composites 30 (2008) 444–454

[3] Meenashi D. Cohen and Bryant Mather, 'Sulfate attack on Concrete – Research Needs', ACI Materials Journal, (1991), V. 88, No. 1, 62-68
[4] R.D. Hooton, 'Bridging the gap between research and standards', Cement and Concrete Research, 38 (2008) 247–258
[5] J.A. Larbi, 'Microscopy applied to the diagnosis of the deterioration of brick masonry', Construction and Building Materials 18 (2004) 299–307
[6] Nabil M. Al-Akhras, 'Durability of metakaolin concrete to sulfate attack', Cement and Concrete Research 36 (2006) 1727–1734
[7] S.M. Torres, C.A. Kirk, C.J. Lynsdale, R.N. Swamy and J.H. Sharp, 'Thaumasite–ettringite solid solutions in degraded mortars', Cement and Concrete Research, 34 (2004) 1297–1305
[8] M.J. Shannag and Hussein A. Shaia, 'Sulfate resistance of high-performance concrete', Cement & Concrete Composites 25 (2003) 363–369
[9] Adam Neville, 'The confused world of sulfate attack on concrete', Cement and Concrete Research 34 (2004) 1275–1296
[10] Zhen-Tian Chang, Xiu-Jiang Song, Robert Munn and Marton Marosszeky, 'Using limestone aggregates and different cements for enhancing resistance of concrete to sulphuric acid attack', Cement and Concrete Research 35 (2005) 1486 – 1494
[11] E.F. Irassar, A. Di Maio and O.R. Batic, 'Sulfate Attack On Concrete With Mineral Admixtures', Cement and Concrete Research, Vol. 26, No. 1, pp. 113-123,1996
[12] E.F. Irassar, 'Sulfate attack on cementitious materials containing limestone filler — A review', Cement and Concrete Research, 39 (2009) 241–254
[13] Y.F. Fan, Z.Q. Hub, Y.Z. Zhang and J.L. Liu, 'Deterioration of compressive property of concrete under simulated acid rain environment', Construction and Building Materials, 24 (2010) 1975–1983
[14] M.J. Shannag and Hussein A. Shaia, 'Sulfate resistance of high-performance concrete', Cement & Concrete Composites, 25 (2003) 363–369
[15] Moises Frías, M. Isabel Sanchez de Rojas and Cristina Rodríguez, 'The influence of SiMn slag on chemical resistance of blended cement pastes', Construction and Building Materials, 23 (2009) 1472–1475
[16] D. Planel, J. Sercombeb,, P. Le Bescop, F. Adenot, J.-M. Torrenti, 'Long-term performance of cement paste during combined calcium leaching–sulfate attack: kinetics and size effect', Cement and Concrete Research 36 (2006) 137–143
[17] S.R. Stock, N.K. Naikb, A.P. Wilkinson, and K.E. Kurtis, 'X-ray microtomography (microCT) of the progression of sulfate attack of cement paste', Cement and Concrete Research 32 (2002) 1673–1675
[18] S.A. Hartshorn, J.H. Sharp and R.N. Swamy, 'Thaumasite formation in Portland-limestone cement pastes', Cement and Concrete Research 29 (1999) 1331–1340
[19] Q. Zhou, J. Hill, E.A. ByarsT, J.C. Cripps, C.J. Lynsdale and J.H. Sharp, 'The role of pH in thaumasite sulfate attack', Cement and Concrete Research 36 (2006) 160– 170
[20] S.U. Al-Dulaijan , M. Maslehuddin, M.M. Al-Zahrani, A.M. Sharif, M. Shameem and M. Ibrahim, 'Sulfate resistance of plain and blended cements exposed to varying concentrations of sodium sulfate', Cement & Concrete Composites 25 (2003) 429–437
[21] G. Collett, N.J. Crammond, R.N. Swamya and J.H. Sharp, 'The role of carbon dioxide in the formation of thaumasite', Cement and Concrete Research, 34 (2004) 1599–1612
[22] R.S. Gallop and H.F.W. Taylor, 'Microstructural And Microanalytical Studies Of Sulfate Attack. V. Comparison Of Different Slag Blends', Cement and Concrete Research, Vol. 26, No. 7, pp. 1029-1044, 1996
[23] Seung Tae Lee, Robert Doug. Hooton, Ho-Seop Jung, Du-Hee Park and Chang Sik Choi, 'Effect of limestone filler on the deterioration of mortars and pastes exposed to sulfate solutions at ambient temperature', Cement and Concrete Research 38 (2008) 68–76
[24] T. Bakharev, J.G. Sanjayan and Y.B. Cheng, 'Sulfate attack on alkali-activated slag concrete', Cement and Concrete Research 32 (2002) 211 –216
[25] J. Hill, E.A. Byars , J.H. Sharp, C.J. Lynsdale, J.C. Cripps and Q. Zhou, 'An experimental study of combined acid and sulfate attack of concrete', Cement & Concrete Composites 25 (2003) 997–1003
[26] P. Pipilikaki, M. Katsioti and J.L. Gallias, 'Performance of limestone cement mortars in a high sulfates environment', Construction and Building Materials, 23 (2009) 1042–1049
[27] M.T. Bassuoni and M.L. Nehdi, 'Durability of self-consolidating concrete to sulfate attack under combined cyclic environments and flexural loading', Cement and Concrete Research 39 (2009) 206–226

[28] F. Bellmann and J. Stark, 'Prevention of thaumasite formation in concrete exposed to sulphate attack', Cement and Concrete Research 37 (2007) 1215–1222
[29] D.W. Hobbs and M.G. Taylor, 'Nature of the thaumasite sulfate attack mechanism in field concrete', Cement and Concrete Research 30 (2000) 529-533
[30] M. Nehdi and M. Hayek, 'Behavior of blended cement mortars exposed to sulfate solutions cycling in relative humidity', Cement and Concrete Research 35 (2005) 731–742
[31] Paul Brown, R.D. Hooton and Boyd Clark, 'Microstructural changes in concretes with sulfate exposure', Cement & Concrete Composites 26 (2004) 993–999
[32] Gozde Inan Sezer, Kambiz Ramyar, Bekir Karasu, A. Burak Goktepe and Alper Sezer, 'Image analysis of sulfate attack on hardened cement paste', Materials and Design, 29 (2008) 224–231
[33] H.A.F. Dehwah, 'Effect of sulfate concentration and associated cation type on concrete deterioration and morphological changes in cement hydrates', Construction and Building Materials, 21 (2007) 29–39
[34] F. Bellmann and J. Stark, 'The role of calcium hydroxide in the formation of thaumasite', Cement and Concrete Research, 38 (2008) 1154–1161
[35] H. Saricimen, M. Shameem, M.S. Barry, M. Ibrahim and T.A. Abbasi, 'Durability of proprietary cementitious materials for use in wastewater transport systems', Cement & Concrete Composites 25 (2003) 421–427
[36] A. K. Tamimi and M. Sonebi, 'Assessment of Self-compacting concrete immersed in acidic solutions', Journal of Materials in Civil Engineering, ASCE, Vol. 15, No. 4, 2003, 354-357
[37] Omar S. Baghabra Al-Amoudi, 'Attack on plain and blended cements exposed to aggressive sulfate environments', Cement & Concrete Composites 24 (2002) 305–316
[38] E. Rozière, A. Loukili, R. El Hachem and F. Grondin, 'Durability of concrete exposed to leaching and external sulphate attacks', Cement and Concrete Research xxx (2009) xxx–xxx
[39] E.F. Irassar, A. Di Maio and O.R. Batic, 'Sulfate Attack On Concrete With Mineral Admixtures', Cement and Concrete Research, Vol. 26, No. 1, pp. 113-123,1996
[40] D.D. Higgins and N.J. Crammond, 'Resistance of concrete containing ggbs to the thaumasite form of sulfate attack', Cement & Concrete Composites 25 (2003) 921–929
[41] D.M. Mulenga, J. Stark and P. Nobst, 'Thaumasite formation in concrete and mortars containing fly ash', Cement & Concrete Composites, 25 (2003) 907–912
[42] Michael Thomas, Kevin Folliard, Thano Drimalas and Terry Ramlochan, 'Diagnosing delayed ettringite formation in concrete structures', Cement and Concrete Research, 38 (2008) 841–847
[43] V. Assaad Abdelmseeh, J. Jofriet and G. Hayward, 'Sulphate and sulphide corrosion in livestock buildings, Part I: Concrete deterioration', Biosystems Engineering, (2008 ) 372 – 381
[44] M.T. Bassuoni and M.L. Nehdi, 'Resistance of self-consolidating concrete to sulfuric acid attack with consecutive pH reduction', Cement and Concrete Research 37 (2007) 1070–1084
[45] M.L. Berndt, 'Properties of sustainable concrete containing fly ash, slag and recycled concrete aggregate', Construction and Building Materials, 23 (2009) 2606–2613
[46] Manu Santhanam, Menashi D. Cohen and Jan Olek, 'Mechanism of sulfate attack: A fresh look Part 1: Summary of experimental results', Cement and Concrete Research 32 (2002) 915–921
[47] Bentur A. and Cohen M.D., 'Effect of condensed silica fume on the microstructure of the interfacial zone in Portland cement mortars', Journal of the American Ceramic Society, (1987) V. 70, No. 10, 738-743
[48] Manu Santhanam, Menashi D. Cohen and Jan Olek, 'Mechanism of sulfate attack: a fresh look Part 2. Proposed mechanisms', Cement and Concrete Research 33 (2003) 341–346
[49] Cohen M.D., 'Theories of expansion in sulfoaluminate type expansive cements: schools of thought', Cement and Concrete Research, (1984) V. 13, No. 6, 809-818
[50] Gunnar M. Idorn, 'Innovation in concrete research—review and perspective', Cement and Concrete Research, 35 (2005) 3-10
[51] E.F. Irassar, V.L. Bonavetti, M. Gonza´lez, 'Microstructural study of sulfate attack on ordinary and limestone Portland cements at ambient temperature', Cement and Concrete Research 33 (2003) 31–41
[52] M. Floyd, 'A comparison of classification systems for aggressive ground with thaumasite sulfate attack measured at highway structures in Gloucestershire, UK', Cement & Concrete Composites 25 (2003) 1185–1193

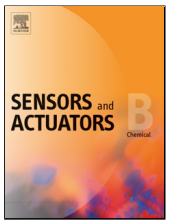
[53] T.I. Longworth, 'Contribution of construction activity to aggressive ground conditions causing the thaumasite form of sulfate attack to concrete in pyritic ground', Cement & Concrete Composites 25 (2003) 1005–1013
[54] S.T. Lee , H.Y. Moon and R.N. Swamy, 'Sulfate attack and role of silica fume in resisting strength loss', Cement & Concrete Composites 27 (2005) 65–76
[55] D.W. Hobbs, 'Thaumasite sulfate attack in field and laboratory concretes: implications for specifications', Cement & Concrete Composites 25 (2003) 1195–1202
[56] N.J. Crammond, 'The thaumasite form of sulfate attack in the UK', Cement & Concrete Composites 25 (2003) 809–818
[57] S.R. Stock, N.K. Naik, A.P. Wilkinson and K.E. Kurtis, 'X-ray microtomography (microCT) of the progression of sulfate attack of cement paste', Cement and Concrete Research 32 (2002) 1673–1675
[58] K. Torii, K. Taniguchi and M. Kawamura, 'Sulfate Resistance Of High Fly Ash Content Concrete', Cement and Concrete Research, (1995) Vol. 25. No. 4, pp. 759-768
[59] Kamile Tosun, Burak Felekoglu, Bülent Baradan, and Akın Altun, 'Effects of limestone replacement ratio on the sulfate resistance of Portland limestone cement mortars exposed to extraordinary high sulfate concentrations', Construction and Building Materials, 23 (2009) 2534–2544
[60] Omar Saeed Baghabra Al-Amoudi, 'Performance of 15 reinforced concrete mixtures in magnesium-sodium sulphate environments', Construction and Building Materials, Vol. 9. No. 3, 149-158, 1995
[61] P. E. Grattan-Bellew, 'Microstructural investigation of deteriorated Portland cement concretes', Construction and Building Materials, Vol. 10, No. 1, pp. 3-16, 1996
[62] Thidar Aye, Chiaki T. Oguchi and Yasuhiko Takaya, 'Evaluation of sulfate resistance of Portland and high alumina cement mortars using hardness test', Construction and Building Materials, 24 (2010) 1020–1026
[63] T. Vuka, R. Gabrovsek and V. Kaucic, 'The influence of mineral admixtures on sulfate resistance of limestone cement pastes aged in cold MgSO4 solution', Cement and Concrete Research, 32 (2002) 943–948
[64] Cohen M.D. and Bentur A., 'Durability of Portland cement silica fume pastes in magnesium sulfate and sodium sulfate solutions', ACI Materials Journal, (1988) V. 85, No. 3, 148-157*63
[65] George J. Klir and Bo Yuan, 'Fuzzy sets and fuzzy logic', Prentice Hall of India Private Limited, New Delhi, Fifth reprint, 2001

**AUTHORS**

Alok Verma is an Associate Professor in Civil Engineering at Delhi Technological University, Delhi (India). His research interests include durability of cement based materials, mathematical analysis of random phenomena and effect of earthquakes on structures.

Mukesh Shukla, a Ph.D in Structural Engineering, is Director of a prestigious engineering institute at Ujjain, in the state of Madhya Pradesh in India. His research interests include durability of concrete and ferrocement structures.

A. K. Sahu is a Professor in Civil Engineering at Delhi Technological University, Delhi (India). His research interests include durability of concrete and masonry structure




Title: Carboxylated multiwalled carbon nanotubes based biosensor for aflatoxin detection

Author: <ce:author id="aut0005" biographyid="vt0005"> Chandan Singh<ce:author id="aut0010" biographyid="vt0010"> Saurabh Srivastava<ce:author id="aut0015" biographyid="vt0015"> Md Azahar Ali<ce:author id="aut0020" biographyid="vt0020"> Tejendra K. Gupta<ce:author id="aut0025" biographyid="vt0025"> Gajjala Sumana<ce:author id="aut0030" biographyid="vt0030"> Anchal Srivastava<ce:author id="aut0035" biographyid="vt0035"> R.B. Mathur<ce:author id="aut0040" biographyid="vt0040"> Bansi D. Malhotra<ce:footnote id="fn1"><ce:note-para id="npar0005">These authors have contributed equally.</ce:note-para></ce:footnote>

PII: S0925-4005(13)00472-3
DOI: http://dx.doi.org/doi:10.1016/j.snb.2013.04.040
Reference: SNB 15405

To appear in: *Sensors and Actuators B*

Received date: 27-12-2012
Revised date: 7-4-2013
Accepted date: 10-4-2013

Please cite this article as: C. Singh, S. Srivastava, M.A. Ali, T.K. Gupta, G. Sumana, A. Srivastava, R.B. Mathur, B.D. Malhotra, Carboxylated multiwalled carbon nanotubes based biosensor for aflatoxin detection, *Sensors and Actuators B: Chemical* (2013), http://dx.doi.org/10.1016/j.snb.2013.04.040

# Carboxylated multiwalled carbon nanotubes based biosensor for aflatoxin detection

Chandan Singh[a £], Saurabh Srivastava[a,b £], Md Azahar Ali[a], Tejendra K. Gupta[c], Gajjala Sumana[a], Anchal Srivastava[b], R. B. Mathur[c], Bansi D. Malhotra[a,d*]

[a]Department of Science & Technology, Centre on Biomolecular Electronics, Biomedical Instrumentation Section, CSIR-National Physical Laboratory, New Delhi-110012, India.
[b]Department of Physics, Banaras Hindu University, Varanasi, U.P.-221005, India.
[c]Physics and Engineering of Carbon, Material Physics and Engineering Division, CSIR-National Physical Laboratory, New Delhi-110012, India-110012.
[d]Department of Biotechnology, Delhi Technological University, Delhi-110042, India.



* Author for correspondence e-mail.bansi.malhotra@dce.ac.in or bansi.malhotra@gmail.com;Tel.:+91117871043/1022 ,Extension 1609; Fax: +91112781023
£ These authors have contributed equally.

**Abstract**

We report results of studies relating to the development of an electrochemical immunosensor based on carboxylated multiwalled carbon nanotubes (c-MWCNTs) electrophoretically deposited onto indium tin oxide (ITO) glass. This c-MWCNTs/ITO electrode surface has been functionalized with monoclonal aflatoxin $B_1$ antibodies (anti-$AFB_1$) for the detection of aflatoxin-$B_1$ using electrochemical technique. Electron microscopy, X-ray diffraction and Raman studies suggest the successful synthesis of c-MWCNTs and the Fourier transform infra-red spectroscopic (FT-IR) studies reveal its carboxylic functionalized nature. The proposed immunosensor shows high sensitivity (95.2μA $ng^{-1}mL\ cm^{-2}$), improved detection limit (0.08 ng $mL^{-1}$) in the linear detection range of 0.25-1.375 ng $mL^{-1}$. The low value of association constant (0.0915 ng $mL^{-1}$) indicates high affinity of immunoelectrode towards aflatoxin ($AFB_1$).

**Keywords**: Aflatoxin $B_1$, Multiwalled carbon nanotubes, Electrochemical detection, Electrophoretic deposition

## 1. Introduction

Aflatoxins are a group of secondary fungal metabolites that are produced by *Aspergillus flavus* and *Aspergillus parasiticus* under certain conditions [1]. The four most important aflatoxins have been designated as $B_1$, $B_2$, $G_1$ and $G_2$. The International Agency for Research on Cancer (IARC) has categorized aflatoxin $B_1$ as a group 1 human carcinogen and aflatoxins $G_1$, $G_2$, and $B_2$ as group 2, as possible human carcinogens [2-3]. These toxins exhibit carcinogenic, mutagenic and teratogenic properties and are isolated from various agricultural products [4]. Aflatoxin $B_1$ ($AFB_1$) can enter the food chain mainly by ingestion via dietary route in humans and animals. The intake of $AFB_1$ over a long period of time, even at very low concentration, may be highly dangerous [5]. The maximum permitted (per kg) amount of $AFB_1$ by European community legislation is 2µg above which it may lead to liver cancer that is known to be the fifth most commonly occurring cancer in the world [6]. Due to the widespread prevalence of aflatoxins, efforts are being made to develop rapid and sensitive methods for their detection. The conventional techniques such as thin layer chromatography (TLC), high performance liquid chromatography (HPLC), gas liquid chromatography (GLC) and mass spectrometry have been used for aflatoxin detection [7-10]. However, these techniques are known to be expensive, require increased amount of samples and are time consuming. In this context, the development of an electrochemical biosensor for food toxin (aflatoxin $B_1$) detection has recently aroused much interest due to its high sensitivity, fast detection, high signal-to-noise ratio and simplicity [11, 12]. Electrochemical immunosensor based on screen-printed electrode has been proposed for the detection of aflatoxins in barley and milk [13, 14]. Many nanostructured materials including platinum, gold and nickel have recently also been proposed for $AFB_1$ detection [15-17].

Multiwalled carbon nanotubes (MWCNTs) based electrochemical biosensor may perhaps be advantageous due to their several captivating properties such as high conductivity, large surface area, fast electron transfer properties, biocompatibility and high mechanical strength [18, 19]. Recent studies have shown that use of MWCNTs may result in increased electrochemical activity of different biomolecules and hence these may be used to obtain enhanced electron transfer of desired proteins [20, 21]. In addition, ease of surface functionalization, ballistic electron transport and the loading of greater amount of biomolecules both on outer and inner surface may result in increased redox current many folds as compared to that of the glassy carbon electrode [22, 23]. MWCNTs are known to be new kinds of quantum wires, whose electronic states lie in the vicinity of the Fermi level and are quite different from those of free electrons and hence exhibit large current-carrying capacity. And the interesting electrochemical properties such as increased voltammetric current, increased heterogeneous electron-transfer rate, insignificant surface fouling property and excellent "electrocatalytic" effect towards the oxidation/reduction of various proteins make MWCNTs an ideal candidate for application to biosensing [24].

The desired biomolecules can be both covalently or non-covalently linked with MWCNTs surface. The non-covalent functionalization of MWCNTs may result in decreased stability and durability of the sensor due to prevailing electrostatic interactions [25, 26]. In this context, the utilization of carboxylated MWCNTs (c-MWCNTs) for covalent attachment of the aminated antibodies via strong amide bond formation may perhaps lead to increased stability of the biosensor. Keeping this in mind, we demonstrate the development of a novel electrophoretically deposited c-MWCNTs platform for the covalent attachment of anti-$AFB_1$ for sensitive detection of aflatoxin $B_1$ using electrochemical technique. Among the various techniques for fabrication of thin film of carbon nanotubes, the electrophoretic deposition

(EPD) of CNTs has been found to offer several advantages such as uniformity of deposition, control of thickness, micro structural homogeneity and simplicity [23, 27].

## 2 Experimental

### *2.1 Materials and methods*

Aflatoxin $B_1$ ($AFB_1$), anti-aflatoxin $B_1$ mouse monoclonal antibodies (anti-$AFB_1$), bovine serum albumin (BSA), N-hydroxysuccinimide (NHS) and N-ethyl-N'-(3-dimethylaminopropyl carbodiimide) (EDC) have been purchased from Sigma-Aldrich USA. All other chemicals are of analytical grade and have been used without further purification. Deionized water has been used in all the buffer and solution preparations.

### *2.2 Synthesis and functionalization of MWCNTs*

Chemical vapour deposition method has been utilized for the preparation of multiwalled carbon nanotubes (MWCNTs) using a mixture of ferrocene and toluene that act as a catalyst and hydrocarbon source, respectively [28]. These nanotubes are purified and functionalized through refluxing in concentrated nitric acid/sulphuric acid solution, generating a large number of COOH groups on the MWCNTs surface [29].

### *2.3 Fabrication of c-MWCNTs/ITO electrodes*

Carboxyl functionalized multiwalled carbon nanotubes (c-MWCNTs) have been deposited onto ITO substrate using electrophoretic deposition (EPD) technique [30]. The stock solution of c-MWCNTs (50 $mgdL^{-1}$) has been prepared in acetonitrile using ultrasonication (50 W, 0.25 A) for 3-4 h. Further, 100 µl of this solution is dispersed in 10 mL of acetonitrile to prepare c-MWCNTs colloidal suspension suitable for electrophoretic deposition. A constant DC voltage source having two electrode systems is used for

electrophoretic deposition. Surface charge on c-MWCNTs is generated by mixing $10^{-5}$ to $10^{-4}$ mol of magnesium nitrate [Mg $(NO_3)_2.6H_2O$] in the colloidal suspension that act as an electrolyte [30]. Pre-cleaned ITO glass (sheet resistance $30\Omega cm^{-1}$) and platinum foil are used as anode and cathode, respectively [Fig.1]. The electrodes having separation of 1 cm are immersed into the colloidal suspension containing c-MWCNTs and a constant electric field having intensity as 110 V/cm is applied for about 2 min. The c-MWCNTs/ITO electrodes ($0.65cm^2$), are removed from the suspension and are washed with deionized water followed by drying at room temperature (298K).

*2.4 Immobilization of monoclonal antibodies of aflatoxin (anti-$AFB_1$) onto c-MWCNTs/ITO electrodes*

Anti-$AFB_1$ solution having concentration of 10µg/mL has been prepared in phosphate buffer (PB, pH 7.4). 10 µL of this solution is uniformly spread on c-MWCNTs/ITO electrode surface and is incubated for 6 h under humid conditions at room temperature ($25^0$C). Prior to immobilization, the c-MWCNTs/ITO electrode is activated using EDC (0.4M) as the coupling agent and NHS (0.1M) as activator (Fig. 1) [31]. The $NH_2$ group of anti-$AFB_1$ is covalently bound with COOH terminal of MWCNTs via strong amide bond (CO-NH) formation. Bovine serum albumin (BSA) solution (1mg/mL) has been used to block non-specific active sites of the electrode. The BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode is then washed with PBS and stored at $4^0$C, when not in use.

*2.5 Characterization techniques*

The fabricated electrodes such as c-MWCNTs/ITO and anti-$AFB_1$/MWCNTs/ITO are characterized using Fourier transformed infra-red Spectroscopy (FT-IR, Perkin-Elmer, model spectrum BX using ATR accessory) X-ray Diffractrometer (Bruker), Raman Spectroscopy

(HR800 LabRam, Horiba/Jobin-Yvon), Scanning electron microscopy (SEM LEO 440), Transmission electron microscopy (HR-TEM, Tecnaii-G2F30 STWIN), UV-Visible spectrophotometer (Perkin-Elmer). Electrochemical characterization has been conducted using Autolab Potentiostat/Galvanostat (Eco Chemie, Netherlands) in presence of phosphate buffer saline (PBS; 50 mM, 0.9% NaCl, pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$ (redox species) using a three-electrode cell with Ag/AgCl as a reference electrode and Pt foil as the counter electrode.

## 3. Results and discussions

### *3.1 Structural studies*

Figure 2A demonstrates results of the Raman spectroscopic studies of MWCNTs and carboxylated MWCNTs. The peak seen at 1345 $cm^{-1}$ is attributed to D-band, arising due to the disordered or $sp^3$ hybridized carbons in the nanotubes walls. The other characteristic peak seen at 1580 $cm^{-1}$ is assigned to the G-band that is related to tangential stretching of the $sp^2$-bonded carbon atoms in graphene-like structure. The ratio between D-band and G-band intensities ($I_D/I_G$) can be used to obtain information on nature of defects in MWCNTs. Raman spectra of c-MWCNTs shows that the intensity ratio increases as compared to that of MWCNTs. The ratio between D and G-band intensities ($I_D/I_G$) of MWCNTs and carboxylated MWCNTs is estimated to be as 0.7186 and 0.8038, respectively. This significant increase (12%) in the $I_D/I_G$ ratio of c-MWCNTs is perhaps due to the generation of carboxyl sites in the treated carbon nanotubes [32, 33].

Figure 2B shows the X-ray diffraction pattern obtained for the c-MWCNTs. The intense characteristic peak seen at 25.9$^o$ corresponds to the (002) reflection plane of c-MWCNTs. The diffraction peaks found at 42.7$^o$, 53.5$^o$ and 77.5$^o$ are indexed to the (100), (004) and

(110) reflections. The full width half maximum (FWHM) corresponding to (002) plane has been estimated as 1.005° with interplanar spacing of 3.4 Å.

Figure 3A shows results of UV-visible studies of dispersed c-MWCNTs and anti-$AFB_1$-MWCNTs in water . The absorption peak seen at 260 nm arises due to electronic $\pi$-$\pi$* transition of aromatic C-C bonds of MWCNTs (Fig. 3A, curve a). The UV-Visible studies obtained for anti-$AFB_1$ functionalized c-MWCNTs lead to change in intensity indicating biofunctionalization of carbon nanotubes (Fig.3A, curve b).

FT-IR spectra of c-MWCNTs/ITO and BSA/Anti-$AFB_1$/MWCNTs/ITO films are shown in Fig.3B. The spectrum obtained for c-MWCNTs/ITO film shows a strong peak at 1243 $cm^{-1}$ that is assigned to the stretching vibrations of C-O, the peak seen at 1492$cm^{-1}$ is due to C=C stretching and the 800$cm^{-1}$ peak is assigned to C-H bending vibrations. The stretching vibration seen at 1730 $cm^{-1}$ is attributed to the carboxyl group, indicating COOH functionalization of MWCNTs [curve, a]. After anti-$AFB_1$ immobilization onto c-MWCNTs/ITO surface, the characteristic peak found at 1559 $cm^{-1}$ is attributed to amide I. And the intense and broad peaks seen in the region 3100 $cm^{-1}$ may arises due to amide B indicating immobilization of antibodies (anti-$AFB_1$) [curve, b].

*3.2 Microscopic studies*

Fig. 4 shows scanning electron micrograph of c-MWCNTs/ITO and anti-$AFB_1$/MWCNTs/ITO electrode surfaces. The majority of c-MWCNTs appear to be entangled with tubular structure onto ITO surface and forms network like structure (image a). The diameter of c-MWCNTs varies from 40 nm to 100 nm with length over a few microns. After immobilization of anti-$AFB_1$ onto c-MWCNTs surface, the morphology changes to nanoporous structure and the globular structure on the surface indicates immobilization of anti-$AFB_1$ onto the MWCNTs/ITO surface (image b). The uniform and highly distributed

globular structure reveals relatively high loading of antibodies onto the surface and is attributed to the covalent binding between COOH group of nanotubes and $NH_2$ groups in the antibodies. The observed morphological changes of c-MWCNTs are due to antibodies funtionalization present on the c-MWCNTs/ITO electrode surface. The high resolution-transmission electron microscopy image [Inset 4a] shows the lattice fringes of c-MWCNTs with an interplanar spacing as 3.4 Å for (002) graphitic planes indicating high crystallinity.

*3.3 Electrochemical studies*

The cyclic voltammetry (CV) studies of c-MWCNTs/ITO electrode and anti-$AFB_1$/MWCNTs/ITO bioelectrode have been investigated in PBS (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$ in the range of -0.5 V to +0.7 V at the scan rate of 50 $mVs^{-1}$. The curve (a) [Fig.5A] shows well-defined redox peaks arising due to oxidation and reduction of $[Fe(CN)_6]^{3-/4-}$ in presence of c-MWCNTs/ITO electrode. The anodic and cathodic current peaks are found at $4.68\times10^{-4}$A and $4.11\times10^{-4}$A with peak- to- peak separation of 0.211V. It has been observed that the peak current in case of c-MWCNTs/ITO electrode is higher as compared to that of other electrode. This is perhaps due to presence of defect sites in c-MWCNTs that allow straight electron pathway for mass transportation towards electrode in the bulk solution. The magnitude of current in the case of anti-$AFB_1$/MWCNTs/ITO (curve, b) [Fig.5A] decreases to $3.07\times10^{-4}$A with peak-to-peak separation of 0.362 indicating hindered charge transfer due to insulating nature of antibodies. The response current further decreases slightly after BSA conjugation since BSA blocks most of the non-specific active sites. The oxidation peak potential of MWCNTs/ITO has been found at 0.26 V that is shifted to a higher value (0.31 V) for anti-AFB1/ MWCNTs/ITO electrode. This may be attributed to the macromolecular structure of the antibodies that perhaps obstructs the electron transfer owing to their insulating nature. These results confirm the functionalization of

MWCNTs/ITO electrode with the antibodies. This shift in the oxidation peak potential is not prominent after BSA conjugation (0.31 V).

The cyclic voltammetric studies have been performed on the BSA/Anti-$AFB_1$/MWCNTs bioelectrode as a function of scan rate in the range of 20-100mV/s [Fig. S1, see supplementary information sheet]. The magnitudes of anodic and cathodic peak currents have been found to vary proportionally as a function of square root of scan rate [Eq. (1-2)] as shown in [Fig. S1, inset (i)]. The redox peak potential is found to be shifted (anodic peak potential towards positive potential side and cathodic peak potential towards negative potential side) as the scan rate increases from 20 to 100 mV/s [Fig. S1, inset (ii)]. The proportional increase in the anodic and cathodic peak potential as a function of square root of scan rate [Eq. (3-4)] indicates a diffusion controlled process.

$$I_a\,[A] = 4.32\times10^{-5}\,[A] + 4.19\times10^{-5}\,[A^2s/mV]^{1/2} \times \{\text{scan rate (mV/s)}\}^{1/2};\ R^2 = 0.998 \quad (1)$$

$$I_c\,[A] = -9.15\times10^{-5}\,[A] - 3.21\times10^{-5}\,[A^2s/mV]^{1/2} \times \{\text{scan rate (mV/s)}\}^{1/2};\ R^2 = 0.992 \quad (2)$$

$$V_{ap} = 1.7\times10^{-1}(V) + 2.2\times10^{-2}(V^{1/2}s^{-1/2})*[\text{scan rate (mV/s)}]^{1/2};\ R^2 = 0.998 \quad (3)$$

$$V_{cp} = 1.9\times10^{-2}(V) - 2.35\times10^{-2}\,(V^{1/2}s^{-1/2})*[\text{scan rate (mV/s)}]^{1/2};\ R^2 = 0.995 \quad (4)$$

Further, the surface concentration of BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode has been calculated using Brown-Anston Model [Eq. (5)].

$$I_p = \frac{n^2F^2I^*AV}{4RT} \quad (5)$$

where, n is number of electrons transferred (1), F is Faraday constant (96485.5 $Cmol^{-1}$), I* is surface concentration ($molcm^{-2}$), A is surface area of the electrode (0.65$cm^2$), V is scan rate (20-100 mV/s), R is gas constant (8.314 $mol^{-1}K^{-1}$), and T is room temperature (298K). The surface concentration of BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode is estimated to be as $3.92\times10^{-8}$ $mol/cm^2$. Diffusion coefficient of redox species for BSA/Anti-

$AFB_1$/MWCNTs/ITO bioelectrode has been calculated to be as $3.79\times10^{-7}$ $cm^2/s$ using Randles-Sevcik equation [Eq. 6]

$$I_p = (2.69\times10^5)n^{3/2}AD^{1/2}\,v^{1/2}c \qquad (6)$$

where $I_p$ is peak current (A), n is electron stoichiometry, D is diffusion coefficient ($cm^2/s$), c is concentration ($mol/cm^2$) and *v* is scan rate (V/s).

*3.4 Electrochemical response studies*

The response of BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode has been investigated as a function of $AFB_1$ concentration using CV as shown in Fig. [5B]. It appears that the defects and oxygenated groups present on the c-MWCNTs surface are known to be responsible for the enhanced electron transfer rate and perhaps provide electrocatalytic effect towards the oxidation/reduction of a wide variety of compound including proteins as found in literature [20, 21]. It has been found that the response current for BSA/Anti-$AFB_1$/MWCNTs/ITO electrode increases as the concentration of the antigen ($AFB_1$) increases. It may perhaps be attributed to strong affinity of the antigens towards the antibodies may promote spatial orientation which may provide an easy conducting path for electron transfer to the electrode surface as reported in literature [34-36]. Alternately, it may perhaps be related to the formation of antigen-antibody complex between $AFB_1$ and anti-$AFB_1$ onto MWCNTs/ITO surface that acts as the electron transfer accelerating layer enabling facile charge transfer to the electrode surface [35]. The calibration plot obtained for BSA/Anti-$AFB_1$/MWCNTs/ITO electrode as a function of $AFB_1$ concentration follows Eq. (7) [Fig.5B, inset].

$$I(A) = 2.34\times10^{-4}A + 2.38\times10^{-5}A\ mL/ng \times AFB1\ concentration\ (ng/mL);\ R^2=0.996 \qquad (7)$$

Further, we have performed the control experiment using the c-MWCNTs/ITO electrode as a function of $AFB_1$ concentration conducted in PBS containing 5 mM $[Fe(CN)_6]^{3-/4-}$ (Fig. 5C). It has been observed that the response current obtained for c-MWCNTs/ITO does not significantly change as a function of $AFB_1$ concentration (Inset, Fig. 5 C).

This BSA/Anti-$AFB_1$/MWCNTs/ITO biosensor exhibits very high sensitivity as 95.2 $\mu A\ ng^{-1} mL\ cm^{-2}$ in the linear detection range of 0.25-1.375 ng/mL that may be attributed to excellent electrochemical properties and ballistic electron transport in MWCNTs [37]. The lower detection limit (0.08 ng/mL) has been estimated using $3\sigma/m$ criteria; where $\sigma$ is the standard deviation of the calibration plot whereas m is the obtained sensitivity. The low value of the association constant for BSA/Anti-$AFB_1$/MWCNTs/ITO immunoelectrode has been found as 0.0915 ng/mL that indicates higher affinity of anti-$AFB_1$ with aflatoxin ($AFB_1$). The response time has been estimated to be as 15 s [data not shown]. The sensing characteristic of proposed BSA/Anti-$AFB_1$/MWCNTs/ITO immunosensor has been summarized in Table 1 along with those reported in literature. It may be noted that the obtained detection limit for this immunosensor is better than those reported in literature [Table 1]. It may perhaps be attributed to very large surface area of the MWCNTs and the increased functional groups resulting in high loading of antibodies. Besides this, fabricated immunosensor exhibits the higher sensitivity (Table 1).

*3.5 Stability, Selectivity, and reproducibility studies*

The storage stability of this biosensor has been determined at a regular interval of one week and it has been found that it retains 92% current response within 45 days after which the current slightly decreases to 70 % [Fig. S2, see supplementary information sheet]. The selectivity studies of this immunosensor has been investigated in the presence of another food

toxin namely ochratoxin-A (1.0 ng/dL) using CV [Fig. 6 (A)]. However, no significant change in the current response as evident by very low relative standard deviation (0.54 %) indicating that this immunosensor is highly selective. The reproducibility of the BSA/Anti-$AFB_1$/MWCNTs/ITO biosensor using different working electrodes that fabricated via the same set of procedure with concentration of $AFB_1$ (0.5 ng/mL) has been investigated [Fig. 6 (B)]. It has been observed that this biosensor shows good reproducibility for different electrodes (six) with constant sensor surface area as evidenced by the relative standard deviation (RSD) of 2.0% (mean value = 227.6 μA). The low RSD (2.0%) of this fabricated BSA/Anti-$AFB_1$/MWCNTs/ITO immunosensor indicates good precision. It is found that this immunoelectrode shows good accuracy as evident by low RSD value (0.16 %, n=8), under eight repeated measurements.

## 4. Conclusions

We have fabricated a BSA/Anti-$AFB_1$/MWCNTs/ITO immunosensing platform for aflatoxin detection using electrochemical cyclic voltammetry technique. The c-MWCNTs are deposited onto ITO substrate using electrophoretic deposition and used for monoclonal anti-$AFB_1$ immobilization. The results of electrochemical studies of the proposed immunosensor indicates high sensitivity (95.2μA $ng^{-1}mL\ cm^{-2}$) and improved detection limit (0.08 ng/mL). In addition, this biosensor shows good reproducibility and stability as 45 days. Efforts should be made to utilize these carboxylated multi-walled carbon nanotubes based electrode for detection of other food toxins including ochratoxins (A & B), vomitoxin and citrinin etc.

## Acknowledgements

We thank Director NPL, India for the facilities. C. Singh, S. Srivastava and Md. Azahar Ali are thankful to DST and CSIR, India for providing financial support. The financial

support received by Department of Science and Technology, India (Grant No. DST/TSG/ME/2008/18 and GAP-081132) and Indian Council of Medical Research, India (Grant No. ICMR/5/3/8/91/GM/2010-RHN) is gratefully acknowledged.

**Table 1** Comparison of the sensing characteristics of BSA/Anti-$AFB_1$/MWCNTs/ITO biosensor with some of those reported in literature.

| Electrode | LDL | Stability | Response time | Sensitivity | Ref |
|---|---|---|---|---|---|
| BSA/aAFB1-C-AuNP/MBA/Au | 0.179 $ngmL^{1}$ | ..... | 60 s | 45 µA ng $mL^{-1}$ | 16 |
| a-AFB1/DMSO/RnNi-film/ITO | 0.327 $ngmL^{-1}$ | 60 days. | 5 s | 59.l µA/ng $mL^{-1}$ | 17 |
| Pt/CNT | 0.50 $ngmL^{-1}$ | ..... | 44 s | 0.11 µA/$ngmL^{-1}$ | 15 |
| BSA/Anti-AFB1/MWCNTs/ITO | 0.08 $ngmL^{-1}$ | 45 days | 15 s | 95.2 µA/$ngmL^{-1}$ | Present work |

**References**

[1] S. J. Daly, G. J. Keating, P. P. Dillon, B. M. Manning, R.O'Kennedy, H.A. Lee, M.R. A. Morgan, Development of surface plasmon resonance-based immunoassay for Aflatoxin B1, J. Agric. Food Chem.48 (2000) 5000-5004.

[2] R. J. Cole, R. H. Cox, Handbook of toxic fungal metabolites. Academic Press, New York 1981.

[3] International Agency for Research on Cancer, 1993. IARC Monographs on the Evaluations of Carcinogenic Risks to Human, vol. 56. IARC, Lyon, pp 489–521.

[4] O.M.Moss, Risk assessment for aflatoxins in foodstuffs. Int. Biodeteior. Biodegrad. 50(2002) 137–142.

[5] M. Miraglia, C. Brera, M.Colatosti, Application of biomarkers to assessment of risk to human health from exposure to mycotoxins. J. Microchem. 54 (1996) 472–477.

[6] D. M. Parkin, F. Bray, J. Ferlay, P. Pisani, Estimates of the worldwide incidence of 25 major cancers in 1990, Int. J. Cancer. 80 (1999) 827-41.

[7] E. Chiavaro, C. Dall'Asta, G. Galaverna, A. Biancardi, E. Gambarelli, A. Dossena, R. Marchelli, New reversed-phase liquid chromatographic method to detect aflatoxins in food and feed with cyclodextrins as fluorescence enhancers added to the eluent. J Chromatogr A. 937 (2001) 31-40.

[8] S. Nawaz, R. D. Coker, S. J. Haswell, HPTLC-A valuable chromatographic tool for the analysis of aflatoxins, Planar Chromatogr. 8 (1995) 4-9.

[9] AOAC, International Performance Tested Methods, 2004. Toxin tests kits (www.aoac.org).

[10] W. Horwitz, Official Method of Analysis of AOAC International, 17$^{th}$ ed. AOAC International, Gaithershurg, MD 2000.

[11] S. Srivastava,V. Kumar, A.Ali, P.R. Solanki, A. Srivastava, G.Sumana, P.S. Saxena, A.G. Joshi, B.D. Malhotra. Electrophoretically deposited reduced graphene oxide platform for food toxin detection, Nanoscale (DOI: 10.1039/C3NR32242D).

[12] R. M. Pemberton, R. Pittson, N. Biddle, G. A. Drago, J. P. Hart, Studies towards the development of a screen-printed carbon electrochemical immunosensor array for mycotoxins: A sensor for aflatoxin $B_1$, Anal. Lett. 39(2006) 1573–1586.

[13] N. Paniel, A. Radoi, J.L. Marty, Development of an electrochemical biosensor for the detection of aflatoxin $M_1$ in milk, Sensors 10 (2010) 9439-9448.

[14] S. Piermarini , L. Micheli , N.H.S. Ammida, G. Palleschi, D. Moscone, Electrochemical immunosensor array using a 96-well screen-printed microplate for aflatoxin $B_1$ detection, Biosensors. Bioelectron. 22 (2007) 1434–1440

[15] S. C. Li, J. H. Chen, H. Cao, D. S. Yao, D. L. Liu, Amperometric biosensor for aflatoxin $B_1$ based on aflatoxin-oxidase immobilized on multiwalled carbon nanotubes, Food Control 22 (2011) 43-49.

[16] A. Sharma, Z. Matharu, G. Sumana, P. R. Solanki, C. G. Kim, B. D. Malhotra, Antibody immobilized cysteamine functionalized-gold nanoparticles for aflatoxin detection, Thin Solid Films 519 (2010) 1213-18.

[17] P. Kalita, J. Singh, M. K. Singh, P. R. Solanki, G. Sumana, B. D. Malhotra, Ring like self assembled Ni nanoparticles based biosensor for food toxin detection, Appl. Phys. Lett. 100 (2012) 093702-5.

[18] J. Wang, Y. Lin, Functionalized carbon nanotubes and nanofibers for biosensing applications, Trends Analyt Chem. 27 (2008) 619–626.

[19] S. Viswanathan, L. C. Wu, M. R. Huang, J. Ho, Electrochemical immunosensor for cholera toxin using liposomes and poly(3,4-ethylenedioxythiophene)-coated carbon nanotubes, Anal. Chem. 78 (2006) 1115-21.

[20] M. Musameh, J. Wang, A. Merkoci, Y. Lin, Low-potential stable NADH detection at carbon-nanotube-modified glassy carbon electrodes, Electochem. Commun. 4 (2002) 743- 46.

[21]J. J. Gooding, R. Wibowo, J. Q. Liu, W. Yang, D. Losic, S. Orbons, F. J. Mearns, J. G. Shapter , D. B. Hibbert, Protein electrochemistry using aligned carbon nanotube arrays, J. Am. Chem. Soc. 125 (2003) 9006-07.

[22] J. Wang, M. Musameh, Carbon nanotube/teflon composite electrochemical sensors and biosensors, Anal. Chem. 75 (2003) 2075-79.

[23] V. Semet, V. T. Binh, D. Guillot, K. B. K. Teo, M. Chhowalla, G. A. J. Amaratunga, W. I. Milne, P. Legagneux, D. Pribat, Reversible electromechanical characteristics of individual multiwall carbon nanotubes, Appl. Phys. Lett. 87 (2005) 223103-5.

[24] M. Pumera, The electrochemistry of carbon nanotubes: fundamentals and applications,Chem. Eur. J. 15 (2009)4970-78.

[25] J. X. Wang, M. X. Li, Z. J. Shi, N. Q. Li, Z. N. Gu, Direct electrochemistry of cytochrome c at a glassy carbon electrode modified with single-wall carbon nanotubes, Anal. Chem.74 (2002)1993-97.

[26] J. V. Veetil, K. M. Ye, Development of immunosensors using carbon nanotube Biotechnol. Progr 23 (2007) 517-31.

[27] B. Gao, G. Z. Yue, Q. Qiu, Y. Cheng, H. Shimoda, L. Fleming, O. Zhou, Fabrication an electron field emission properties of carbon nanotube films by electrophoretic deposition, Adv. Mat. 13 (2001) 1770-73.

[28] R. B. Mathur, S. Chaterjee, B. P. Singh, Growth of carbon nanotubes on carbon fibre substrates to produce hybrid/phenolic composites with improved mechanical properties, Compos. Sci. Technol. 68 (2008) 1608-15.

[29] V. Datsyuk, M. Kalyva, K. Papagelis, J. Parthenios, D. Tasis, A. Siokou, I. Kallitsis , C. Galiotis, Chemical oxidation of multiwalled carbon nanotubes, Carbon 46 (2008) 833-40.

[30] A. R. Boccaccini, J. Chao, R. A. Judith, J. C. Boris, E. Thomas, J. Minay, S. P. Milo, Electrophoretic deposition of carbon nanotubes, Carbon 44 (2006) 3149 -60.

[31] S. Srivastava, P. R. Solanki, A. Kaushik, Md. A. Ali, A. Srivastava, B. D. Malhotra, A self assembled monolayer based microfluidic sensor for urea detection, Nanoscale 3 (2011) 2971-7.

[32] E. R. Edwards, E. F. Antunes, E. C. Botelho, M. R. Baldan , E. J. Corat, Evaluation of residual iron in carbon nanotubes purified by acid treatments, App. Surf. Sci. 258 (2011) 641-48.

[33] H. B. Zhang, G. D. Lin, Z. H. Zhou, X. Dong ,T. Chen, Raman spectra of MWCNTs and MWCNT-based $H_2$-adsorbing system, Carbon 40 (2002) 2429 -36.

[34] J. Okuno, K. Maehashi, K. Kerman, Y. Takamura, K. Matsumoto, E. Tamiya, Label-free immunosensor for prostate-specific antigen based on single-walled carbon nanotube array-modified microelectrodes, Biosens. Bioelectron. 22 (2007) 2377–2381.

[35] A. Kaushik, P. R. Solanki, M .K. Pandey, S. Ahmed, B. D. Malhotra, Cerium oxide-chitosan based nanobiocomposite for food borne mycotoxin detection, Appl. Phys. Lett. 95(2009) 173703-05.

[36] S. Viswanathan, L.-C. Wu, M.-R. Huang, J. A. Ho, Electrochemical immunosensor for cholera toxin using liposomes and poly(3,4-ethylenedioxythiophene)-coated carbon nanotubes, Anal. Chem. 78(2006) 1115-1121.

[37] R. A. Gustavo, M. D. Rubianes, M. C. Rodriguez, N. F. Ferreyra, G. L. Luque, M. L. Pedano, S. A. Miscoria, C. Parrado, Carbon nanotubes for electrochemical biosensing, Talanta 74 (2007) 291-07.

**Figure captions**

Fig. 1 Schematic representation of c-MWCNTs based biosensor for aflatoxin $B_1$ detection

Fig. 2 (A) Raman spectra of pristine MWCNTs and COOH functionalized MWCNTs and (B) XRD spectrum of c-MWCNTs

Fig. 3(A) UV-Spectra of c-MWCNTs (a) and anti-$AFB_1$/MWCNTs (b)

Fig. 3(B) FTIR spectra of c-MWCNTs/ITO (a) and BSA/Anti-$AFB_1$/MWCNTs/ITO (b)

Fig. 4 (A) SEM images of c-MWCNTs/ITO film (inset: high resolution TEM image of c-MWCNTs) and (B) anti-AFB1/MWCNTs/ITO film.

Fig. 5(A) Cyclic voltammetry characterization of c-MWCNTs/ITO (curve a), Anti-$AFB_1$/MWCNT/ITO (curve b) and BSA/Anti-$AFB_1$/MWCNTs/ITO (curve c) conducted in PBS (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$, (B) CV response of BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode at different concentration of $AFB_1$, conducted in PBS (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$ ; Inset: calibration plot obtained between the oxidation peak current and $AFB_1$ concentration, (C) Control experiment of the c-MWCNTs/ITO electrode as a function of $AFB_1$ concentration; Inset: oxidation peak current *vs* $AFB_1$ concentration plot of the c-MWCNTs/ITO electrode.

Fig. 6(A) CV response of BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrode in the presence of ochratoxin-A conducted in PBS (pH 7.4) containing 5 mM $[Fe(CN)_6]^{3-/4-}$ and (B) CV response of different BSA/Anti-$AFB_1$/MWCNTs/ITO bioelectrodes fabricated via the same set of procedure with $AFB_1$ (0.5 ng/mL).

**Biographies**

**Chandan Singh** received his M.Tech degree in biotechnology from Jaypee University of Information Technology, Waknaghat, H.P in 2011. He is presently working as DST-SRF at

Department of Science (DST) and Technology Centre on Biomolecular Electronics, Biomedical Instrumentation Section at the National Physical Laboratory, New Delhi, India.

**Saurabh Srivastava** received his MSc degree in Physics from Banaras Hindu University (BHU), India in 2007. He is presently pursuing Ph.D. from Physics department at BHU and also working as a Senior Research Fellow (CSIR) in the DST Centre on Biomolecular Electronics, Biomedical Instrumentation Section, National Physical Laboratory, New Delhi, India. He is actively engaged in the area of development of graphene based immunosensor for toxins detection.

**Md Azahar Ali** received his MSc degree in Electronics from Gauhati University and his M.Tech degree in Bioelectronics from the department of Electronics Engineering, Tezpur University, Assam in 2009. He is presently pursuing Ph.D. from Biomedical Engineering department at IIT Hyderabad, India and also working as a Senior Research Fellow (CSIR) in the DST Centre on Biomolecular Electronics, Biomedical Instrumentation Section at National Physical Laboratory, New Delhi, India. He is actively engaged in the area of development of microfluidic based chip for biomolecules detection.

**Tejendra K. Gupta** received his M. Tech degree from Guru Jambheshwer University, Hisar, Haryana. Currently he is working as Senior Research Fellow (SRF-CSIR) at Physics and Engineering of Carbon, Material Physics and Engineering Division, National Physical Laboratory (CSIR), New Delhi. India.

**Dr. R. B. Mathur** received his PhD from university of Rajasthan in 1976. He is chief scientist at material science and engineering division, National Physical Laboratory, New Delhi, India. His current research activities includes synthesis of carbon nanotubes by Arc

discharge technique and CVD, development of carbon fibres/carbon nanotubes reinforced composites development of advanced carbon materials for energy applications.

**Dr. Gajjala Sumana** received her Ph.D. (1998) from Jiwaji University in chemistry, India. She is currently working as a scientist with the DST Centre on Bimolecular Electronics at the National Physical Laboratory, New Delhi, India. She has a research experience of 10 years in controlled drug delivery, liquid crystal polymers, polymer dispersed liquid crystals and biosensors. She has published more than 50 publications in the area of biomedical applications.

**Dr. Anchal Srivastava** received his Ph.D from Banaras Hindu University, India in 2002. He continued in the same department as a post doctoral fellow and later joined as a lecturer in 2004. He is currently serving as an assistant professor at Department of Physics, BHU. He specializes in the area of synthesis, characterizations and applications of carbon nanomaterials. He has been awarded the Max Planck India fellowship and carried out his research at Max Planck Institute of Solid State Physics, Stuttgart, Germany in Professor Klaus Von Klitzing's group. He has also been awarded the BOYSCAST fellowship from government of India in 2008–09 during which he carried out research on graphene and related materials at Professor P. M. Ajayan's group in Rice University. He has also been a visiting research scholar at Rensselaer Polytechnic Institute, USA in 2006.

**Prof. B. D. Malhotra** received his Ph.D degree in physics from University of Delhi, Delhi, India in 1980. He has published more than 230 papers, has filed 10 patents, has edited/co-edited books on biosensors and polymer electronics, and is currently the Professor at Delhi Technological University, Delhi, India. He has research experience of about 30 years in the field of biomolecular electronics and has supervised 21 Ph.D. students till date. He is a

Fellow of both the Indian National Science Academy and the National Academy of Sciences, India.His current activities including biosensors, conducting polymers, Langmuir-Blodgett films, self-assembled monolayers and nanomaterials etc.

## Figures

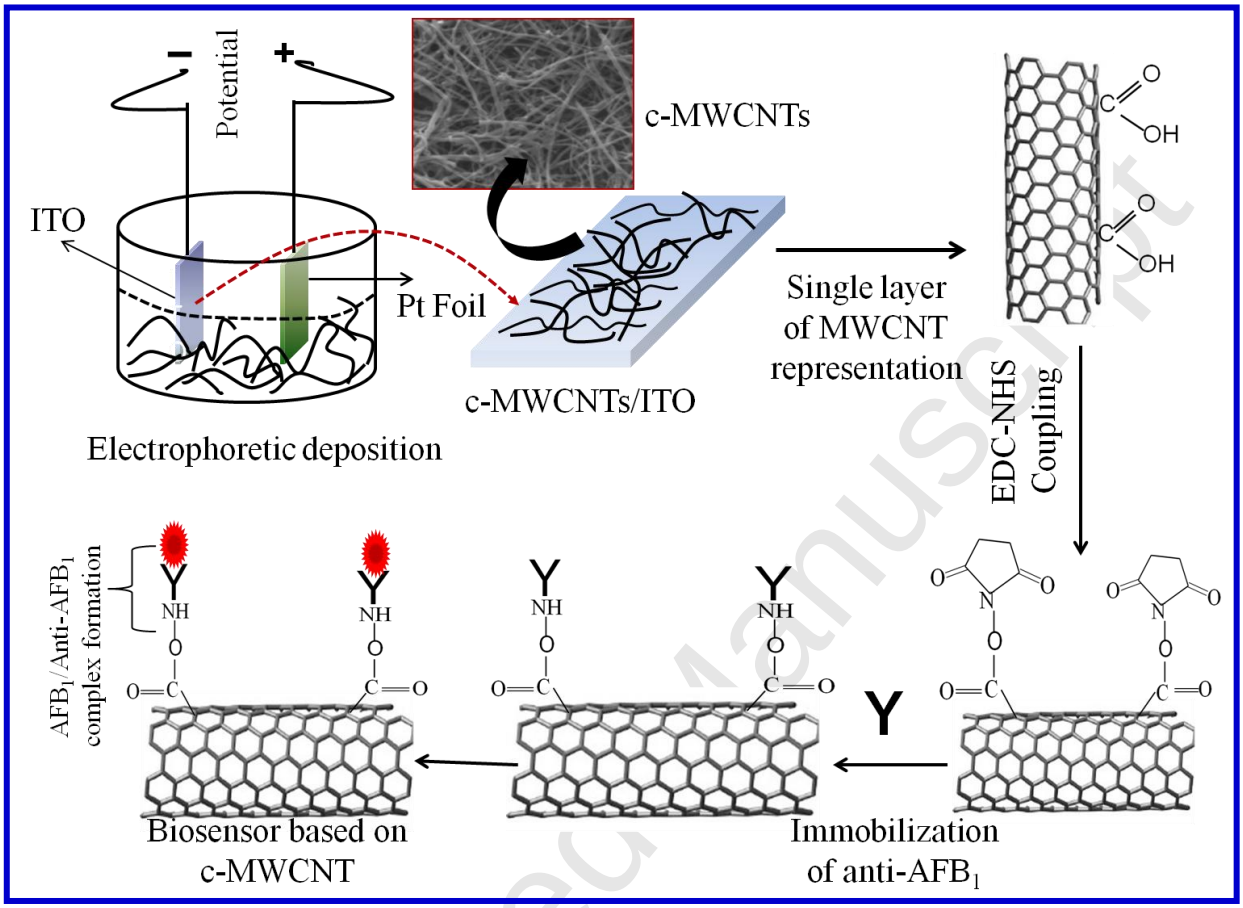



**Fig. 1**

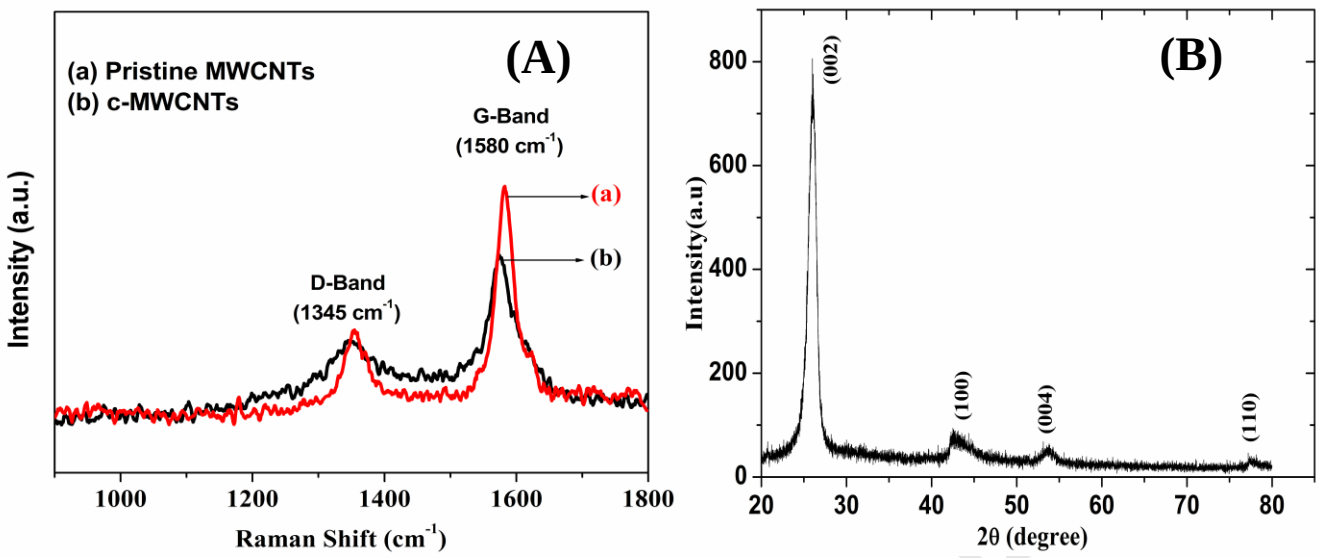



**Fig. 2**

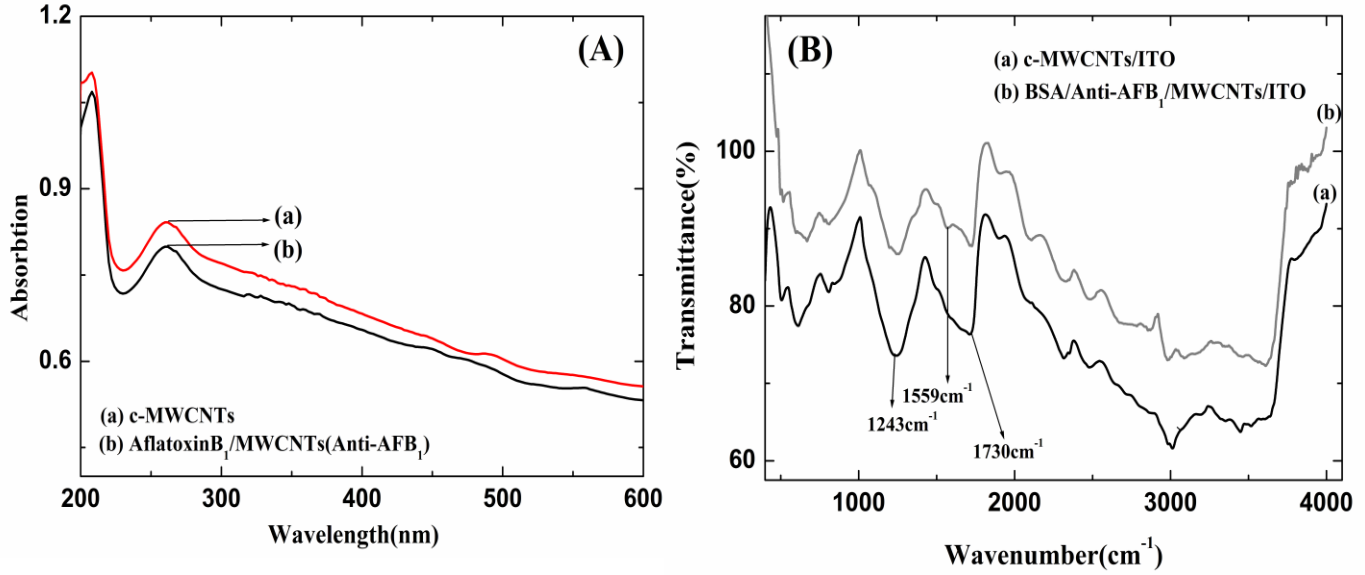



**Fig. 3**

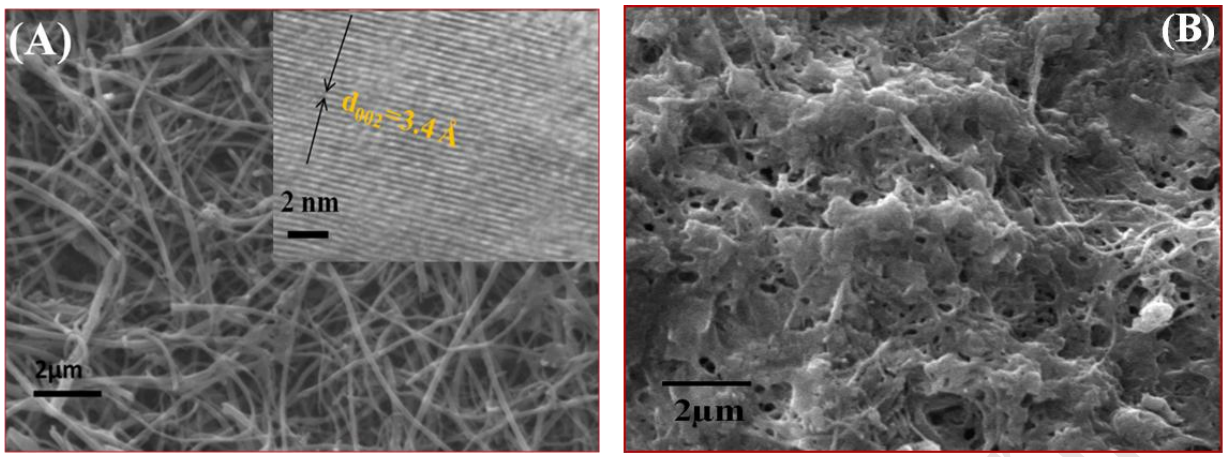



**Fig. 4**

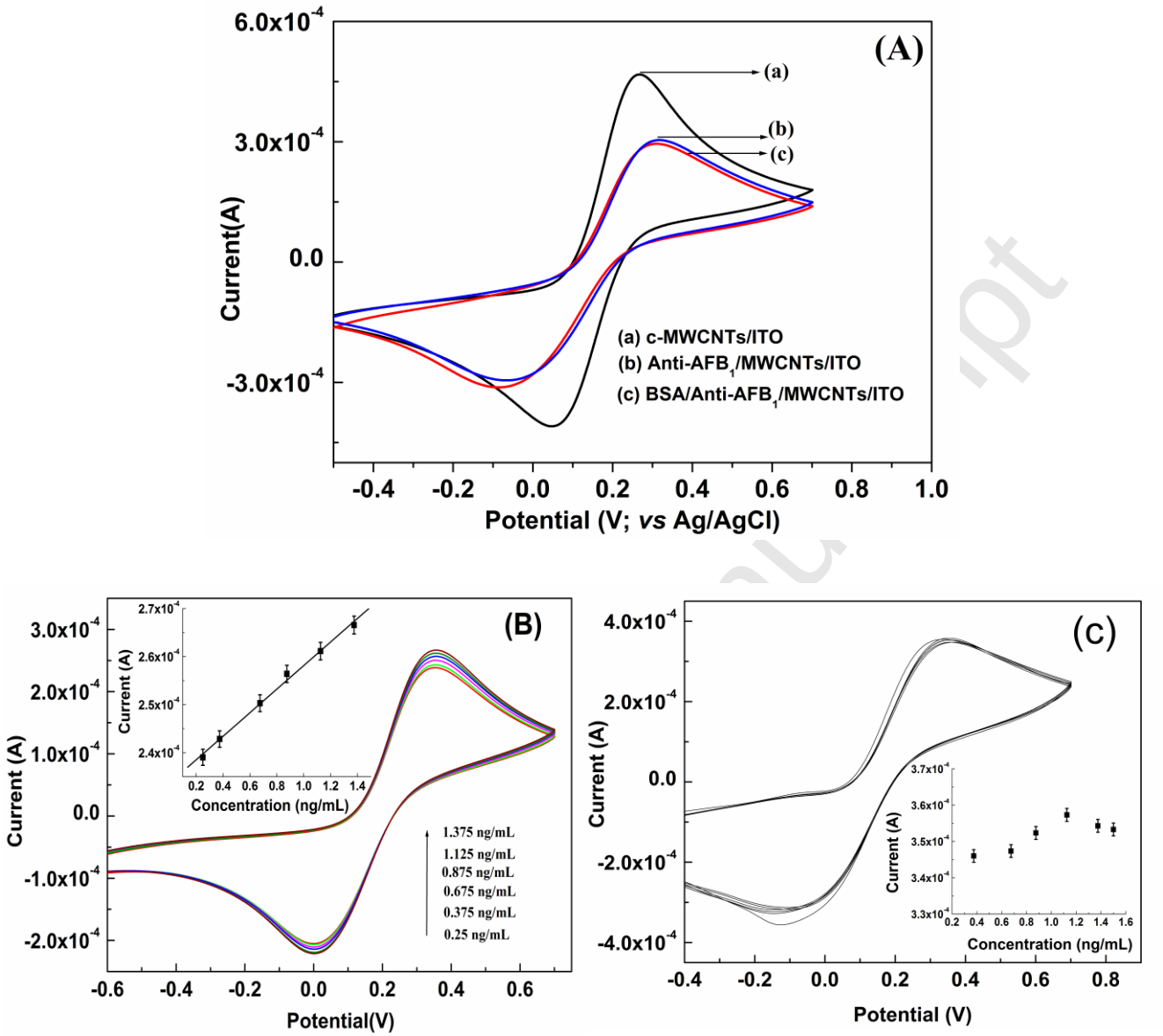



**Fig. 5**

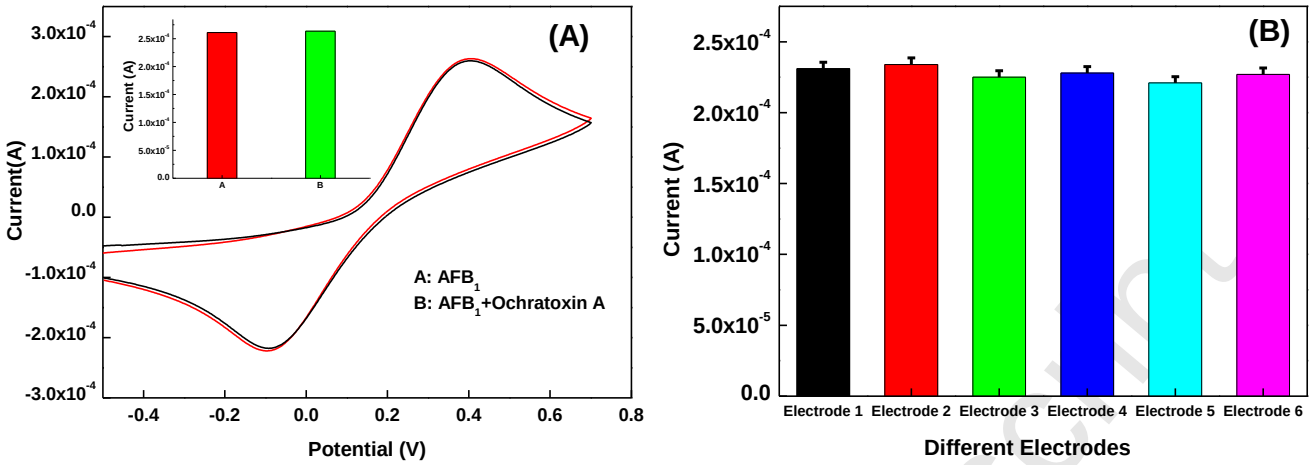



**Fig. 6**

Hindawi Publishing Corporation
ISRN Electronics
Volume 2013, Article ID 181869, 6 pages
http://dx.doi.org/10.1155/2013/181869

*Research Article*

# CDBA Based Universal Inverse Filter

**Rajeshwari Pandey,[1] Neeta Pandey,[1] Tushar Negi,[2] and Vivek Garg[2]**

[1] *Department of Electronics and Communication, Delhi Technological University, Bawana Road, Delhi 110042, India*
[2] *Department of Electrical Engineering, Delhi Technological University, Bawana Road, Delhi 110042, India*

Correspondence should be addressed to Rajeshwari Pandey; rajeshwaripandey@gmail.com

Received 30 November 2012; Accepted 3 February 2013

Academic Editors: H.-C. Chien, C. W. Chiou, and E. Tlelo-Cuautle



Current difference buffered amplifier (CDBA) based universal inverse filter configuration is proposed. The topology can be used to synthesize inverse low-pass (ILP), inverse high-pass (IHP), inverse band-pass (IBP), inverse band-reject (IBR), and inverse all-pass filter functions with appropriate admittance choices. Workability of the proposed universal inverse filter configuration is demonstrated through PSPICE simulations for which CDBA is realized using current feedback operational amplifier (CFOA). The simulation results are found in close agreement with the theoretical results.

## 1. Introduction

Inverse filters are commonly used in communication [1], speech processing, audio and acoustic systems [2, 3], and instrumentation [4] to reverse the distortion of the signal incurred due to signal processing and transmission. The transfer characteristics of the system that caused the distortion should be known a priori and the inverse filter to be used should have a reciprocal transfer characteristic so as to result in an undistorted desired signal. Literature review on inverse filter suggests that numerous well-established [5] methods for digital inverse filter design do exist but analog inverse filter design remained unexplored area as is evident from the limited availability of analog inverse filter circuits/design methods [6–14] until recently. However recent research trend suggests that the area is now gaining a renewed interest.

A brief account of the complete literature on analog inverse filter is presented here. Reference [6] presents a general method for obtaining the inverse transfer function for linear dynamic systems and the inverse transfer characteristic for nonlinear resistive circuits using nullors. The realization procedures for the current-mode FTFN-based inverse filters from the voltage-mode op-amp-based RC filters are presented in [7, 8]. The procedure outlined in [7] is applicable to planar circuits only as it uses RC : CR dual transformation, whereas the method presented in [8] makes use of adjoint transformation and thus is applicable to nonplanar circuits [12]. Single FTFN based inverse filters proposed in [9–12] present inverse filters using current feedback operational amplifier (CFOA). All the circuits presented in [10, 11] provide single inverse filter function; however [12] presents a topology which can realize inverse low-pass (ILP), inverse high-pass (IHP), and inverse band-pass (IBP) filter functions by appropriate admittance choice. In [13, 14] inverse all-pass (IAP) filters have been implemented using current difference transconductance amplifier (CDTA) and current conveyors, respectively. This study reveals that no universal inverse filter configuration has been proposed in the literature so far, to the best of the authors' knowledge. Therefore the aim of this paper is to present a current differencing buffered amplifier (CDBA) based universal inverse filter topology which realizes all five inverse filter functions, namely, ILP, IHP, IBP, IBR and IAP by appropriate admittance selection.

## 2. Circuit Description

Inherent wide bandwidth which is virtually independent of closed-loop gain, greater linearity, and large dynamic range are the key performance features of current mode technique [15]. The CDBA being a current processing analog building block inherits the advantages of current mode technique. In addition, it is free from parasitic capacitances [16] as its input terminals are internally grounded. Thus this active bolck is appropriate for high frequency operation. The circuit symbol

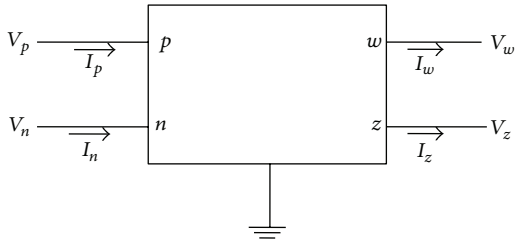



Figure 1: Block diagrammatic representation of CDBA.

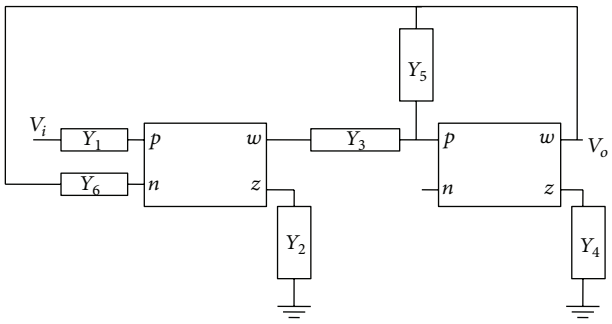



Figure 2: Proposed inverse filter configuration.

of CDBA is shown in Figure 1, and the port characteristics are given by

$$\begin{bmatrix} I_z \\ V_w \\ V_p \\ V_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_z \\ I_w \\ I_p \\ I_n \end{bmatrix}. \tag{1}$$

The proposed inverse filter configuration is shown in Figure 2. Routine analysis of the circuit of Figure 2 results in the following transfer function:

$$\frac{V_0(s)}{V_i(s)} = \frac{N(s)}{D(s)} = \frac{Y_1Y_3}{Y_2Y_4 + Y_3Y_6 - Y_2Y_5}, \tag{2}$$

where

$$\begin{aligned} N(s) &= Y_1Y_3, \\ D(s) &= Y_2Y_4 + Y_3Y_6 - Y_2Y_5. \end{aligned} \tag{3}$$

With the admittance choices of $Y_1 = sC_1 + G_1$ and $Y_3 = sC_3 + G_3$, the $N(s)$ can be expressed as

$$N(s) = s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3. \tag{4}$$

And the appropriate admittance choices for $Y_2$, $Y_4$, $Y_5$, and $Y_6$, as shown in Table 1, would result in the required denominator functions $D(s)$ and hence the required inverse filter responses.

Table 1

| Response type | $Y_2$ | $Y_4$ | $Y_5$ | $Y_6$ |
|---|---|---|---|---|
| ILP | $G_2$ | $G_4$ | 0 | 0 |
| IHP | $sC_2$ | $sC_4$ | 0 | 0 |
| IBP | $sC_2$ | $G_4$ | 0 | 0 |
| IBR | $G_2$ | $G_4$ | $sC_5$ | $sC_6$ |
| IAP | $G_2$ | $G_4$ | $sC_5$ | $sC_6$ |

Using the admittance choices given in Table 1, the ILP, IHP, and IBP response can, respectively, be expressed as

$$\begin{aligned} \frac{V_0}{V_i} &= \frac{1}{\left(G_2G_4\right)/\left(s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3\right)}, \\ \frac{V_0}{V_i} &= \frac{1}{\left(s^2C_2C_4\right)/\left(s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3\right)}, \\ \frac{V_0}{V_i} &= \frac{1}{\left(sC_2G_4\right)/\left(s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3\right)}. \end{aligned} \tag{5}$$

For admittance choices suggested for IBR and IAP in Table 1, the $D(s)$ can be written as

$$D(s) = s^2C_3C_6 + s\left(C_6G_3 - C_5G_2\right) + G_2G_5. \tag{6}$$

The resulting transfer function can be expressed as

$$\frac{V_0(s)}{V_i(s)} = \frac{1}{\dfrac{s^2C_3C_6 - s\left(C_5G_2 - C_6G_3\right) + G_2G_4}{s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3}} \tag{7}$$

which represents an IBR response if $C_5 = C_6 = C_1$ and $G_1G_3 = G_2G_5$. The response will be IAP if $C_5G_2 - C_6G_3 = C_1G_3 + C_3G_1$ which can be easily obtained by choosing

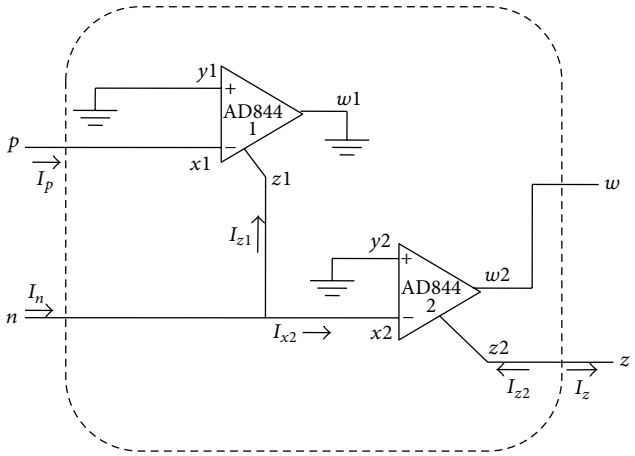



Figure 3: CDBA realization with AD844 [17].

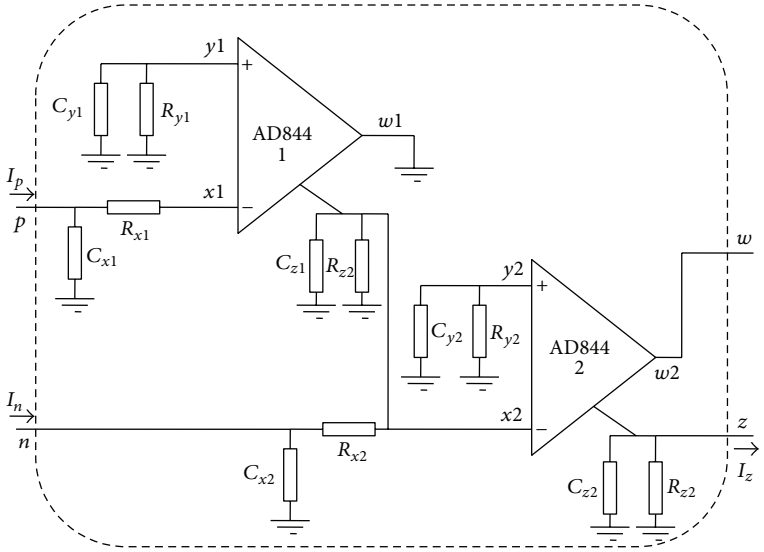



Figure 4: CDBA realization with nonideal model of AD844.

a suitable value of $C_5$. If $C_3 = C_6 = C_1 = C$, then $C_5 = 3C$ yields an IAP response provided $G_1 = G_2 = G_3$.

The resonant angular frequency ($\omega_0$) and the quality factor ($Q_0$) are given by (8) and (9), respectively, for all the responses

$$\omega_0 = \sqrt{\left(\frac{G_1 G_3}{C_1 C_3}\right)}, \tag{8}$$

$$Q_0 = \frac{\sqrt{C_1 C_3 G_1 G_3}}{(C_1 G_3 + G_1 C_3)}, \tag{9}$$

(whereas $H_{ILP}$, $H_{IHP}$, and $H_{IBP}$, the gain constants for ILP, IHP, and IBP responses, respectively, are given by

$$H_{ILP} = \frac{G_2 G_4}{G_1 G_3}, \qquad H_{IHP} = \frac{C_2 C_4}{C_1 C_3}, \qquad H_{IBP} = \frac{C_2 G_4}{(C_1 G_3 + C_3 G_1)}. \tag{10}$$

## 3. Sensitivity Analysis

The passive sensitivities of $\omega_0$ and $Q_0$ for the proposed configuration can be expressed as

$$S_{G_1}^{\omega_0} = S_{G_3}^{\omega_0} = \frac{1}{2}, \qquad S_{G_1}^{\omega_0} = S_{C_3}^{\omega_0} = -\frac{1}{2},$$
$$S_{G_1}^{Q_0} = S_{C_3}^{Q_0} = \frac{1}{2} - \frac{C_3 G_1}{(C_1 G_3 + C_3 G_1)}, \tag{11}$$
$$S_{G_3}^{Q_0} = S_{C_1}^{Q_0} = \frac{1}{2} - \frac{C_1 G_3}{(C_1 G_3 + C_3 G_1)}.$$

It is clearly observed from (11) that the passive sensitivities are lower than 1/2 in magnitude and hence the proposed universal inverse filter configuration may be termed as insensitive.

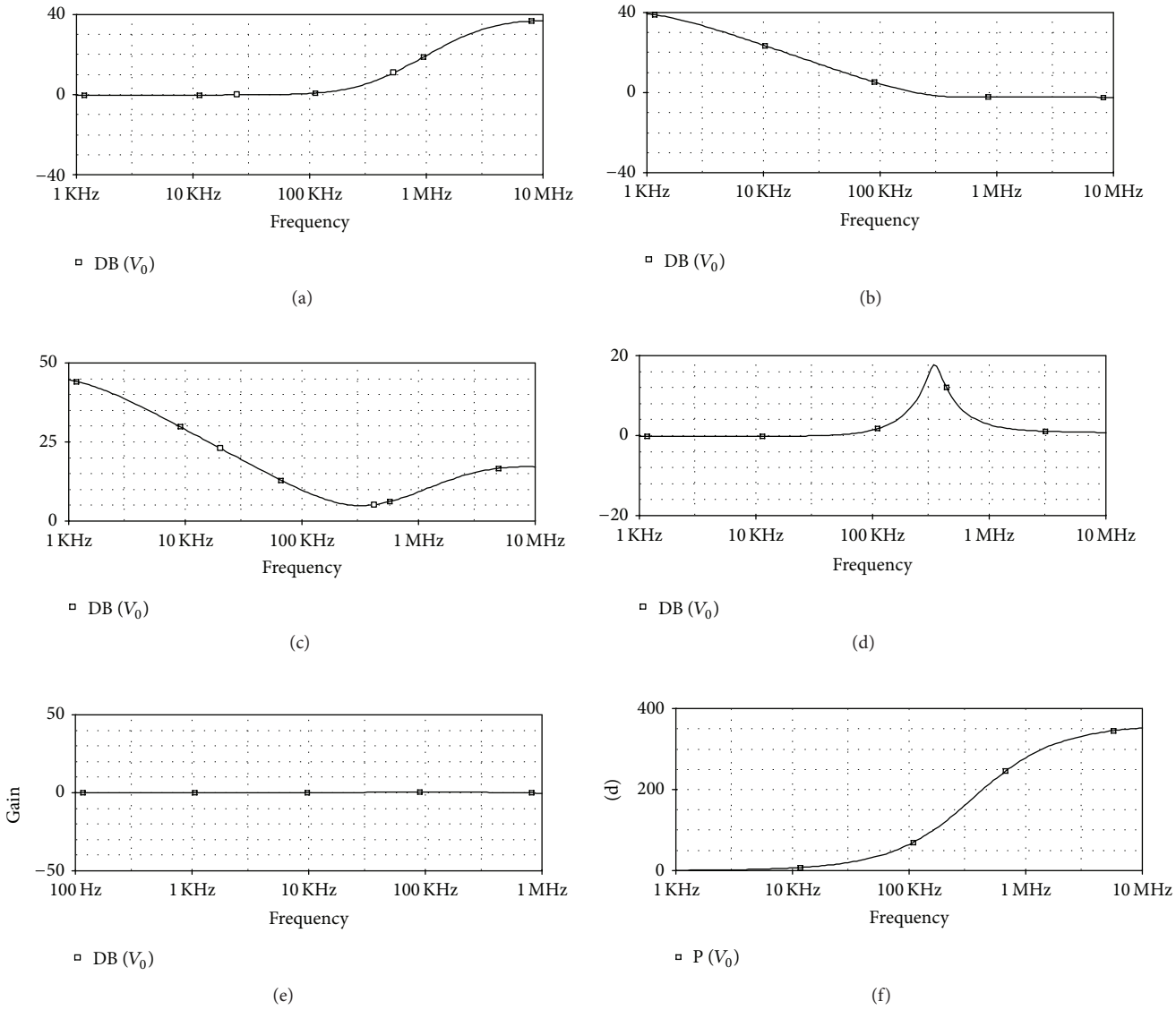



Figure 5: (a) Inverse low-pass response. (b) Inverse high-pass response. Inverse band-pass response. (d) Inverse band reject response. (e) Inverse all-pass magnitude response. (f) Inverse all-pass phase response.

## 4. Realizing a CDBA and Associated Nonideality Analysis

For the proposed configuration, the CDBA was realized using AD844 CFOA IC as shown in Figure 3 [17]. Ideally the input resistance at the $x$ terminal is zero and is infinite at the $z$ terminal. From Figure 3 various currents can be calculated as

$$\begin{aligned} I_{z1} &= I_p, \\ I_{x2} &= I_n - I_{z1}, \\ I_{z2} &= I_{x2}. \end{aligned} \tag{12}$$

Therefore the current from $z$ terminal $I_z$ can be calculated as

$$I_z = -I_{z2} = \left(I_p - I_n\right). \tag{13}$$

And the output voltage $V_w$ is given as

$$V_w = V_z. \tag{14}$$

In analysis so far, ideal characteristics of the CFOA have been considered. However, the effect of the parasitics of the CFOA needs to be taken into consideration for performing nonideality analysis [18–21]. For this, the model of AD844 [18] which includes a finite input resistance $R_x$ in series with $C_x$ at port-$x$, the $z$-port parasitic impedance ($R_z \| C_z$), and the $y$-port parasitic impedance ($R_y \| C_y$) is used. Using this nonideal model for CFOA, the CDBA structure of Figure 3

modifies to Figure 4. The nonideal transfer function of ILP from Figure 4 can be expressed as

$$\frac{V_0}{V_i} = \left(s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3\right) / \\ G_2'G_4'\left(1 + \frac{sC_{z1}}{G_2'}\right)\left(1 + \frac{sC_{z2}}{G_4'}\right) \\ \times\left(1 + \frac{s}{G_{x1}}\left(\left(C_{x1} + C_1\right) + G_1\right)\right) \\ \times\left(1 + \frac{s}{G_{x2}}\left(\left(C_{x2} + C_3\right) + G_3\right)\right), \tag{15}$$

where $G_2' = 1/(R_2 \| R_{z1})$ and $G_4' = 1/(R_4 \| R_{z2})$.

Considering $G_{x1} \gg G_2$ and $G_{x2} \gg G_5$, (15) modifies to

$$\frac{V_0}{V_i} = \left(s^2C_1C_3 + s\left(C_1G_3 + C_3G_1\right) + G_1G_3\right) / \\ G_2'G_4'\left(1 + \frac{sC_{z1}}{G_2'}\right)\left(1 + \frac{sC_{z2}}{G_4'}\right) \\ \times\left(1 + \frac{s\left(G_{x1} + C_1\right)}{G_{x1}}\right)\left(1 + \frac{s\left(G_{x2} + C_3\right)}{G_{x2}}\right). \tag{16}$$

It is clear from (16) that nonidealities of CFOA introduce parasitic poles in the transfer function. The deviation from the ideal behavior so caused can be kept small if all the external resistors are chosen to be much larger than $R_x$ but much smaller than $R_y$ and $R_z$. Similarly external capacitors should be chosen to be much larger than $C_y$ and $C_z$. Nonideal transfer functions for IHP and IBP can also be deduced in a similar manner.

## 5. Simulation Results

The proposed theoretical predictions are validated through simulations using PSPICE macromodel of CFOA AD844 IC as shown in Figure 3. Supply voltages used are ±10 V. The proposed inverse filter configuration is designed with equal value components. All the resistances of value 10 KΩ and capacitors of value 50 pF are chosen resulting in a theoretical $f_0$ of 318.5 KHz. Simulated frequency magnitude responses for ILP, IHP, IBP, IBR, and IAP are shown from Figure 5(a) to Figure 5(e), respectively, whereas Figure 5(f) shows the phase response for IAP. The simulated $f_0$ for all the responses is found to be 316.3 KHz and is in close agreement to the theoretical value.

## 6. Conclusion

A current difference buffered amplifier (CDBA) based universal inverse filter configuration is proposed. Appropriate admittance selections allow using the proposed topology as one of the five inverse filter configurations, namely, ILP, IHP, IBP, IBR, and IAP. Workability of the proposed universal inverse filter configuration is demonstrated through PSPICE simulations for which CDBA is realized using current feedback operational amplifier (CFOA). The simulation results are found in close agreement with the theoretical results.

## References


[1] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 658–672, 1997.

[2] A. Watanabe, "Formant estimation method using inverse-filter control," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 4, pp. 317–326, 2001.

[3] O. Kirkeby and P. A. Nelson, "Digital filter design for inversion problems in sound reproduction," *Journal of the Audio Engineering Society*, vol. 47, no. 7-8, pp. 583–595, 1999.

[4] Z. Zhang, D. Wang, W. Wang, H. Du, and J. Zu, "A group of inverse filters based on stabilized solutions of fredholm integral equations of the first kind," in *Proceedings of the IEEE International Instrumentation and Measurement Technology Conference*, pp. 668–671, May 2008.

[5] G. John Proakis and G. Dimitris Manolakis, *Digital Signal Processing*, Prentice Hall, New York, NY, USA, 4th edition, 2007.

[6] A. Leuciuc, "Using nullors for realisation of inverse transfer functions and characteristics," *Electronics Letters*, vol. 33, no. 11, pp. 949–951, 1997.

[7] B. Chipipop and W. Surakampontorn, "Realization of current-mode FTFN-based inverse filter," *Electronics Letters*, vol. 35, no. 9, pp. 690–692, 1999.

[8] H. Y. Wang and C. T. Lee, "Using nullors for realisation of current-mode FTFN-based inverse filters," *Electronics Letters*, vol. 35, no. 22, pp. 1889–1890, 1999.

[9] M. T. Abuelma'atti, "Identification of cascadable current-mode filters and inverse-filters using single FTFN," *Frequenz*, vol. 54, no. 11-12, pp. 284–289, 2000.

[10] S. S. Gupta, D. R. Bhaskar, R. Senani, and A. K. Singh, "Inverse active filters employing CFOAs," *Electrical Engineering*, vol. 91, no. 1, pp. 23–26, 2009.

[11] S. S. Gupta, D. R. Bhaskar, and R. Senani, "New analogue inverse filters realised with current feedback op-amps," *International Journal of Electronics*, vol. 98, no. 8, pp. 1103–1113, 2011.

[12] H.-Y. Wang, S.-H. Chang, T.-Y. Yang, and P.-Y. Tsai, "A novel multifunction CFOA-based inverse filter," *Circuits and Systems*, vol. 2, pp. 14–17, 2011.

[13] N. A. Shah, M. Quadri, and S. Z. Iqbal, "High output impedance current-mode allpass inverse filter using CDTA," *Indian Journal of Pure and Applied Physics*, vol. 46, no. 12, pp. 893–896, 2008.

[14] N. A. Shah and M. F. Rather, "Realization of voltage-mode CCII-based allpass filter and its inverse version," *Indian Journal of Pure and Applied Physics*, vol. 44, no. 3, pp. 269–271, 2006.

[15] C. Toumazu, F. J. Lidgey, and D. G. Haigh, *Analogue IC Design: The Current Mode Approach*, IEEE Circuits and Systems Series, Peringrinus, 1990.

[16] C. Acar and S. Ozoguz, "A new versatile building block: current differencing buffered amplifier suitable for analog signal-processing filters," *Microelectronics Journal*, vol. 30, no. 2, pp. 157–160, 1999.

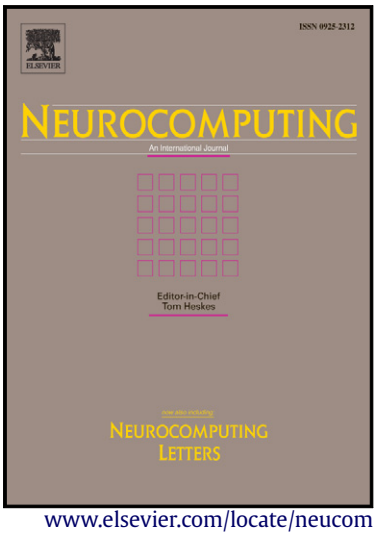
[17] S. Özcan, A. Toker, C. Acar, H. Kuntman, and O. Çiçekoglu, "Single resistance-controlled sinusoidal oscillators employing current differencing buffered amplifier," *Microelectronics Journal*, vol. 31, no. 3, pp. 169–174, 2000.

[18] AD844 Current Feedback Op-Amp Data Sheet, Analog Devices Inc., Norwood, NJ, USA, 1990.

[19] C. Sánchez-López, F. V. Fernández, E. Tlelo-Cuautle, and S. X. D. Tan, "Pathological element-based active device models and their application to symbolic analysis," *IEEE Transactions on Circuits and Systems I*, vol. 58, no. 6, pp. 1382–1395, 2011.

[20] E. Tlelo-Cuautle, C. Sánchez-López, and D. Moro-Frías, "Symbolic analysis of (MO)(I)CCI(II)(III)-based analog circuits," *International Journal of Circuit Theory and Applications*, vol. 38, no. 6, pp. 649–659, 2010.

[21] R. Trejo-Guerra, E. Tlelo-Cuautle, C. Sánchez-López, J. M. Muñoz-Pacheco, and C. Cruz-Hernández, "Realization of multiscroll chaotic attractors by using current-feedback operational amplifiers," *Revista Mexicana de Fisica*, vol. 56, no. 4, pp. 268–274, 2010.




www.elsevier.com/locate/neucom

# Color Segmentation by Fuzzy Co-clustering of chrominance color features

Madasu Hanmandlu, Om Prakash Verma, Seba Susan, V.K. Madasu

PII: S0925-2312(13)00336-6
DOI: http://dx.doi.org/10.1016/j.neucom.2012.09.043
Reference: NEUCOM13301

To appear in: *Neurocomputing*

Received date: 19 December 2011
Revised date: 19 September 2012
Accepted date: 22 September 2012

Cite this article as: Madasu Hanmandlu, Om Prakash Verma, Seba Susan, V.K. Madasu, Color Segmentation by Fuzzy Co-clustering of chrominance color features, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2012.09.043

# Color Segmentation by Fuzzy Co-clustering of chrominance color features

Madasu Hanmandlu, Senior Member IEEE, Om Prakash Verma , Seba Susan and V.K. Madasu



*Abstract*—**This paper presents a novel color segmentation technique using fuzzy co-clustering approach in which both the objects and the features are assigned membership functions. An objective function which includes a multi-dimensional distance function as the dissimilarity measure and entropy as the regularization term is formulated in the proposed fuzzy co-clustering for images (FCCI) algorithm. The chrominance color cues a* and b* of CIELAB color space are used as the feature variables for co-clustering. The experiments are conducted on 100 natural images obtained from the Berkeley segmentation database. It is observed from the experimental results that the proposed FCCI yields well formed, valid and high quality clusters, as verified from Liu's F-measure and Normalized Probabilistic RAND index. The proposed color segmentation method is also compared with other segmentation methods namely Mean-Shift, NCUT, GMM, FCM and is found to outperform all the methods. The bacterial foraging global optimization algorithm gives image specific values to the parameters involved in the algorithm.**

***Index Terms*— Fuzzy Clustering, Co-clustering, Object membership, Feature membership, Validity measure, Bacterial Foraging, Color segmentation**

## I. INTRODUCTION

The segmentation of color images is a potential area of research due to its practical significance in various fields. Image segmentation partitions the image into regions/segments such that pixels belonging to a region are more similar to each other than those belonging to different regions. Clustering is a well known approach for segmenting images. It strives to assess the relationships among patterns of the data set by organizing them into groups or clusters such that patterns within a cluster are more similar to each other than those belonging to different clusters. Many algorithms for both hard and fuzzy clustering have

M. Hanmandlu (email: mhmandlu@iitd.ac.in) and Seba Susan (email: sebasusan@indiatimes.com) are with IIT Delhi, New Delhi, India, O.P. Verma (email: opverma.dce@gmail.com) is with Delhi Technological Univ., Delhi, India, and V.K. Madasu(email: v.madasu@uq.edu.au) is with Queenland Univ,, Brisbane, Australia.

been developed to achieve this purpose. In hard clustering, data is divided into crisp clusters, where each data belongs to exactly one cluster. In fuzzy clustering, the data points can belong to more than one cluster, and associated with each of the points are membership grades that indicate the degree to which the data points belong to the different clusters. Clustering in the color domain gives improved segmentation results since color components carry more information than the gray scale components.

Several techniques have been proposed in the field of color segmentation. Histogram based segmentation of color [1] is one of the existing techniques. But it doesn't guarantee contiguity of the resulting regions. Edge detection based techniques [2] pose the difficulty of determining the boundary of an image due to the ambiguity of the response of a weak edge. Recently, Arbelaez *et al.* in [3] have proposed a hierarchical segmentation obtained from the output of a contour detector which overcomes the difficulties of weakly linked boundaries. In [4] color segmentation by region growing and merging is investigated. One drawback of the conventional region growing technique is the selection of the seed point and the order in which regions grow or merge. In [5], the problem of seed selection is solved by using the relaxation labeling technique which yields satisfactory results. Recent techniques for region growing use automated seed selection process as in [6] which uses a fuzzy similarity and fuzzy distance based approach. In [7], after the region growing of similar color, Markov Random Fields (MRF) are applied to improve the results. However, it is observed that some homogeneous regions may get disconnected due to the MRF process. Blobworld[8], a popular image segmentation and retrieval algorithm groups pixels into regions by modeling the joint distribution of color texture and position features by a mixture of Gaussians with parameters being decided by the expectation maximization algorithm. However, the resulting blobs may not contain all the details of objects and also may not distinguish an object which is not visually distinct. Further an iterative post processing step is required to correct the mis-alignment of object boundaries. Mean-Shift filtering [9] and Graph partitioning[10,11] methods and their hybrids[12] perform clustering in feature–space and are found to be effective for color segmentation. But they are very sensitive to the parameters like color bandwidth (Mean-Shift) and the threshold edge length (Graph method). Neural network based approaches [13,14] for image segmentation like Competitive Learning Neural Network (CLNN) and Self Organizing of Kohonen Feature Map (SOFM) avoid complex programming but usually consume a lot of training time. Other significant works on image segmentation include: Watershed technique[15] based on the morphological watershed transform, segmentation using K-nearest neighbor (K-NN) technique[16] which is sensitive to the choice of reference sample and JSEG[17]-a segmentation algorithm based on color and texture. Various combinations of popular segmentation algorithms like region merging and graph partitioning[18], mean-shift and region merging[19], watershed and Kohonen SOM[20] have been suggested together with their advantages. In [21], color segmentation is carried out by applying a set of fuzzy if-then-rules on 200 fixed color samples. The Fuzzy C-Means (FCM) clustering method, a popular choice for color segmentation has been investigated in the works of [22]. The results are quite good but for the computational complexity and sensitivity to the initialization. Several variants of FCM are summarized in [14]. In [23] fuzzy set theory and

maximum fuzzy entropy principle are used to convert the image to the fuzzy domain and a Space Scale filter is used to analyze the homogeneity histogram to find the appropriate segments. Fuzzy co-clustering algorithm with its dual fuzzy (object and feature) membership functions was originally derived for document clustering, examples being FCCM, FCoDoK [24,25] and robust versions PFCC[26], RFCC[27]. The co-clustering done so far on images [28,29] is limited to indexing of images for Content based image retrieval (CBIR) in which low level semantic features derived from image histogram are the feature variables for clustering.

In this paper the Fuzzy co-clustering approach is adapted for the segmentation of natural images. An algorithm for the Fuzzy Co-clustering of images (FCCI) is developed by incorporating the distance between each feature data point and the feature cluster center as the dissimilarity measure and the entropies of the objects and features as the regularization terms in the objective function. To prove the effectiveness of our approach we apply the FCCI algorithm for the segmentation of color images with successful results. Some preliminary work on color segmentation of histo-pathological images is reported in [30] and this serves as a precursor to the main work. The CIELAB color space is favored for our experiments due to the wide range of colors possible and its closeness to the human perception system[31], and its chrominance color vector {a*,b*} is proved to be the best feature combination for the segmentation task [32]. It is found from the experimental results that the color segmentation results obtained by the proposed technique are of high quality with respect to both Liu's evaluation measure and NPR index which are global segmentation evaluation measures and also outperforms over other popular color segmentation methods. The choice of the number of clusters for the experiment is determined in a novel manner by plotting Xie and Beni's cluster validity [33] as the number of clusters is increased and by checking for the first local minima of the curve. The resulting segmentation offers a good tradeoff between color differencing and human perception.

The organization of the paper is as follows: The Fuzzy Co-Clustering algorithm for Images (FCCI) is introduced in Section II. While minimizing the objective function in FCCI, Bacterial Foraging is adapted for global learning of parameters. Later in Section III, the proposed algorithm is applied for the color segmentation together with the state of the art comparisons. Finally conclusions from over-all results are given in Section IV.

## II. FUZZY CO-CLUSTERING ALGORITHM FOR IMAGES

**Motivation for the algorithm and related work:**

Co-clustering simultaneously clusters both objects and features together [24]. This provides two membership functions: the partition or object membership function and the ranking or the feature membership function. The latter serves to filter out the relevant features only during the computation of the object membership function and thus solves the problem of sparseness of data by reducing the dimensionality. The co-clustering algorithm is thus suited to applications with large dimensions and is found to be apt for our experiments on multi-feature color images. The problem of outliers is also minimized by using feature membership function [26]. The problem with using only the feature memberships is that it may lead to coincident/overlapping clusters therefore highlighting the need for both feature and object memberships. Further we include the distance function of feature data points from the feature cluster centroids in the co-clustering process to create richer co-clusters than other fuzzy co-clustering algorithms. The inclusion of the distance factor in the degree of aggregation reduces the optimization problem to a minimization one. In this work co-clustering is integrated with the Fuzzy approach with a view to obtaining distinct clusters [24,26]. Both the object and feature memberships in the proposed method are fuzzy, i.e. the object membership is calculated when different clusters compete for a data point and feature memberships are defined when different features compete for a cluster. Thus we have two constraints on the two fuzzy memberships (object and feature memberships) in our method.

Therefore the aim is to have a co-clustering algorithm with the following advantages:

1. It must be insensitive to initialization and form distinct clusters. (Fuzzy clustering)
2. It should perform well in high dimensions and provide well defined clusters. (Co-clustering)
3. It should minimize the impact of outliers to improve the accuracy of co-clustering. (ranking/feature memberships)
4. Its objective function should integrate the distance measure of input features w.r.t. feature centroids into the entropy regularization framework.
5. It must be reasonably fast enough.

Several maximum entropy clustering algorithms and their variants are available in the literature [34,35]. One such approach of interest to the present work is the variant of FCM which props on entropy regularization [36]. It involves the minimization of the following objective function,

$$J_{FCM} = \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} Dist(x_i, p_c) + T_U \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} \log u_{ci} \quad (1)$$

subject to the constraint

$$\sum_{c=1}^{C} u_{ci} = 1, \qquad u_{ci} \in [0,1], \quad \forall i = 1,\ldots,N \quad (2)$$

Where the symbols represent:

---

C,N: The number of clusters and data points respectively

$u_{ci}$: Fuzzy membership function

$T_U$ : The weight factor in the entropy term

$Dist(x_i, p_c)$ : The dissimilarity term equal to the square of the Euclidean distance between pixel $x_i$ and cluster center $p_c$.

---

The first term in the R.H.S of (1) denotes the effective squared distance; the second term is the entropy which serves as a regulating factor during the minimization process. The proposed approach aims at co-clustering in the entropy framework of FCM. For this we begin by replacing the distance function $Dist(x_i, p_c)$ with $Dist(x_{ij}, p_{cj})$ in the proposed objective function. $Dist(x_{ij}, p_{cj})$ is computed for each feature j=1,2,..,K separately.

**Formulating the objective Function:**

Let $X = \{x_1, x_2, ..., x_i, ...., x_N\} \in \mathbb{R}^K$ be the set of N data points associated with an image I of size $N_1 \times N_2 = N$ , and K is the dimension of the feature space associated with each data point. Let $x_{ij}$ denote the j[th] feature of the i[th] data point , $P = \{p_{cj}\}$ be the set of feature cluster centers and $D_{cij} = Dist(x_{ij}, p_{cj})$ be the square of the Euclidean distance between feature data point $x_{ij}$ and the feature cluster centroid $p_{cj}$ given by

$$D_{cij} = d^2(x_{ij}, p_{cj}) = (x_{ij} - p_{cj})^2 \quad (3)$$

Let $u_{ci}$ denote the object membership of the i[th] data point to cluster c, $U = \{u_{ci}\}$ be the C×N object membership function matrix of image I, $v_{cj}$ denote the feature membership defined as the membership of feature j to the c[th] cluster and $V = \{v_{cj}\}$ be the corresponding C×K feature membership matrix for image I.

Including the feature membership function $v_{cj}$ in the first term of (1) and replacing the distance function by $D_{cij} = Dist(x_{ij}, p_{cj})$ yields $\sum_{c=1}^{C} \sum_{i=1}^{N} \sum_{j=1}^{K} u_{ci} v_{cj} D_{cij}$, which is regarded as the degree of aggregation of the proposed objective

function $J_{FCCI}$. Separate entropy regularizing terms $\sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} log u_{ci}$ and $\sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} log v_{cj}$ for the object and feature membership functions constitute the second and third terms of $J_{FCCI}$ respectively. Minimizing these two terms is equivalent to maximizing the fuzzy entropies $-\sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} log u_{ci}$ and $-\sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} log v_{cj}$. The entropies are maximized when the fuzzy memberships $u_{ci}$ and $v_{cj}$ are uniformly distributed according to their constraints i.e. $u_{ci} = \frac{1}{C}$ and $v_{cj} = \frac{1}{K}$.

The objective function $J_{FCCI}$ resulting from combining all the above terms is:

$$J_{FCCI}(\mathbf{U,V,P}) = \sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ci} v_{cj} D_{cij} + T_U \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} log u_{ci} + T_V \sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} log v_{cj} \quad (4)$$

subject to the following constraints:

$$\sum_{c=1}^{C} u_{ci} = 1, \qquad u_{ci} \in [0,1], \quad \forall i = 1,...,N \quad (5)$$

$$\sum_{j=1}^{K} v_{cj} = 1, \qquad v_{cj} \in [0,1], \quad \forall c = 1,...,C \quad (6)$$

The minimization of the first term in (4) assigns to the object a higher membership value taking into account the feature cluster center it is closest to and which is more relevant than other features for that particular cluster. The inner product $\{v_{cj} D_{cij}\}$ assigns a higher weight to the distance function pertaining to the prominent features and a lower weight to that of the irrelevant features. The first term therefore denotes the effective squared distance. The second and third entropy regularization terms combine all $u_{ci}$'s and $v_{cj}$'s separately. These contribute to the fuzziness in the resulting clusters. $T_U$ and $T_V$ are the weighting parameters that specify the degree of fuzziness. Increasing $T_U$ and $T_V$ increases the fuzziness of the clusters.

**Deriving the update equations:**

The constrained optimization problem of FCCI can now be defined from (4) by applying the Lagrange multipliers $\lambda_i$ and $\gamma_c$ to constraints (5) and (6) respectively as shown below.

$$J(\mathbf{U,V,P}) = \sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ci} v_{cj} D_{cij} + T_U \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} log u_{ci} + T_V \sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} log v_{cj} + \sum_{i=1}^{N} \lambda_i \left(\sum_{c=1}^{C} u_{ci} - 1\right) + \sum_{c=1}^{C} \gamma_c \left(\sum_{j=1}^{K} v_{cj} - 1\right) \quad (7)$$

Taking the partial derivative of J(U,V,P) in (7) with respect to U and setting the gradient to zero we have,

$$\frac{\partial J}{\partial \mathbf{U}} = \sum_{j=1}^{K} v_{cj} D_{cij} + T_U (1 + \log u_{ci}) + \lambda_i = 0 \quad (8)$$

Subjecting $u_{ci}$ derived from (8) to the constraint in (5) the formula for computing the object membership function $u_{ci}$ reduces to,

$$u_{ci} = \frac{e^{(-\sum_{j=1}^{K} \frac{v_{cj} D_{cij}}{T_U})}}{\sum_{c=1}^{C} e^{(-\sum_{j=1}^{K} \frac{v_{cj} D_{cij}}{T_U})}} \tag{9}$$

In a similar manner, taking the partial derivative of J(U,V,P) with respect to V and setting the gradient to zero we have,

$$\frac{\partial J}{\partial \boldsymbol{V}} = \sum_{i=1}^{N} u_{ci} D_{cij} + T_V (1 + \log v_{cj}) + \gamma_c = 0 \tag{10}$$

Applying the constraint (6) to $v_{cj}$ derived from (10), we obtain the formula for the feature membership function $v_{cj}$ as:

$$v_{cj} = \frac{e^{(-\sum_{i=1}^{N} \frac{u_{ci} D_{cij}}{T_V})}}{\sum_{j=1}^{K} e^{(-\sum_{i=1}^{N} \frac{u_{ci} D_{cij}}{T_V})}} \tag{11}$$

Taking the partial derivative of J(U,V,P) with respect to P and setting the gradient to zero we have:

$$\frac{\partial J}{\partial \boldsymbol{P}} = v_{cj} \sum_{i=1}^{N} u_{ci} x_{ij} - v_{cj} p_{cj} \sum_{i=1}^{N} u_{ci} = 0 \tag{12}$$

Solving (12) yields the formula for $p_{cj}$ as :

$$p_{cj} = \frac{\sum_{i=1}^{N} u_{ci} x_{ij}}{\sum_{i=1}^{N} u_{ci}} \tag{13}$$

The solution of the constrained optimization problem in (7) can be approximated by Picard iteration or Alternating Optimization (AO) [37] through (9), (11) and (13) which are the update equations for the object, feature memberships and the cluster centroids respectively in each iteration. Optimal partitions U* of X can be obtained by solving for (U*,V*,P*) at the local minima of $J_{FCCI}$. The proof of convergence of the FCCI algorithm to a local optimum is given in Appendix I. Since U*,V*and P* are unknowns, the objective function in (4) is neither concave nor convex and usually has many local optima. To find a global optimum of the constrained optimization problem, the FCCI algorithm is further given as a learning step to the Bacterial Foraging algorithm which optimizes the values of the weight parameters $T_U$ and $T_V$ .

**Pseudo-code of FCCI Algorithm**

1.Initialize the parameters $T_U$, $T_V$, maximum error limit ε and maximum number of iterations $\tau_{max}$.

2. Set iteration number $\tau = 1$.

3. Initialize $u_{ci}$ such that $0 \le u_{ci} \le 1$.

4. REPEAT

5. Calculate $p_{cj}$ using (13)

6. Calculate $D_{cij}$ using (3).

7. Calculate $v_{cj}$ using (11).

8. Calculate $u_{ci}$ using (9).

9. Calculate τ= τ+1.

10. UNTIL

$max(|u_{ci}(\tau) - u_{ci}(\tau - 1)|) \le \varepsilon$ or τ= $\tau_{max}$

---

Since all our experiments are found to converge within 200 iterations, $\tau_{max}$ =200 and the maximum error limit ε taken to be $10^{-2}$.

## III. COLOR IMAGE SEGMENTATION USING FCCI

### *A. Algorithm for color segmentation*

**Review of Xie and Beni's Cluster validity S:**

According to Xie and Beni[33] the validity function S of the clusters for the worst case is defined by:

$$S = \frac{\sigma / N}{d_{min}^{\,2}} \tag{14}$$

The $d_{min}$ in (14) is evaluated from

$$d_{min} = \min_{\forall c} \{ \sum_{j=1}^{K} (p_{(c+1)j} - p_{cj})^2 \} \tag{15}$$

where $d_{min}$ is the minimum distance between the cluster centroids $p_{cj}$ for cluster c = 1,..,C and feature *j* and σ is the maximum variation among all the clusters C, given by:

$$\sigma = \max_{\forall c} \{ \sum_{i=1}^{N} u_{ci}^{\,2} \sum_{j=1}^{K} (x_{ij} - p_{cj})^2 \} \tag{16}$$

To determine the number of clusters based on the above Cluster Validity S a flowchart is given in Fig. 1 which checks for the occurrence of the first local minima of S.

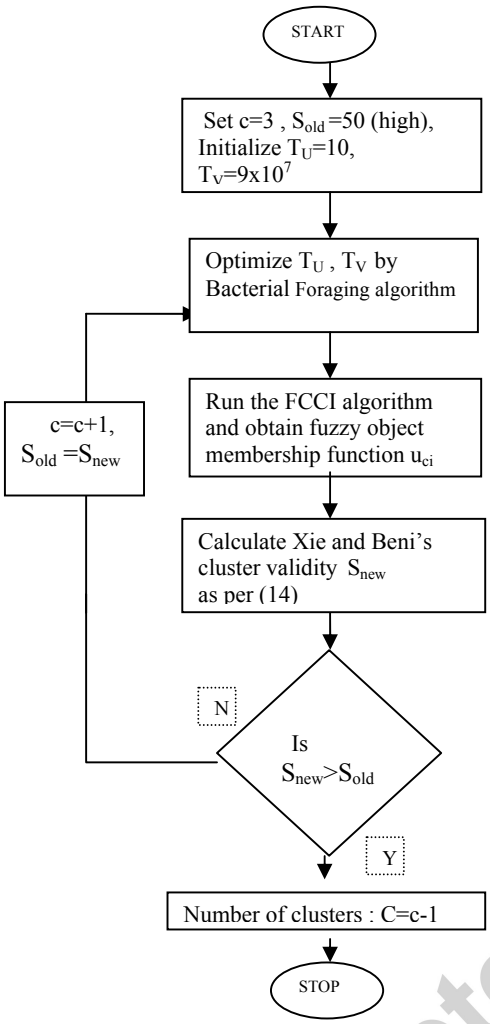



**Fig 1.** : *Flowchart for determining the number of clusters*



**Algorithm for Color Image Segmentation using FCCI**

The algorithm is outlined as follows:

---

1. Obtain the three dimensional RGB input image
2. Convert RGB color space into the CIELAB color space with color dimensions K= 2, i.e . {a*,b*}.
3. Perform 2D to 1D transformation [38] (by lifting the elements columnwise) to generate data point $x_{ij}$ in the $j^{th}$ dimension, j=1,2 for each pixel i=1,…,N, where N is the size of the data. This step is important since the computations become simpler when data is one dimensional rather than two dimensional.
4. Determine the number of clusters C as per the flowchart in Fig.1.

5. Run the FCCI algorithm for C clusters and obtain the object $u_{ci}$ membership function.
6. Defuzzify $u_{ci}$ into clusters.

---

*B. Bacterial Foraging for the global minimum*

The Bacterial Foraging (BF)[39] based on the bacterial chemotactic behavior of *E.Coli* is used for optimizing the values of fuzzy parameters $T_U$ and $T_V$ in the FCCI algorithm. The BF algorithm initially accepts a set of initial values from the user before optimizing these to global minimum values by subsequent iterations. The following initial values: $T_U$ =10, $T_V$ = 9 x $10^7$ are assumed for the color segmentation experiments. The choice of these initial values is made by conducting a set of random experiments by hit and trial. Bacterial Foraging is treated as an optimization process [40], where each bacterium seeks to maximize the energy obtained per unit time spent on foraging. Suppose that θ is the position of a bacterium and let $J_D(\theta)$ represent the combined effects of attractants and repellents from the environment, for e.g. $J_D(\theta)<0$, $J_D(\theta)=0$, $J_D(\theta)>0$, representing that the bacterium at location is in nutrient rich, neutral, and noxious environments, respectively.

Chemotaxis is a foraging behavior that implements a type of optimization where bacteria try to climb up the nutrient concentration (i.e. find lower and lower values of $J_D(\theta)<0$ termed as swim), avoid noxious substances (for $J_D(\theta)>0$ termed as tumble), and search for ways out of neutral media (avoid being at positions where $J_D(\theta)=0$).

At the end of the required number of chemotaxis steps an assessment is made about the health of the bacteria by sorting them in the ascending order. Half of the healthy bacteria is replicated by assigning the same location and the other half is eliminated. This operation constitutes the reproduction step.

After the end of the desired number of reproduction steps, each bacterium may be eliminated or dispersed with some probability. This step is known as Elimination and dispersal and is meant shake up the bacteria so as to move them to better locations.

The initialization for the Bacterial Foraging algorithm is done as per the guidelines in [39]:

1. Set the number of bacteria B=50
2. The number of parameters w to be optimized are 2 : $T_U$, $T_V$
3. Swimming length Ns = 4
4. $N_c$, the number of iterations in a chemotactic loop is set to 100
5. $N_{re}$, the number of reproduction steps is set to 2.
6. $N_{ed}$, the number of elimination and dispersal events is set to 2.
7. The probability $p_{ed}$ that each bacterium will be eliminated/dispersed is set to 0.25
8. Location of each bacterium $L(w,B, N_c, N_{re}, N_{ed})$ .

*C. Results of color segmentation*

**Segmentation Evaluation Indices**

The quality of the segmentation is generally judged by two types of indices: the *goodness methods* such as Liu's F-measure which ascertains the color difference in the CIELAB color space and also penalizes the formation of large number of segments, and the *discrepancy methods* which ascertain the quality with respect to some reference result like ground truth images for example the RAND index. The above two types of quality measures are used together to judge the efficiency and practicality of the proposed algorithm.

1. *Liu's Evaluation Measure (F)*: The performance of color segmentation is evaluated using Liu and Yang's [41] evaluation function F:

$$F(I) = \frac{1}{1000(N_1 \times N_2)} \sqrt{G} \sum_{i=1}^{G} \frac{e_i^2}{\sqrt{A_i}} \tag{17}$$

where I is the segmented image and $N_1 \times N_2$ is the image size, G is the number of regions of the segmented image, $A_i$ is the area and $e_i$ is the average color error of the $i^{th}$ region where $e_i$ is defined as the sum of Euclidean distances between the {a*,b*} color vector of the pixels of region i and the color centroid attributed to region i in the segmented image. The smaller the value of F(I) the better the segmentation result. We choose Liu's F-factor as one of our evaluation criteria since it gives an accurate measure of the color differencing achieved by the segmentation algorithm and at the same time penalizes large number of regions formed.

2. a.)*Probabilistic RAND index (PR)*: The PR index is a generalization of the RAND index [42] introduced by Unnikrishnan *et al.* in [43]. It allows a comparison of the test segmentation with multiple ground truth images through soft non-uniform weighting of pixel pairs as a function of the variability in the ground truth sets.

Suppose each human k provides information about the segmentation in the form of binary numbers $\prod(l_i^{S_k} = l_j^{S_k})$ for each pair of pixels $(x_i,x_j)$. The set of all perceptually correct segmentations defines a Bernoulli distribution giving a random variable with the expected value denoted as $h_{ij}$. The Probabilistic RAND index (PR) is then defined as:

$$PR(S_{test}, \{S_k\}) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i<j}} [m_{ij}h_{ij} + (1-m_{ij})(1-h_{ij})] \tag{18}$$

Where $m_{ij}$ denotes the event of a pair of pixels i and j having the same label in the test image $S_{test}$:

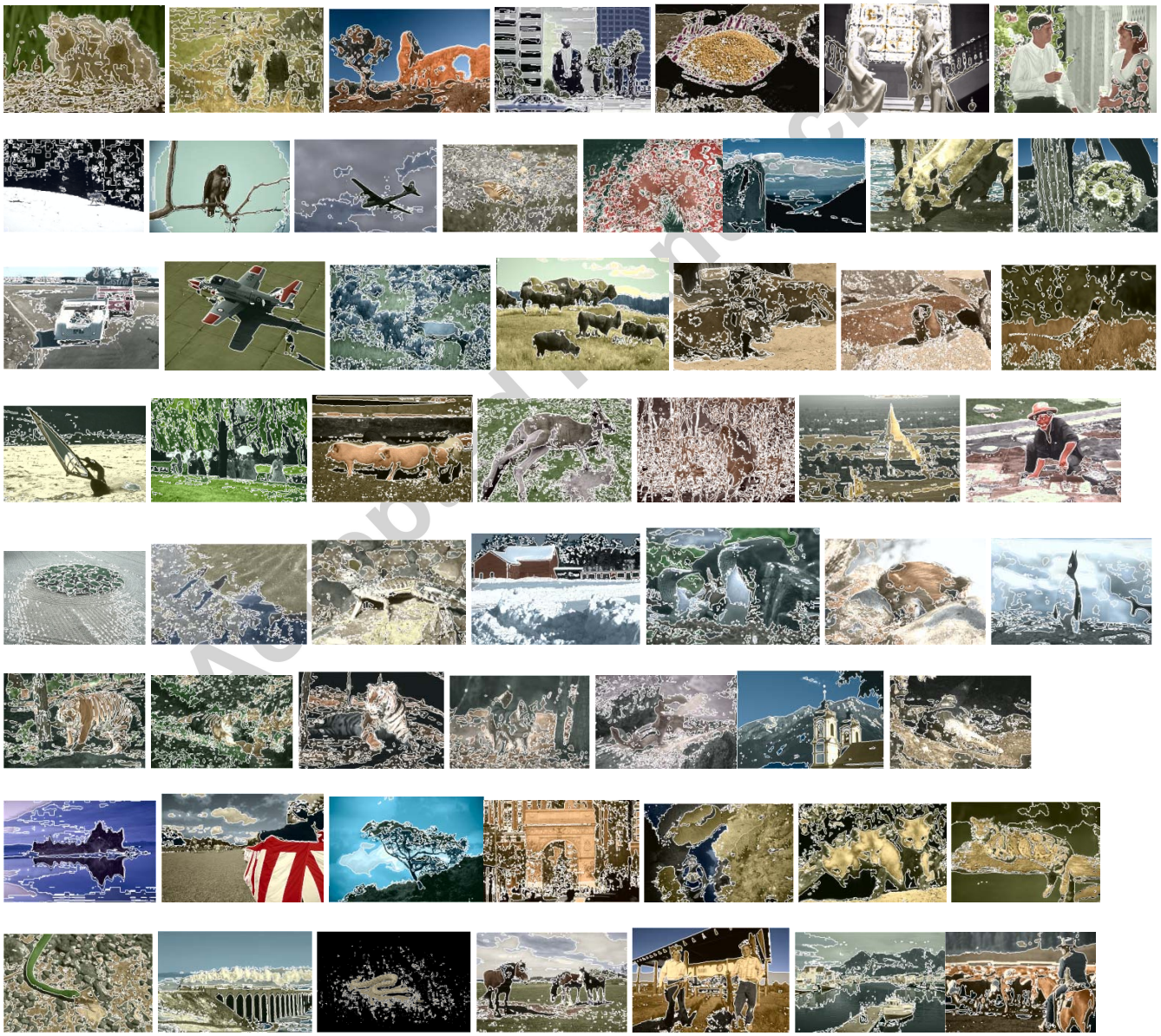
$$m_{ij} = \prod(l_i^{S_{test}} = l_j^{S_{test}}) \tag{19}$$

This measure takes values [0,1], where 0 means no similarities between $S_{test}$ and { $S_1$ ,$S_2$ ,….$S_K$ }, and 1 means all segmentations are identical.

b.) *Normalized Probabilistic RAND index (NPR):* The Normalized PR index by Unnikrishnan *et al.* in [12], is an excellent means of qualitative comparison among image segmentation algorithms. Once the segmentation of all the test images for all the algorithms being compared has been compiled the Normalized PR index is calculated so that a global measure is possible.

$$NormalizedPR = \frac{PR - ExpectedPR}{MaximumPR - ExpectedPR} \tag{20}$$

The above equation assures that the expected value of normalized index is zero providing a wider range. The maximum value of PR, *MaximumPR,* in (20) is taken as 1.



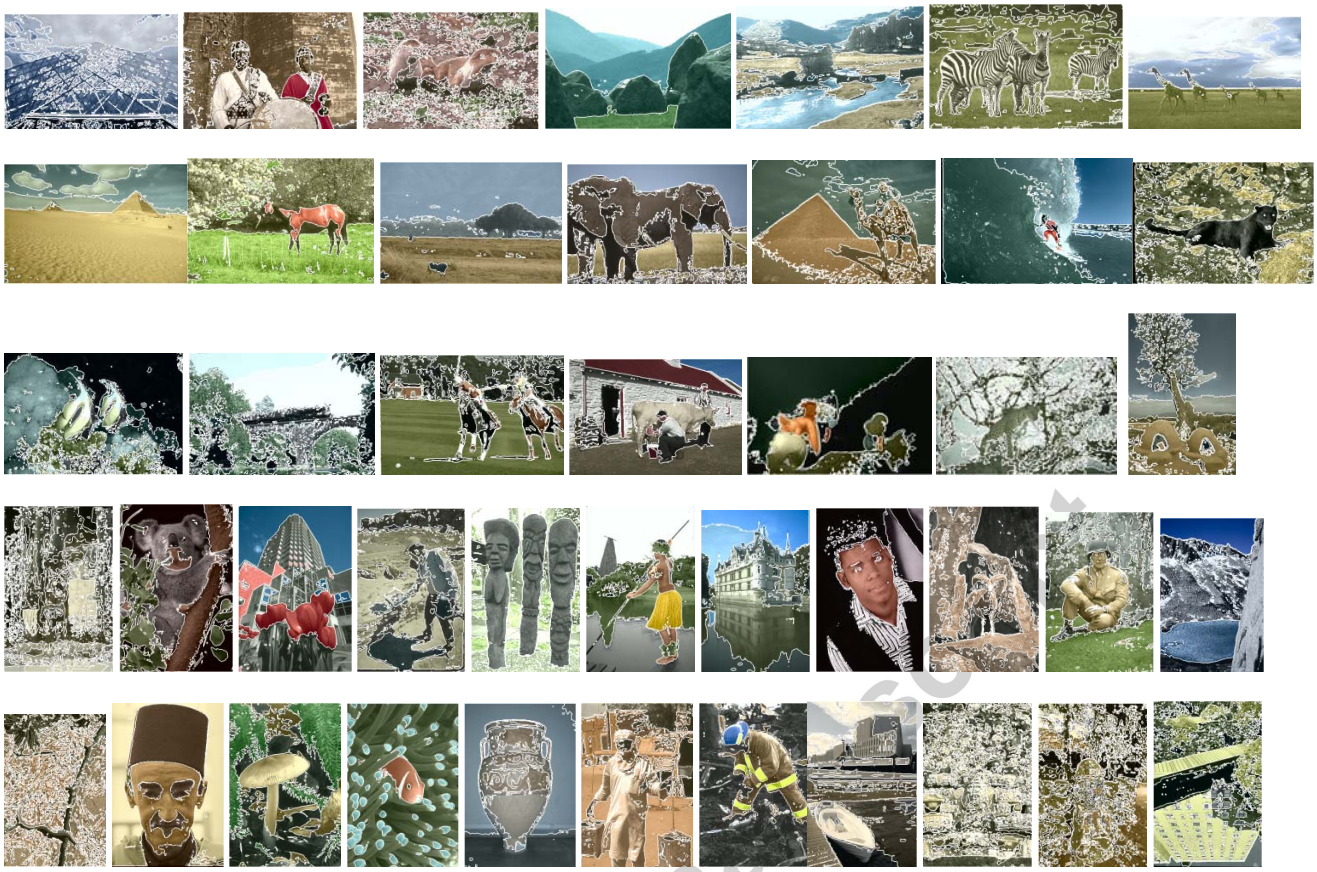
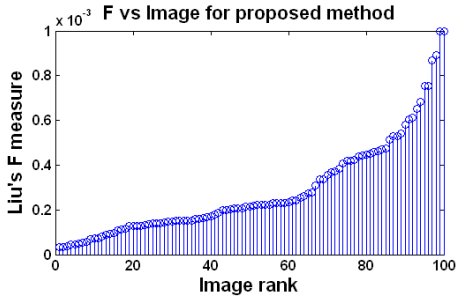


**Fig. 2:** *Color Segmentation Results of 100 test images from Berkeley segmentation database[44] by the proposed method*



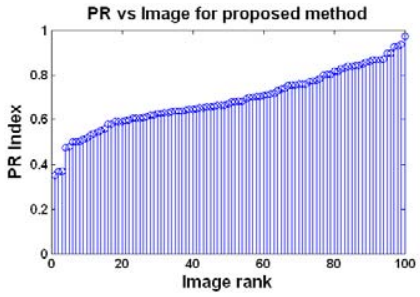




**Fig 3:** *Liu's F-measure and PR index in the increasing order for segmentations of 100 test images from Berkeley dataset[44] by the proposed method*

**Color segmentation results**

In this section experimental results are presented to prove the effectiveness of the proposed color segmentation algorithm on the natural images. For these results MATLAB (ver.7.9) software is run on a Pentium-IV 1.4GHz PC. All images are digitized to 24 bits per pixel in the RGB format. Since the distance between any two points in the RGB space is not proportional to their color

difference, transformation from the RGB space to a uniform color space: CIELAB [31] is performed. The vector {a*,b*} of CIELAB color space contains the total chrominance color information of pixels and is the feature space for our color segmentation experiments. The vector {L*} or luminance vector which decides the darkness or fairness of the image segments is discarded in the clustering process to ensure that the illumination effects do not affect the segmentation process. It is observed that the FCCI algorithm yields highly crisp values of object membership function $u_{ci}$ (close to 0 and 1). On the other hand the feature membership values $v_{cj}$ are highly fuzzy (close to 0.5) due to averaging over the entire dataset. The $v_{cj}$ values however are accentuated by the high values of parameter $T_V$ ($\approx 10^8$) in Eq.(4) creating a considerable influence in the computation of $u_{ci}$, eventually leading to crisp values of $u_{ci}$ after the iterative procedure. This helps in crisp classification during the defuzzification process.

A set of 100 test images is taken from the Berkeley segmentation database [44,45] along with 5-7 ground truth segmentations available for each image in the database for the evaluation of the results. The size of each image is either 321x481 or 481x321 and the average time taken by the FCCI algorithm for each image is approximately 35 seconds. The segmentation of all 100 images by the proposed FCCI algorithm is shown in Fig. 2, with the edges superimposed on original images. The corresponding graphs for Liu's F-factor and the Probabilistic RAND index (PR) is shown in Fig. 3 for 100 images bestowing excellent values for both segmentation evaluation measures. The results show a good match with human ground truth segmentations as indicated by a high value of Probabilistic RAND index (PR), and also efficient color differentiation as indicated by a low value of Liu's evaluation measure F. The number of clusters is determined from the first local minima in the cluster validity S graph (normally <7 clusters for our experiments) as demonstrated by the example shown in Fig. 4.

Some observations made from the results obtained are as follows:

1. *Tradeoff between color differencing and human perception:* In the case of images with distinct colors (Fig.5) there is an excellent correspondence with human perception (high NPR) but the color differentiation is not so good (high Liu's F-factor). In the case of images with indistinct colors (Fig.6), very good color differencing is observed (low Liu's F-factor) but the results appear to be over-segmented and hence NPR value is relatively less. The proposed technique therefore maintains a good tradeoff between the two segmentation evaluation measures.

2. *Sensitivity to parameters*: The algorithm is found to be more sensitive to the values of $T_U$ than $T_V$ since the values of $u_{ci}$ obtained are very crisp and a careful choice of $T_U$ in (4) is required for the algorithm to converge. Replacing the Bacterial Foraging optimization by the Genetic algorithm in our experiments results in a large computational overhead with an initial

population of 100 chromosomes required for acceptable results, while Bacterial Foraging starts giving good results for the initial population of 8 bacteria, thereby, significantly reducing the computational overhead.

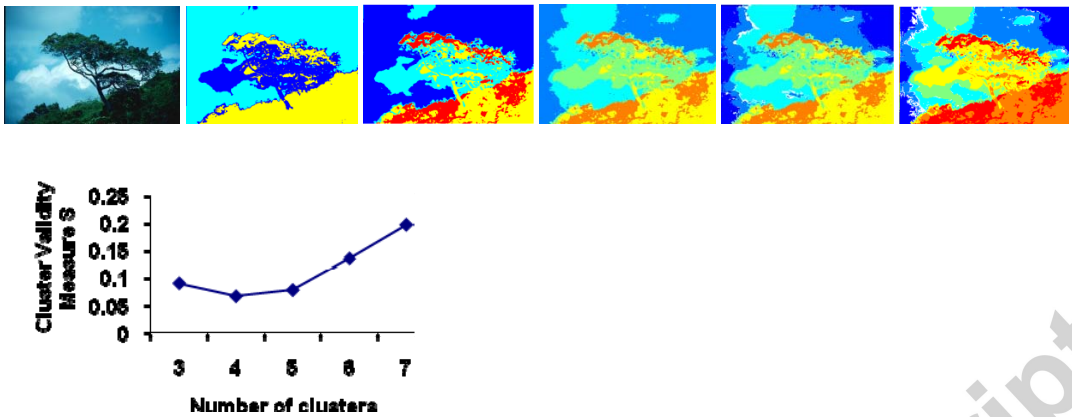



**Fig. 4:** *Original image and Segmentation results (from left to right and top to bottom) of "Tree" image and the corresponding Clustering validity graph as the number of clusters c is varied from 3 to 7. Corresponding values of NPR and Liu's F-factor are:*

*c: 3, 4, 5, 6, 7*

*NPR: 0.5519, 0.5203, 0.1314, -0.213, -0.1332*

*F: 0.00044, 0.000347, 0.000322, 0.000332, 0.000342 Number of clusters is aptly determined to be C=4 (from the first local minima in the graph) as it results in the most optimum combination of NPR and Liu's F-factor*

The value of $T_U$ in our experiments is found to range from 1 to 30 for different images with values close to 1 for images with non-distinct colors (Fig.6) and higher values for visually distinct colors (Fig.5) The valid values of $T_V$ ranges widely from $10^6$ to $10^8$ and do not have a major impact on the resulting clusters since its only function is to contribute to the computation of $u_{ci}$ by scaling $v_{cj}$.

**Fig.5**: *Example segmentations of images with distinct colors-*

*NPR(from left to right): 0.6065, 0.7962, 0.8007, 0.9213*

*F(from left to right) :0.00047, 0.000138, 0.000134, 0.000085*

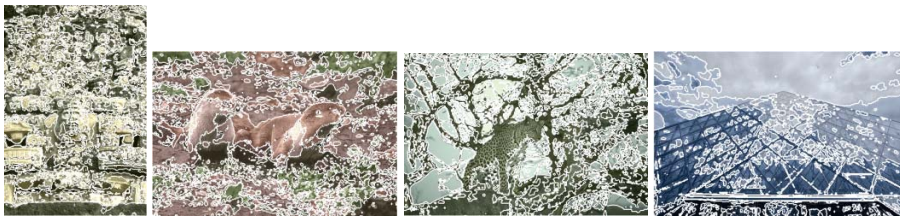


**Fig. 6:** *Example segmentations of images with indistinct/similar colors-*

*NPR(from left to right):-0.284, -0.391, -0.334, -0.098*

*F(from left to right): 0.000053, 0.00023, 0.000057, 0.0000701*

(a)

(b)

**Fig.7.** *Example segmentations with shadows and non-uniform illumination*

*NPR(from left to right):(a)-0.301, 0.314, -0.022 (b) -0.7609*

*F(from left to right):(a) 0.00025, 0.00023, 0.00026 (b) 0.00015*

**Fig.8:** *(a)Original image (b) Under-segmented result*

3. *Complex illumination Patterns*: The algorithm is able to segment natural scenes containing non-uniform illumination efficiently (Fig.7(a)) by segregating shadows from sunlit portions thus agreeing with human perception. However in the cases where the shadows tend to merge with the colors in the scene (Fig.7(b)) result tends to look over-segmented in spite of very good color differencing (low F).

4. *Under-segmentation:* Only in rare cases (1 out of 100) the algorithm fails to segregate extremely indistinct colors as demonstrated by the under-segmented result in Fig.8 due to formation of highly fuzzy clusters.

**Comparisons with other methods:**

The proposed color segmentation technique is compared with some well known methods in literature for unsupervised color segmentation: Fuzzy C-Means (FCM)[22,37], Normalized Graph-Cut (N-CUT) Method[10], Gaussian Mixture model (GMM)[8] and Mean-Shift (MS)[9] segmentation methods. The CIELAB color space is used for all the comparisons. While both mean shift and NCUT graph based method are popular feature space clustering methods, FCM is chosen since it is a widely popular fuzzy clustering method for image segmentation. The parameters for FCM algorithm used in our experiments are: index of fuzziness m=2, maximum error limit $\varepsilon=10^{-2}$, maximum iteration=200. Normalized Cuts graph based segmentation method uses eigenvector techniques to obtain graph partitions. It finds minimum cuts in a graph while minimizing the similarity between different patches. GMM models the color features as a mixture of Gaussian kernels using Expectation Maximization algorithm for estimation of parameters of the Gaussian mixture and is a popular method for image segmentation and retrieval. Figs.(9,10) show the graphs for F and NPR indices for the 100 test images from the Berkeley Segmentation Dataset in the form of histograms for the five methods. It is observed from the graphs that the proposed FCCI algorithm provides the most optimum combination of the lowest values of F-measure (of the order of $10^{-4}$) and sufficiently high values of NPR index among all the five methods proving the efficiency of the proposed color segmentation algorithm by striking a neat balance between color differencing and human perception standards.

Fig.11 shows the color segmented results of the five methods for six randomly selected test images from the Berkeley Segmentation Dataset namely: 'Mud-Huts', 'Plane', 'Eagle', 'Building', 'Wolf', 'Tree'. The corresponding F and NPR segmentation evaluation indices are listed in Table 1 depicting the lowest values of F for the proposed method as compared to all other methods. The NPR readings are also observed from Table 1, to be overall best for the proposed method though the Mean-Shift algorithm performs better for 'Plane' image. It is observed from the segmentation results in Fig.11(b) that the proposed method FCCI forms well defined and interpretable clusters even when the color difference between two regions is not too distinct as in the case of 'Wolf' image. The proposed technique is efficient in segmenting out uniform color regions and gives

less fake boundaries as observed in the case of 'Building' image in Fig.11(b) where the windows of the building are nicely segmented out visually as compared to other methods. An important factor that gives FCCI an edge over FCM clustering technique is the reduced time complexity. The time required by FCM for clustering is 10-15 minutes while the proposed algorithm hardly takes 1 minute for the same. How fuzzy co-clustering scores over fuzzy clustering can be observed by comparing the segmentations of the proposed method (Fig. 11(b)) with that of FCM (Fig.11(d)). The criterion for choosing number of clusters is the same as that for the proposed method (from the local minima in the clustering validity graph). As seen in the 'Tree' image co-clustering forms distinct and correct clusters with criteria being maximum separation between clusters/colors. Thus the green foliage and the blue sky being of distinct colors are grouped into separate clusters by the proposed method whereas in FCM, the parts of sky/clouds are clustered together with green foliage. FCM also suffers from the problem of outliers as observed from the pigmented segmentations in the 'Mud-Huts' example in Fig.11(d). The co-clustering results are improved because of the grading/relevance factor (feature membership function) assigned to each of the color feature (a*,b*) with respect to a particular cluster which leads to correct evaluation of clusters and also solves the problem of outliers. The N-CUT Graph based method does not give a very good correctness with respect to human ground truth images as compared to the proposed method and mean shift method. Also it tends to segregate uniform color regions into large chunks. GMM by Expectation Maximization algorithm is an unsupervised technique resulting in blob like segments. It classifies the entire tree and green foliage of the example 'Tree' in Fig. 11(e) into 1 segment performing well with respect to human evaluation but poorly for color differencing between regions. Moreover, the resulting blobs contain no internal details as demonstrated in the blobs formed in 'Eagle' image in Fig.11(e). The mean shift algorithm performs well corresponding to human ground truths as indicated by a high value of NPR (example-'Plane' image in Fig.11(f)). However it fails in the absence of any dominant colors in the scene as seen in the example of 'Wolf' image where the colors are visually indistinct. The Liu's F-measure values are found to be generally high in the case of both mean shift and N-CUT graph based algorithms. Also the mean shift method is very sensitive to its color bandwidth parameter $h_r$ with a slight change causing large changes in granularity of segmentation. The $h_r$ values for the test images are in the range 1 to 15 while the space bandwidth is $h_s$=15. Though some feature weighted clustering algorithms have been proposed in the past [46] and were applied for image segmentation problems [47], they do not incorporate a separate feature membership to be independently updated in each iteration along with pixel memberships and cluster centroids. Fig. 12 shows the comparison (F and NPR values) between the five methods for the mix of easy and difficult images in Fig. 5 and Fig. 6 respectively taken in the same order. The results affirm that F-values are minimum for FCCI for all images, a fact already established from the graphs in Fig.9, while NPR values are highest for FCCI for all the easy images E1 to E4 and most of the difficult images except for the D2 difficult image for which NCUT, GMM and Mean Shift yield better results. In Fig. 13 we compare the segmentations using some recent fuzzy clustering algorithms apart from FCM namely Spatially weighted FCM [48],

Histogram weighted FCM[49], Possibilistic FCM[50] and Orientation sensitive FCM (OSFCM) [51] on an example image 'Bird' containing three main clusters, the bird, the stones and the grass. The experimental parameters of these algorithms are summarized in Table 2. The results in Fig. 13 indicate that FCCI provides the most optimum segmentation with respect to human evaluation of the scene and is equivalent to FCM in terms of accuracy of the clusters formed. FCM and FCCI form the most visually acceptable clusters of the scene segmenting out the bird nicely in Cluster 2. Hence the reason that FCM among all the existing fuzzy clustering algorithms is most popular for image segmentation purpose. FCCI yields the cleaner image of the two (with lesser number of regions as compared to FCM as seen in Fig.13) complying more with human perception of the scene.

The inlier or bridge pixels are properly clustered by FCCI as compared to all other clustering algorithms as indicated by the region homogeneity tests (criteria being $u_{ci}$>0.5), though the outlier problem of FCM persists. The inlier pixels are defined as those that are equidistant from all centroids and have a membership of 0.5 to all clusters as a result. Since FCCI computes feature memberships as well which evaluates the relationship of a cluster to a feature, it is taken into account the feature value distribution of the inlier to compute its membership to a cluster. However the outlier points or noise pixels are a problem since for these data points the algorithm behaves like FCM and is sensitive though some marginal improvement is noticed from the Table in Fig. 13. Table 3 evaluates the memberships obtained from FCCI, FCM and PFCM algorithms for the $X_{11}$ dataset $\{X_{10} \cup (inlier)\}$ in [50] where $X_{10}$ contains 10 two dimensional data points. The results are found to be best for the proposed FCCI algorithm in terms of crisp values of memberships (1,0) obtained and the definitive values of inlier memberships that is indicative of the clusters ($u_{ci}$ >0.5) to which they belong to. The only shortcoming of FCCI clustering is the response to the outliers which need to be minimized.

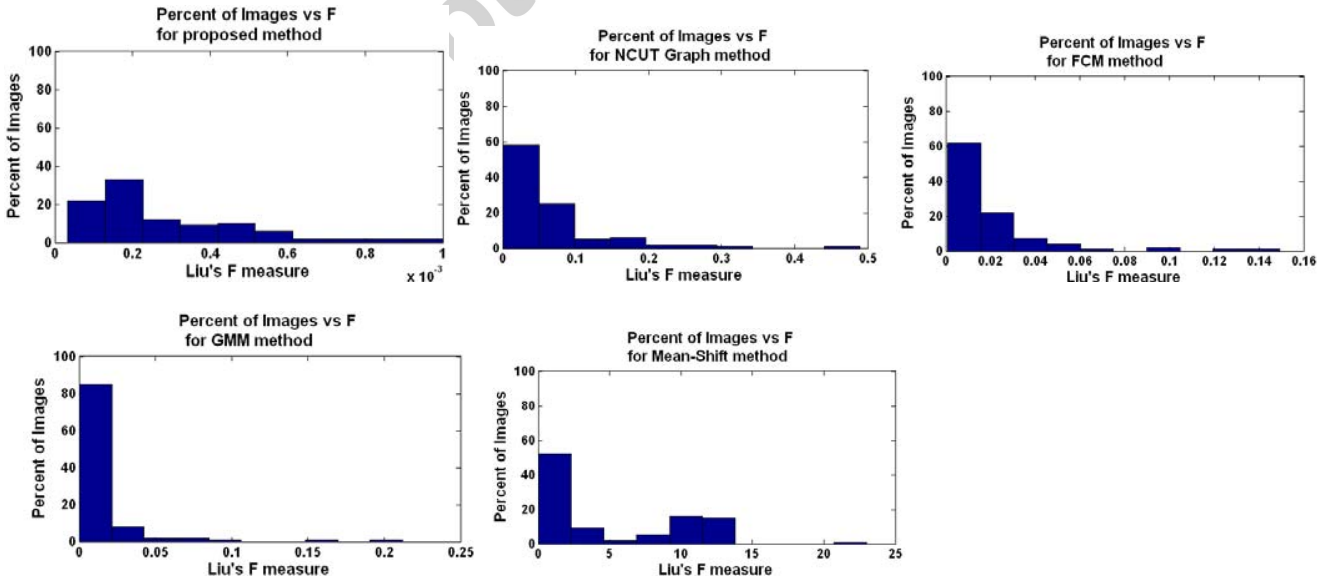



**Fig. 9:** *Histograms of Liu's F-measure achieved for individual images for the proposed method, NCUT Graph based method, FCM, GMM, Mean-shift segmentation methods.*

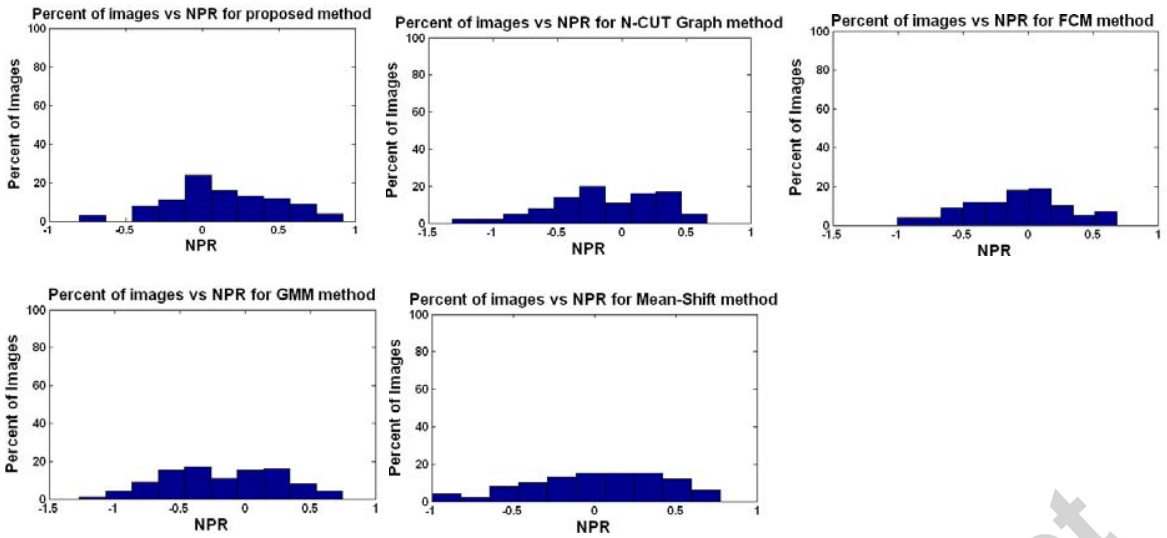
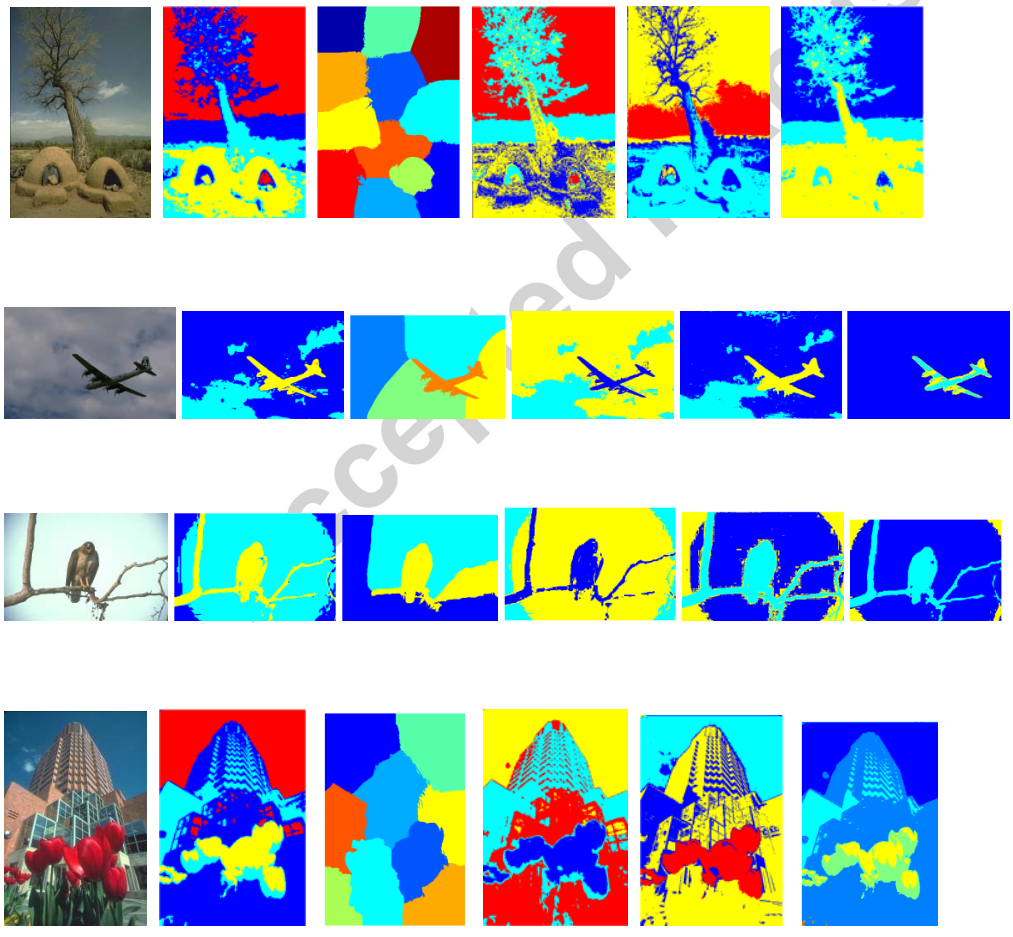
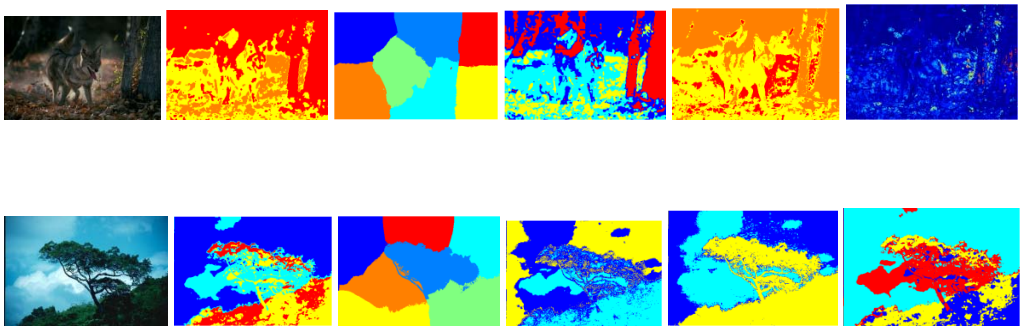



**Fig.10**: *Histograms of NPR Index achieved for individual images for the proposed method, NCUT Graph based method, FCM, GMM, Mean-shift segmentation methods.*



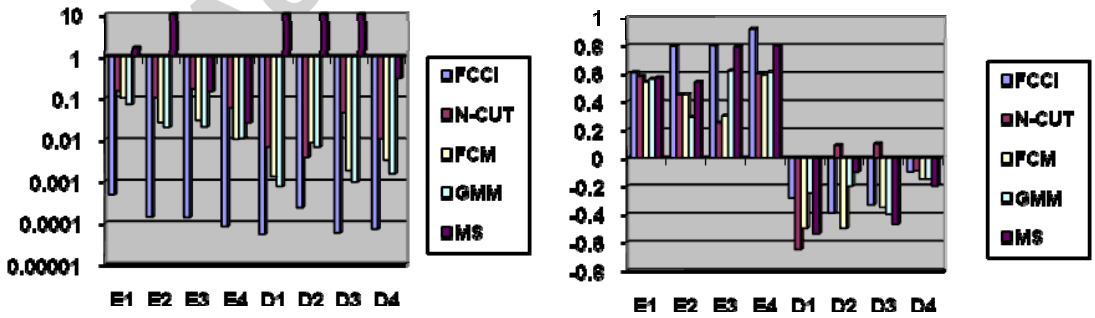


(a) (b) (c) (d) (e) (f)

**Fig. 11**.: *Color segmentation results (a) Original images from Berkeley segmentation database[57] : 'Mud-Huts', 'Plane', 'Eagle', 'Building', 'Wolf', 'Tree' (b)Proposed method [(from top to bottom): $T_U$={10.96,1.15,2.05,9.52,4.81,11.79},$T_V=10^8$] (c)N-CUT (d)FCM (e)GMM (f)Mean Shift segmentations[ Space Bandwidth $h_s$=15, Color bandwidth $h_r$ (from top to bottom):{7,3,4,10,1,8}].*

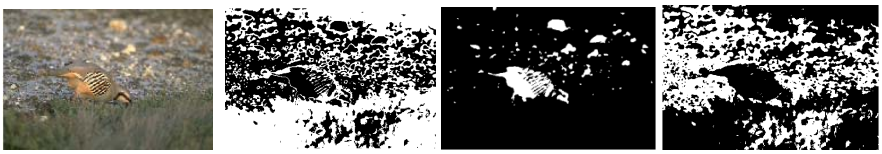
**Table 1** : Liu's F-measure and Normalized Probabilistic RAND index for the six test images: 'Mud-Huts', 'Plane', 'Eagle', 'Building', 'Wolf', 'Tree'

| | *MUD-HUTS* | | *PLANE* | | *EAGLE* | | *BUILDING* | | *WOLF* | | *TREE* | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *F* | *NPR* | *F* | *NPR* | *F* | *NPR* | *F* | *NPR* | *F* | *NPR* | *F* | *NPR* |
| Proposed Method | 0.000128 | 0.3324 | 0.000045 | 0.425 | 0.000142 | 0.4466 | 0.0010 | 0.3496 | 0.000147 | -0.0979 | 0.000447 | 0.5203 |
| N-CUT | 0.060 | 0.1904 | 0.0054 | -0.4135 | 0.0020 | 0.4419 | 0.2506 | 0.2653 | 0.0100 | -0.4611 | 0.0449 | 0.049 |
| FCM | 0.0154 | 0.2514 | 0.0028 | 0.2234 | 0.0061 | 0.4097 | 0.0909 | 0.2403 | 0.0054 | -0.6522 | 0.0086 | 0.09 |
| GMM | 0.0131 | 0.2237 | 0.0035 | 0.1815 | 0.0034 | 0.3476 | 0.0874 | 0.135 | 0.0034 | -0.6996 | 0.0084 | 0.3069 |
| Mean Shift | 0.0075 | 0.3146 | 0.0122 | 0.6562 | 0.3013 | 0.5353 | 1.2487 | 0.3346 | 23.5 | -0.4861 | 0.0225 | 0.3734 |

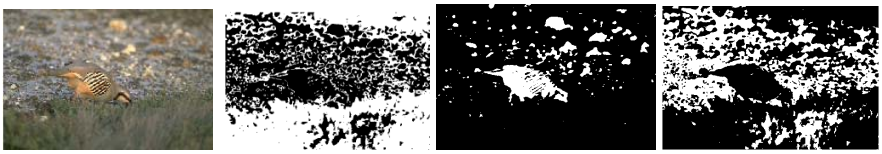



(a) (b)

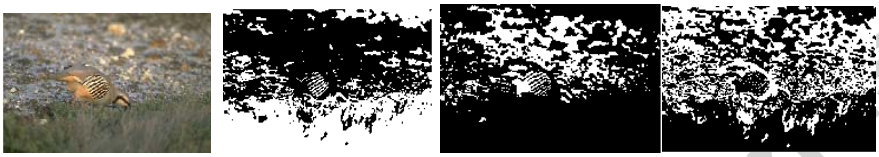
**Fig. 12** The plots for (a) F-value (logarithmic scale) and (b) NPR readings for 4 easy (E1 to E4) images (from Fig. 5) and 4 difficult (D1 to D4) images (from Fig.6) are shown below



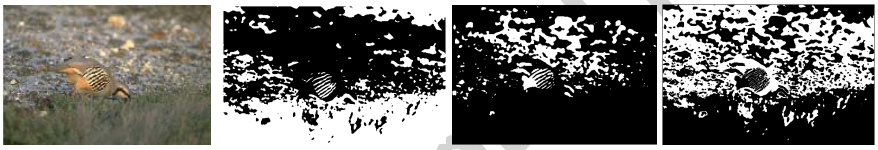
(a)



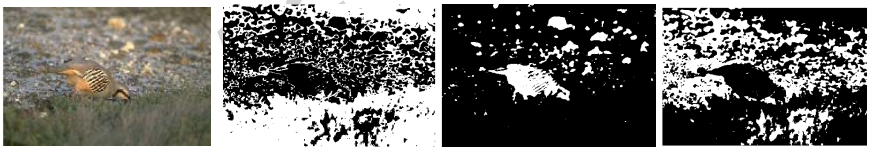
(b)



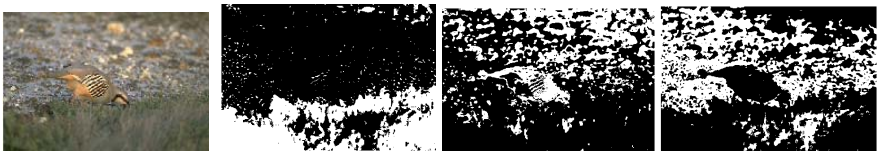
(c)



(d)



(e)



(f)

| *Clustering Evaluation Criteria* | FCCI | FCM | HWFCM | SWFCM | PFCM | OSFCM |
|---|---|---|---|---|---|---|
| Average MSE of clusters (Accuracy of clustering) | **34.7** | **34.6** | 187 | 207 | **34.6** | 52.7046 |
| Total Number of regions segmented (Semantic meaning of clusters) | **530** | 646 | 673 | 822 | 648 | 940 |
| Region Homogeineity test for u>0.5 (inlier test) | **100%** | 95% | 99.9% | 99.87% | 95% | 89.04% |
| Average MSE of clusters with added noise (outlier test) | 300 | 326 | **184** | 194 | 216 | 660 |

(g)

**Fig. 13** The comparison of fuzzy clustering results (with the best values in bold) on the 'Bird' image for three clusters (a) FCCI (b) FCM (c) Histogram Weighted FCM (d) Spatially Weighted FCM (e) PFCM (f) OSFCM (g) Segmentation Evaluation results in terms of Mean Square Error (MSE), Region Homogeneity Test (RH) and Mean Square Error (MSE) in the presence of noise.

Table 2 : Experimental Parameters of the fuzzy clustering algorithms being compared

| Algorithm | Parameters | Parameter Optimization Technique | Clustering Data (from CIEL*a*b* color space) | Computational Complexity * | Dissimilarity Measure * | Average Number of iterations needed for convergence (EL=0.01) | Average Execution Time |
|---|---|---|---|---|---|---|---|
| Fuzzy Co-clustering algorithm for images (FCC1) | $T_U$, $T_V$ | *Bacterial Foraging* | *a*,b** | *ObjectMembership u (CxN) FeatureMembership v (CxK) Centroid p (CxK)* | $v_{cj}D_{cij}$ *(CxNxK)* | *33* | *(with 8 bacteria) <1 minute* |
| Fuzzy C-means algorithm (FCM) | *m=2* | *none* | *a*,b** | *ObjectMembership u (CxN) Centroid p (CxK)* | $D_{ci}$ *(CxN)* | 27 | *10-15 minutes* |
| Histogram weighted FCM (HWFCM) | *m=2* | *none* | L* | *ObjectMembership u (GLxN) Centroid p (Cx1)* | $D_{ci}$ *(CxN)* | 61 | *10 seconds* |
| Spatially weighted FCM (SWFCM) | *m=2* | *none* | L* | *ObjectMembership u (CxN) Centroid p (Cx1)* | $D_{ci}$ *(CxN)* | 100 | *20 seconds* |
| Possibilistic FCM (PFCM) | *m=2,η=1.2,a=b=1* | *none* | *a*,b** | *ObjectMembership u (CxN) Typicality T (CxN) Centroid p (Cx1)* | $D_{ci}$ *(CxN)* | 33 | *20 minutes* |
| OSFCM | *m=2* | *none* | *a*,b** | *ObjectMembership u (CxN) Centroid p (Cx1)* | $D_{ci}$ *(CxN)* | 10 | *20 minutes* |

(*where number of clusters c=1 to C, number of data points i=1:N, number of features j=1:K, number of Gray levels gl=1 to GL with GL<<N, and D=||.||, the Euclidean distance norm)

**Table 3**: Clustering comparison of FCCI, FCM, PFCM- response to inliers and normal data points

| S.No. | Data | | *Fuzzy Co-clustering algorithm for images (FCC1) (Membership values) ($T_U$=0.96, $T_V$=9x10$^7$)* | *Fuzzy C-means algorithm (FCM) (Membership values) (m=2)* | *Possibilistic FCM (PFCM) (Typicality values) (m=2,η=2,a=b=1)* |
|---|---|---|---|---|---|
| | *x* | *y* | *U1, U2* | *U1, U2* | *T1, T2* |
| *1* | *-5* | *0* | ***1,0*** | *0.936, 0.06* | *0.621,0.113* |
| *2* | *-3.34* | *1.67* | ***1,0*** | *0.97 ,0.03* | *0.801,0.165* |
| *3* | *-3.34* | *0* | ***1,0*** | *0.99, 0.01* | *0.953,0.171* |
| *4* | *-3.34* | *-1.67* | ***1,0*** | *0.9 ,0.1* | *0.642,0.157* |
| *5* | *-1.67* | *0* | ***1,0*** | *0.92,0.08* | *0.840,0.278* |
| *6* | *1.67* | *0* | ***0,1*** | *0.08,0.92* | *0.278,0.840* |
| *7* | *3.34* | *1.67* | ***0,1*** | *0.03, 0.97* | *0.165,0.801* |
| *8* | *3.34* | *0* | ***0,1*** | *0.01, 0.99* | *0.171,0.953* |
| *9* | *3.34* | *-1.67* | ***0,1*** | *0.1 ,0.9* | *0.157,0.642* |
| *10* | *5* | *0* | ***0,1*** | *0.06 ,0.94* | *0.113,0.621* |
| *11 (Inlier)* | *0* | *0* | ***0.50031,0.4997*** | *0.5 ,0.5* | *0.49,0.49* |
| ***Cluster centers (Centroids)*** | | | ***-3.03 0*** <br> ***3.03 0*** | ***-3.36 0*** <br> ***3.36 0*** | ***-2.99 0*** <br> ***2.99 0*** |

## IV. CONCLUSIONS

In this paper, the Fuzzy Co-clustering approach based on the simultaneous clustering of both object and feature memberships is used for the color segmentation of natural images. A new objective function is formulated and the update rules are derived. The new algorithm (FCCI) is tried for color segmentation on 100 test images from the Berkeley segmentation dataset yielding precise segmentation. The color vectors {a*,b*} of CIELAB color space are the feature variables for the color segmentation algorithm. The performance is evaluated on the basis of Liu's Function F, and the Normalized Probabilistic RAND (NPR) index. The number of clusters is determined from the first local minima of Xie and Beni's cluster validity curve and is found to produce apt results (low F and high NPR). The proposed method produces accurate color differencing and at the same time adheres to the human perception in segmenting the natural scenes with non-uniform illumination and shading. It is also compared with some of the existing color segmentation techniques and is found to outperform them. The future scope of this work lies in improving the image segmentation in the presence of outliers and exploring other evolutionary algorithms for speedy solutions of the two parameters involved.

## REFERENCES


[1] Y.C.Ohta, T.Kanade, T.Sakai, "Color information for region segmentation", *Comp. Graphics image processing* , 133(1980) pp222-241.

[2]. Jayarammamurthy , S.N and T.L.Huntsberger," Edge and region analysis using fuzzy sets", *Procc. Of IEEE wrkshp. On Lang. aut.* June 1985 , pp.71-75.

[3]. P.Arbelaez et al., "From contours to regions: an empirical evaluation", *CVPR*,2009.

[4] Tremeau, A., and Borel, N, "A region growing and merging algorithm to colour segmentation", *Pattern Recognition*, 1997, 30, (7), pp. 1191–1203

[5]. S.C.Cheng, "Region growing approach to color segmentation using 3-D clustering and relaxation labeling", *IEEE Trans. On Vision & Image Signal proccessing,* Aug. 2003, pp. 270-276.

[6] C.C.Kang, W.J.Wang, "Fuzzy based seeded region growing for image segmentation", *NAFIPS* 2009..

[7]. .Mukherjee, "MRF clustering for segmentation of color images", *Pattern Recognition. Letters,* 23(2002), pp 917-929.

[8]. C.Carson, S.Belongie, H.Greenspan, J.Malik, "Blobworld: Image segmentation using Expectation maximization and its application to image querying", *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol.24, no.8, August 2002.

[9] D. Comaniciu, P.Meer, "Mean-Shift: A robust approach towards feature space analysis", *IEEE Trans. On Pattern Analysis and Machine Intell.*,Vol. 24,no.5, pp 603-619, May 2002.

[10] J.Shi and J.Malik, "Normalized Cuts and image segmentation", *IEEE Trans. On Patern Analysis and Machine Intell.*,Vol.22, no.8, pp 888-905, Aug. 2000.

[11] P.Felzenszwalb,.D.Huttenlocher, "Efficient graph based image segmentation", I*nternational Journal of Computer Vision*, Vol.59, no.2, pp. 167-181, Sept. 2004

[12]. R.Unnikrishnan et al., "Toward objective evaluation of image segmentation algorithms", *IEEE Trans. On Pattern Analysis and Machine Intell*., vol. 29, no.6, June 2007.

[13]. Uchiyama, Arbib,"Color image segmentation using competitive learning", *IEEE Trans. Pattern. Analysis and Mahcine Intell.,* Vol.16, no.12, pp. 1197-1206, 1994.

[14] Cheng,Jiang, Wang,"Colour image segmentation :advances and prospects", Patt. Recog. , 34, pp.1277-1294, 2001.

[15] Vincent, Soille, "Watersheds in digital spaces, an efficient algorithm based on immersion simulations", IEEE Trans. On Patt. Analysis and Machine Intell., vol.9, pp.735-744, 1991.

[16] Tilton, "D-Dimensional formulation and implementation of recursive hierarchical segmentation", NASA, Case no. GSC 15199-1.

[17] Y.Deng, and B.S.Manjunath, "Unsupervised segmentation of color texture regions in images and video", *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23(8), pp. 800-810, Aug. 2001.

[18] Wangenheim et al., "Color image segmentation using an enhanced Gradient network method", Patt. Recog. Letters 30(2009), pp. 1404-1412.

[19]. Q.Luo, Khoshgoftaar, "Unsupervised multiscale color image segmentation based on MDL principle", *IEEE Trans. On system, man, cyber.,* 15(9), pp.100-104, 2006.

[20] S.Ji, H.W.Park,"Image segmentationof color image based on region coherency", *Inter. Conf. on Image Procc*., pp. 80-83, 1998.

[21] Lior Shamir, "Human Perception based color segmentation using fuzzy logic", *Intl. Conf. on Fuzzy sys*.,2006, Vol. 2, pp. 496-505.

[22]. Lim, Y.W., and Lee, S.U, "On the colour image segmentation algorithm based on thresholding and the fuzzy c-means techniques', *Pattern Recognit.*, 1990, 23, (3), pp. 935–952.

[23] H.D.Cheng, J.Li, "Fuzzy homogeneity and scale space approach to color image segmentation, :*ICCV 2000*.

[24] C.H.Oh, K.Honda, H.Ichihashi, "Fuzzy clustering for categorical and multivariate data", *Proc. of IFSA/NAFIPS,* July 2001, Vol. 4, pp 2154-59.

[25] K.Kummamuru, A. Dhawale, R.Krishnapuram, "Fuzzy co-clustering of Documents and Keywords" ,*IEEE Conf. on Fuzzy Systems*, May 2003, Vol.2, pp.772-777.

[26] W.C.Tjhi, L.Chen, "Possibilistic fuzzy co-clustering of large document collections", *Patt. Recog*. 40(2007), pp. 3452-3466.

[27] W.C.Tjhi, L.Chen, "Robust fuzzy co-clustering algorithm", *IEEE Conf. ICICS* 2007.

[28] Guan, Qiu, Yang Xue, " Spectral images and features co-clustering with application to content based image retrieval", *Proc. Of IEEE Wrkshp. On Multimedia & Signal processing*, Oct. 2005,pp.1-4.

[29]. Qiu, "Image and feature co-clustering", *Proc. Of IEEE Conference on pattern Recog.*, 2004,pp.1051-1054.

[30] M. Hanmandlu Seba Susan, V.K.Madasu, B.C.Lovell, ,"Fuzzy Co-clustering of medical images using Bacterial Foraging", *IEEE Conf. on image vision and computing*, New Zealand, Nov 2008.

[31]. Wyszecki G., Stiles W.S., "*Color Science- concepts and methods, Quantitative data and formulae"*,Wiley Inter-Sc. Publns., New York,2000.

[32] G.H.Gomez et al., "Natural image segmentation using the CIELab space", Inter. Conf. on Elect., Comm. Comp., pp107-110, 2009.

[33] X.L.Xie and G.Beni, " A Validity Measure for Fuzzy Clustering*", IEEE Trans. On Patt. Ana. Machine Intell.*, vol.13, pp. 841-847, Aug. 1991.

[34] K.Rose, Gurewitz, Fox,"Statistical mechanics and phase transitions in clustering,"*Phys. Rev. Lett*., 65(8), 1990, pp.945-949.

[35]. E.T. Jaynes," Information theory and statistical mechanics", *Phys. Rev. Lett*, 106(1957), pp. 620-630.

[36] Miyamoto, Mikaidono,"Fuzzy c-means as a regularization and maximum entropy approach", 7th *IFSA World congress*,1997,pp.86-92.

[37] J.C.Bezdek, **Pattern Recognition With Fuzzy Objective Function Algorithms,** New York: Plenum Press 1981.

[38]. M.I.Chacon, L.Aguilar, A.Delgado, "Definition and applications of a Fuzzy image processing scheme", *IEEE Wkshp. On Signal procc*., Oct. 2002, pp.102-107.

[39]. K.M.Passino, "Biomimicry of Bacterial foraging", *IEEE Control Syst. Mag.,* June 2002, pp. 52-67.

[40]. M.Hanmandlu, O.P.Verma, N.Krishna Kumar, M.Kulkarni, " A Novel Optimal Fuzzy System for color image enhancement using bacterial foraging", *IEEE Trans. On Measurements and Instrumentation,* vol. 58, no.8, pp. 2867-2879.

[41]. J.Liu, Yang,"Multi resolution color image segmentation*", IEEE Trans. On Patt. Analysis and Machine Intell.* Vol. 16, no.7, pp. 530-549, 1994.

[42]. W.M.Rand, "Objective criteria for the evaluation of clustering methods", *J. Am. Statistical Assoc*., vol. 66, no.336, pp.846-850, 1971.

[43]. R.Unnikrishnan, M.Herbert, "Measures of similarity", *Procc. IEEE Workshop Computer vision apll*., 2005.

[44]. The Berkeley segmentation dataset and benchmark. [Online] **http://www.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/**

[45]. D.Martin, C.Fowlkes, D.Tal, J,Malik, "A Database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", Proc. On intl. conf. on Comp. vision, July 2001, vol. 2, pp. 416-423.

[46] Chenglong Tang, Shigang Wang, Wei Xu. "New fuzzy c-means clustering model based on the data weighted approach." *Data& Knowledge Engineering* ,2010,69(9): 881-900

[47] Xiao, Ho, Bargeila, "Automatic brain MRI segmentation scheme based on feature weighting factors selection on fuzzy c-means clustering algorithms with Gaussian smoothing", *International Journal of Computational Intelligence in Bioinformatics and Systems Biology*, Volume 1 Issue 3, February 2010

[48] KS Chuang , H.L.Tzeng, S. Chen, J.Wu, T-J. Chen, "Fuzzy c-means clustering with spatial information for image segmentation ", *Computerized Medical Imaging and Graphics* 30 (2006) 915.

[49] Shilong Wang, Yuru Xu and Yongjie Pang, "A fast underwater optical image segmentation algorithm based on a histogram weighted fuzzy c-means improved by PSO" *Journal of Marine Science and Application,* Volume 10, Number 1 (2011), 70-75.

[50]N.pal, K.Pal, J.Bezdek, "A Possibilistic fuzzy c-means clustering algorithm", *IEEE Trans. On Fuzzy Systems*, 13(4), 517-530,2005.

[51] Schmid P. "Segmentation of Digitized Dermatoscopic Images by Two-Dimensional Color Clustering." *IEEE Trans. on Medical Imaging*, 18(2): 164-171. 1999

APPENDIX I

The proof of convergence of the FCCI algorithm is shown below:

**Theorem 1**: The updated values of $u_{ci}$ given by Eq.(9) never increase the objective function in every iteration. .

*Proof*: Consider the objective function as a function of $u_{ci}$ alone.

$$J_{FCCI}(U) = \sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ci} v_{cj} D_{cij} + T_U \sum_{c=1}^{C}\sum_{i=1}^{N} u_{ci} \log u_{ci} + constant \tag{31}$$

where, $constant = T_V \sum_{c=1}^{C}\sum_{j=1}^{K} v_{cj} \log v_{cj}$

Also, the product $v_{cj}D_{cij}$ may be considered as constant. To prove theorem 1 we have to prove that U*, i.e the updated values of $u_{ci}$ given by Eq.(9) are the local minima of the objective function $J_{FCCI}(U^*)$ provided that the constraints in (5) and(6) are satisfied. For this we need to prove that the Hessian matrix $\Delta^2 J_{FCCI}(U^*)$ is positive definite.

$$\Delta^2 J_{FCCI}(U) = \begin{bmatrix} \frac{\partial^2 J_{FCCI}(U)}{\partial u_{11}\partial u_{11}} & \cdots & \frac{\partial^2 J_{FCCI}(U)}{\partial u_{11}\partial u_{CN}} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 J_{FCCI}(U)}{\partial u_{CN}\partial u_{11}} & \cdots & \frac{\partial^2 J_{FCCI}(U)}{\partial u_{CN}\partial u_{CN}} \end{bmatrix} = \begin{bmatrix} \frac{T_U}{u_{11}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{T_U}{u_{CN}} \end{bmatrix} \tag{32}$$

At U*, $u_{ci} \geq 0$ and $T_U$ is always assigned a positive value. Therefore the Hessian matrix $\Delta^2 J_{FCCI}(U^*)$ is positive definite. We have proved the first necessary condition $\frac{\partial J_{FCCI}(u_{ci})}{\partial u_{ci}} = 0$ and the second sufficient condition: $\Delta^2 J_{FCCI}(U^*)$ is positive definite. Therefore $u_{ci}^*$ updated is indeed a local minima of $J_{FCCI}(U)$ and it never increases the objective function value.

**Theorem 2**: For every iteration the updated values of $v_{cj}$ given by Eq.(11) never increases the objective function.

*Proof*: Proof is similar to proof of Theorem 1.

**Theorem 3**: The following constraint is satisfied by $J_{FCCI}$ in (4):

$$J_{FCCI} \geq T_U \times N \times \log \frac{1}{C} + T_V \times C \times \log \frac{1}{K}$$

*Proof:* Since the minimum value of $u_{ci}$ and $v_{cj}$ is 0, and $D_{cij} \geq 0$, the first term of $J_{FCCI}$ reduces to :

$$\sum_{c=1}^{C}\sum_{i=1}^{N}\sum_{j=1}^{K} u_{ci} v_{cj} D_{cij} \geq 0 \tag{33}$$

The second and third terms denote the entropy values and maximum value of entropy occurs when $u_{ci}=1/C$ and $v_{cj}=1/K$. In view of these the point of minima is:

$$J_{FCCI} \geq T_U \times N \times \log \frac{1}{C} + T_V \times C \times \log \frac{1}{K} \tag{34}$$

**Corollary**: Theorems 1-2 prove that the updated equations of FCCI point to a local minima of the Objective function and Theorem 3 indicates the lower limit of $J_{FCCI}$.

APPENDIX II

The CIELAB color space is obtained from the RGB color space by the following transformation:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.000 & 0.010 & 0.990 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{35}$$

The features of CIELAB are derived from :

$$L^* = \begin{cases} 116(Y')^{\frac{1}{3}} - 16 & if Y' > 0.008856 \\ 903.3Y' & otherwise \end{cases} \tag{36}$$

$$a^* = 500\left[ K_1^{\frac{1}{3}} - K_2^{\frac{1}{3}} \right], \tag{37}$$

$$b^* = 200\left[ K_2^{\frac{1}{3}} - K_3^{\frac{1}{3}} \right] \tag{38}$$

where,

$$K_i = \begin{cases} \phi_i & if \phi_i > 0.008856 \\ 7.787\phi_i + \frac{16}{116} & otherwise \end{cases} \qquad \text{for i=1,2,3} \tag{39}$$

$$\phi_1 = X' = \frac{X}{X_0}, \quad \phi_2 = Y' = \frac{Y}{Y_0}, \quad \phi_3 = Z' = \frac{Z}{Z_0} \tag{40}$$

The $X_0$ , $Y_0$ and $Z_0$ are the values of X,Y,Z for the reference white, respectively. The reference white is defined as { R=G=B=255}.

**Author Biography**

Madasu Hanmandlu

**Madasu Hanmandlu** (M'02) received the B.E. degree in Electrical Engineering from Osmania University, Hyderabad, India, in 1973, the M.Tech. degree in power systems from R.E.C. Warangal, Jawaharlal Nehru Technological University (JNTU), India, in 1976, and the Ph.D. degree in control systems from Indian Institute of Technology, Delhi, India, in 1981. From 1980 to 1982, he was a Senior Scientific Officer in Applied Systems Research Program (ASRP) of the Department of Electrical Engineering, IIT Delhi. He joined the EE department as a lecturer in 1982 and became Assistant Professor in 1990, an Associate Professor in 1995 and finally a Professor in 1997. He was with Machine Vision Group, City University, London, from April –November, 1988, and Robotics Research Group, Oxford University, Oxford from March-June, 1993, as part of the Indo-UK research collaboration. He was a Visiting Professor with the Faculty of Engineering (FOE), Multimedia University, Malaysia from March 2001 to March 2003. He worked in the areas of Power Systems, Control, Robotics and Computer Vision, before shifting to fuzzy theory. His current research interests mainly include Fuzzy Modeling for Dynamic Systems and applications of Fuzzy logic to Image Processing, Document Processing, Medical Imaging, Multimodal Biometrics, Surveillance and Intelligent Control. He has authored a book on Computer Graphics in 2005 under PBP publications and also has well over 220 publications in both conferences and journals to his credit. He has guided 20 Ph.Ds and 120 M.Tech students. He has handled several sponsored projects. He was an Associate Editor of both Pattern Recognition Journal (Jan. 2005-Mar. 2011) and of IEEE Transactions on Fuzzy Systems and a reviewer to other journals such as Pattern Recognition Letters, IEEE Transactions on Image Processing and Systems, Man and Cybernetics( Jan. 2007-Jan.2010). He is a senior member of IEEE and is listed in Reference Asia; Asia's who's who of Men and Women of achievement; 5000 Personalities of the World (1998), American Biographical Institute. He was a Guest Editor of Defense Science Journal for the special issue on" Information sets and Information Processing" September, 2011

O.P.Verma

Om Prakash Verma received his B.E. degree in Electronics and communication Engineering from Malaviya National Institute of Technology, Jaipur, India, M. Tech. degree in Communication and Radar Engineering from Indian Institute of Technology (IIT), Delhi, India and PhD in the area of applications of soft and evolutionary in image processing from University of Delhi, Delhi, India. From 1992 to 1998 he was assistant professor in Department of ECE at Malaviya National Institute of Technology, Jaipur, India. He joined Department of Electronics & Communication Engineering, Delhi Technological University (formerly Delhi College of Engineering) as Associate Professor in 1998. Currently, he is Professor and Head, Department of Information Technology Delhi Technological University, Delhi India. He is also the author of more than 30 publications in both international conference proceeding and journal. He has guided more than 30 M. Tech student for their thesis and presently 5 PhD research scholars are working under his supervision. He has authored a book on Digital Signal Processing in 2003. His present research interest includes: Applied soft computing, Artificial intelligent, Evolutionary computing, Image Processing, Digital signal processing. He is also a Principal investigator of an Information Security Education Awareness project, sponsored by Department of Information Technology, Government of India.

Seba Susan
She is currently a student of doctoral studies in Electrical Engineering Department in the Indian Institute of Technology, Delhi. Her areas of interest include Image processing, Pattern Recognition, Fuzzy-Neural Networks.

V.K.Madasu
Vamsi Krishna Madasu obtained Bachelor of Technology degree in Electronics & Communication Engineering with distinction from Jawaharlal Nehru Technological University, India in 2002 and PhD in Electrical Engineering from the University of Queensland, Australia in 2006. From 2006-2008, he was a Research Associate in the School of Engineering Systems at Queensland University of Technology where he developed innovative Image Processing and Fuzzy Logic based technologies for diverse industrial applications. Currently, he is a Senior Research Officer at Tetra Q, University of Queensland, Australia working in the field of Medical Image Analysis. Vamsi is a member of IEEE, Computer Society and is listed in Who's Who in the World.

Madasu Hanmandlu

Om Prakash Verma

Seba Susan

V.K.Madasu

# Condition Based Maintenance Modeling for Availability Analysis of a Repairable Mechanical System

**RachnaChawla**

*Department of Mechanical and Automation Engineering*

*Maharaja Agrasen Institute of Technology, Delhi, India*

rachnapareva@gmail.com

**Girish Kumar**

*Department of Mechanical and Production Engineering*

*Delhi Technological University, Delhi, India*

girish.kumar154@gmail.com

**Abstract- This paper deals with the condition based maintenance modeling for availability analysis of repairable mechanical systems using MARKOV analysis. Maintenance actions are selected out of four actions namely no repair, minor maintenance, imperfect maintenance and major maintenance. The various probabilities of selecting the maintenance depend upon the level of degradation. Thus system MARKOV model is developed incorporating these aspects, i.e. multi-state degradation, periodic inspection, condition based maintenance actions and random failures. The solution of the model is obtained analytically by solving system of ordinary differential equations by Ranga-Kutta method using MATLAB software. The proposed methodology is implemented for centrifugal pump. The suggested approach helps in gauging and assessing availability and hence is useful for the engineers in enhancing the overall availability of the system. It is also helpful for maintenance engineers in deciding suitable maintenance and replacement policies. We are optimizing the condition monitoring interval to maximize the system availability.**

**Keywords – Markov, CBM, Ranga-Kutta**

## I . INTRODUCTION

Reliability has always been an important aspect in the assessment of industrial products and/or equipment. Good product design is of course essential for products with high reliability. However, no matter how good the product design is, products deteriorate over time since they are operating under certain stress or load in the real environment, often involving randomness. Maintenance has, thus, been introduced as an efficient way to assure a satisfactory level of reliability during the useful life of a physical asset.

The earliest maintenance technique is basically breakdown maintenance (also called unplanned maintenance, or run-to-failure maintenance), which takes place only at breakdowns. A later maintenance technique is time-based preventive maintenance (also called planned maintenance), which sets a periodic interval to perform preventive maintenance regardless of the health status of a physical asset. With the rapid development of modern technology, products have become more and more complex while better quality and higher reliability are required. This makes the cost of preventive maintenance higher and higher. Eventually, preventive maintenance

has become a major expense of many industrial companies. Therefore, more efficient maintenance approaches such as condition-based maintenance (CBM) are being implemented to handle the situation.

CBM is a maintenance program that recommends maintenance actions based on the information collected through condition monitoring. CBM attempts to avoid unnecessary maintenance tasks by taking maintenance actions only when there is evidence of abnormal behaviours of a physical asset. A CBM program, if properly established and effectively implemented, can significantly reduce maintenance cost by reducing the number of unnecessary scheduled preventive maintenance operations.

A CBM program consists of three key steps (see Fig. 1):

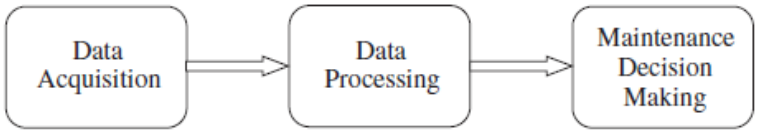



Fig.1. Three steps in a CBM program.

1. Data acquisition step (information collecting), to obtain data relevant to system health.
2. Data processing step (information handling), to handle and analyse the data or signals collected in step 1 for better understanding and interpretation of the data.
3. Maintenance decision-making step (decision-making), to recommend efficient maintenance policies.

The remaining paper is organised in the following manner:

Section II deals with system modelling. In section III the solution of the system model is obtained. In section IV results are discussed sensitive analysis is carried out to optimise the condition monitoring interval. Finally section V concludes the paper.

## II. SYSTEM MODELLING

### *A. Degradation*

Whenever a system or a model is in working it degrades with time. The degradation is gradual not sudden. We are trying to study a mode that follows this kind of failure.

In degradation modeling we study a system that is prone to degradation and mostly we study the systems where reliability is critical. As shown in the figure is such a system.

There are four stages shown. Fresh component is given the stage D1, then with time it degrades to a stage D2 and so on and finally it goes to a failure state. We will be studying the degradation rate from one stage to the other for all the stages.

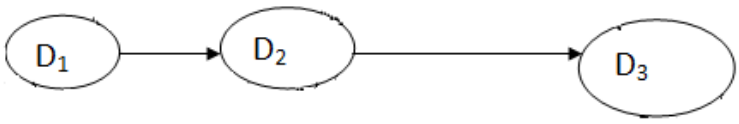



Fig.2. Multi State degradation

*B. Inspection*

Inspection is a way to see the health of the system and deciding whether the system requires repair/maintenance or not. Now there are two types of inspections:

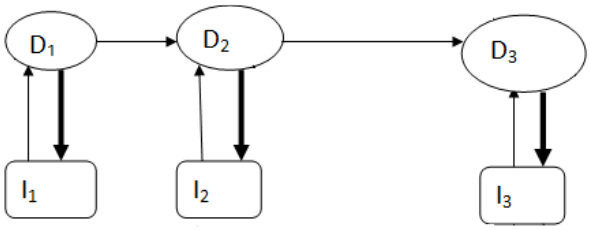



Fig 3. Periodic Inspection at every stage

Online: In this we need not to stop the system for inspection so the availability of the system is more, and

Offline: In this we need to stop the system for inspection.

As described in the above figure we take the system further and do periodic inspections at each state defined. These inspections help us in maintain the system by doing timely repairs and maintenance.

*C. Condition Based Maintenance*

Condition based maintenance (CBM), shortly described, is maintenance when need arises. This maintenance is performed after one or more indicator shows the equipment is going to fail or that equipment performance is deteriorating.

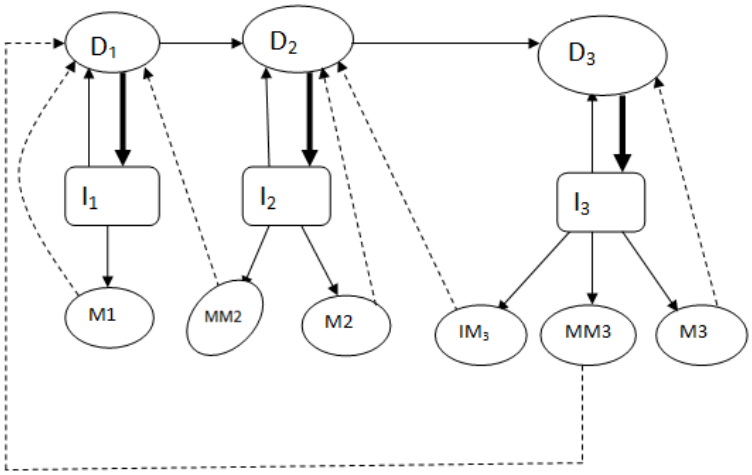



Fig 4.Condition based maintenance

In our system model, we have described three types of maintenance on the basis of the requirement of the system. The three types of maintenance are: Minor maintenance, intermediate maintenance and major maintenance.

In stage D1, our system is new, thus , we need not require much maintenance for it. Therefore we have kept the probability for our system to undergo minor maintenance to be 0.1 and the probability that system would go back to the stage D1 without any maintenance to be 0.9.

Similarly in stage D2 as our system is in continuous working state, it deteriorates and thus its efficiency decreases and the need to repair it or maintain it increases as compared to the system in stage D1. Due to this reason we have decreased the probability that the system would go back to stage D2 without any repair from 0.9 to 0.7 and the probability that the system would require maintenance has been increased from 0.1 to 0.3.

Finally, when our system moves from stage D2 to D3, it deteriorates further giving rise to the need to repair it in order to increase its availability. Therefore the probability is that he system requires minor repair or intermediate repair or major repair or no repair has been altered again the probability that the system would require major maintenance has been changed to 0.2.The probability that the system would require intermediate repair has been changed to 0.4. The probability that the system would require minor maintenance has been change to 0.2 and finally the probability that the system would go back to stage D3 without any repair has been changed to 0.2.

*D. Random Failure*

Random failure is defined as the situation Condition in which the system fails due to some random causes. These random causes can be anything from natural calamity to human error. Random failures can also occur due to voltage fluctuations, manufacturing defects, problem in system components, etc.

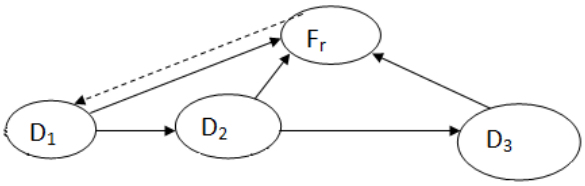



Fig 5. Random failure

Random failure causes the system to go in offline mode thereby bringing its availability to 0.

## III. Solution of System Model

After developing the model as above, we will now obtain the solution using analytical approach(Markov Analysis). The set of ordinary differential equation are provided below:

1. $dP_{d1}/dt = -\lambda_{d1d2}P_{d1}(t) - \lambda_{d1i1}P_{d1}(t) - \lambda d_{1fr}P_{d1}(t) + \mu_{i1d1}P_{i1}(t) + \mu_{m1d1}P_{m1}(t) + \mu_{mm2d1}P_{mm2}(t) + \mu_{mm3}P_{mm3}(t) + \mu_{frd1}P_{fr}(t)$
2. $dP_{d2}/dt = \lambda_{d1d1}P_{d1}(t) - \lambda_{d2d3}P_{d2}(t) - \lambda_{d2fr}P_{d2}(t) - \lambda_{d2i2}P_{d2}(t) + \mu_{i2d2}P_{i2}(t) + \mu_{m2d2}P_{m2}(t) + \mu_{im3d2}P_{im3}(t)$
3. $dP_{d3}/dt = \lambda_{d2d3}P_{d2}(t) - \lambda_{d3i3}P_{d3}(t) - \lambda_{d3fr}P_{d3}(t) + \mu_{m3d3}P_{m3}(t) + \mu_{i3d3}P_{i3}(t)$
4. $dP_{i1}/dt = \lambda_{d1i1}P_{d1}(t) - \lambda_{i1m1}P_{i1}(t) - \mu_{i1d1}P_{i1}(t).$
5. $dP_{i2}/dt = \lambda_{2i2}P_{d2}(t) - \lambda_{i2m2}P_{i2}(t) - \lambda_{i2mm2}P_{i2}(t) - \mu_{i2d2}P_{i2}(t)$
6. $dP_{i3}/dt = \lambda_{3i3}P_{d3}(t) - \lambda_{i3m3}P_{i3}(t) - \lambda_{i3mm3}P_{i3}(t) - \lambda_{i3m3}P_{i3}(t) - \mu_{i3d3}P_{i3}(t)$
7. $dP_{Fr}/dt = \lambda_{d1fr}P_{d1}(t) + \lambda_{dfr}P_{d2}(t) + \lambda_{d3fr}P_{d3}(t) - \mu_{frd1}P_{fr}(t)$
8. $dP_{m1}/dt = \lambda_{i1m1}P_{i1}(t) - \mu_{m1d1}P_{m1}(t)$
9. $dP_{m2}/dt = \lambda_{i2m2}P_{i2}(t) - \mu_{m2d2}P_{m2}(t)$
10. $dP_{mm2}/dt = \lambda_{i2mm2}P_{i2}(t) - \mu_{mm2d1}P_{mm2}(t)$
11. $dP_{m3}/dt = \lambda_{i3m3}P_{i3}(t) - \mu_{m3d3}P_{m3}(t)$
12. $dPim3/dt = \lambda_{i3m3}P_{i3}(t) - \mu_{im3}P_{i3}(t)$
13. $dP_{mm3}/dt = \lambda_{i3mm3}P_{i3}(t) - \mu_{mm3d1}P_{mm3}(t)$

## IV. RESULT AND SENSITIVITY ANALYSIS

In this section results are given in tabular form and sensitivity analysis is carried out.

Table -1: Distribution parameters for failure/Repair/Inspection Interval Transition

| S.no. | Transition | PARAMETER | VALUE |
|---|---|---|---|
| 1 | $D_1D_2$ | $\lambda_{D1D2}$ | 0.00025 |
| 2 | D2D3 | $\lambda_{D2D3}$ | 0.00067 |
| 3 | D1I1 | $\lambda_{D1I1}$ | 0.004 |
| 4 | I1M1 | $\Lambda_{I1M1}$ | 0.5 |
| 5 | D2I2 | $\lambda_{D2I2}$ | 0.00595 |
| 6 | I2M2 | $\lambda_{I2M2}$ | 0.25 |
| 7 | I2MM2 | $\lambda_{I2MM2}$ | 0.25 |
| 8 | D3I3 | $\lambda_{D3I3}$ | 0.01 |
| 9 | I3M3 | $\lambda_{I3M3}$ | 0.125 |
| 10 | I3MM3 | $\lambda_{I3MM3}$ | 0.125 |
| 11 | I3IM3 | $\lambda_{I3IM3}$ | 0.125 |
| 12 | D1Fr | $\lambda_{D1Fr}$ | 0.00002 |
| 13 | D2Fr | $\lambda_{D2Fr}$ | 0.00002 |
| 14 | D3Fr | $\lambda_{D3Fr}$ | 0.00002 |
| 15 | I1D1 | $\mu_{I1D1}$ | 0.005 |
| 16 | I2D2 | $\mu_{I2D2}$ | 0.025 |
| 17 | I3D3 | $\mu_{I3D3}$ | 0.0125 |
| 18 | M1D1 | $\mu_{M1D1}$ | 0.05 |
| 19 | M3D2 | $\mu_{M2D2}$ | 0.025 |
| 20 | MM2D1 | $\mu_{MM2D1}$ | 0.0125 |
| 21 | M3D3 | $\mu_{M3D3}$ | 0.016 |
| 22 | IM3D2 | $\mu_{IM3D2}$ | 0.01 |
| 23 | MM3D1 | $\mu_{MM3D1}$ | 0.0625 |
| 24 | FrD1 | $\mu_{FrD1}$ | 0.02 |

*Source of data- www.barringer.com

### *A. Sensitivity Analysis*

Sensitivity refers to the change in the result obtained when one or more independent parameters considered in the calculations are varied. Sensitivity Analysis is a technique to check the sensitivity of the solution obtained. For that, keeping other factors constant, one of the parameters is varied.

### *B. Varying the Inspection Interval*

In the beginning we change the periodic inspection time at I1 keeping those at I2 and I3 constant. We observe that as we decrease the periodic time, the availability of the component decreases. This is so because in the beginning the component is new and the frequent inspection lead to time wastage and increases the possibility of minor repair work on the component. Thus decrease its availability. As shown in the table below:

Table -2 Sensitivity analysis for system availability varying inspection interval for degradation stage 1.

| S.No. | I1(hrs) | I2(hrs) | I3(hrs) | Availability |
|---|---|---|---|---|
| 1. | 50 | 150 | 100 | 0.8807 |
| 2. | 100 | 150 | 100 | 0.9194 |
| 3. | 200 | 150 | 100 | 0.9607 |
| 4. | 300 | 150 | 100 | 0.9731 |

| | | | | |
|---|---|---|---|---|
| 5. | 400 | 150 | 100 | 0.9792 |
| 6. | 600 | 150 | 100 | 0.9859 |
| 7. | 800 | 150 | 100 | 0.9893 |
| 8. | 1000 | 150 | 100 | 0.9991 |
| 9. | 1100 | 150 | 100 | 0.9922 |
| 10. | 1200 | 150 | 100 | 0.9931 |
| 11. | 1300 | 150 | 100 | 0.9934 |
| 12. | 1500 | 150 | 100 | 0.9942 |

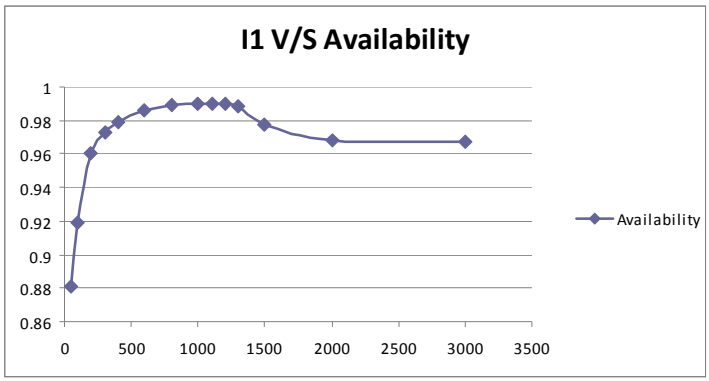



Fig 6.Inspection Interval for I1

Next, we change the periodic inspection time at I2 keeping those at I1 and I3 constant. We observe that when we increase the periodic inspection time there is very slight increase in availability of the component. This is so , because the system has degraded to an extent that it needs frequent inspection to increase the availability of the component.

Table -3 Sensitivity analysis for system availability varying inspection interval for degradation stage 2.

| S.No. | I1(hrs) | I2(hrs) | I3(hrs) | Availability |
|---|---|---|---|---|
| 1. | 250 | 50 | 100 | 0.9667 |
| 2. | 250 | 100 | 100 | 0.9695 |
| 3. | 250 | 150 | 100 | 0.9677 |
| 4. | 250 | 200 | 100 | 0.9682 |
| 5. | 250 | 400 | 100 | 0.9689 |
| 6. | 250 | 600 | 100 | 0.9692 |
| 7. | 250 | 800 | 100 | 0.9694 |
| 8. | 250 | 1000 | 100 | 0.9694 |

| | | | | |
|---|---|---|---|---|
| 9. | 250 | 1500 | 100 | 0.9696 |
| 10. | 250 | 2000 | 100 | 0.9697 |
| 11. | 250 | 10000 | 100 | 0.9700 |
| 12. | 250 | 20000 | 100 | 0.9697 |

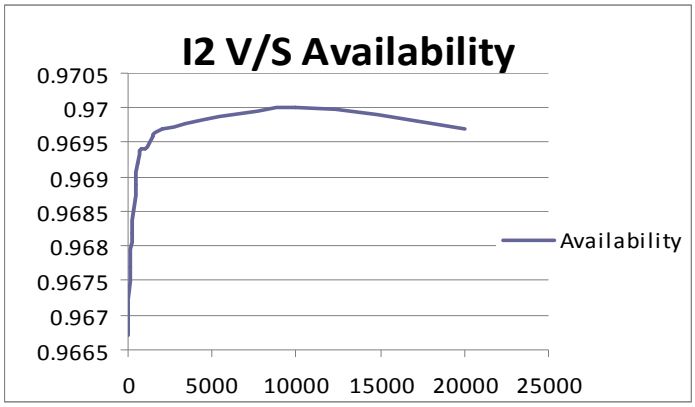



Fig 7. Inspection Interval for I2

Next, we change the periodic inspection time at I3 keeping those at I1 and I2 constant. We observe that, as we increase the periodic inspection time the availability of the component merely increases. This is so, because the component has degraded to a higher level and need frequent inspection.

Table -4 Sensitivity analysis for system availability varying inspection interval for degradation stage 3.

| S.No. | I1(hrs) | I2(hrs) | I3(hrs) | Availability |
|---|---|---|---|---|
| 1. | 250 | 150 | 25 | 0.9676 |
| 2. | 250 | 150 | 50 | 0.9676 |
| 3. | 250 | 150 | 100 | 0.9677 |
| 4. | 250 | 150 | 150 | 0.9677 |
| 5. | 250 | 150 | 200 | 0.9677 |
| 6. | 250 | 150 | 300 | 0.9678 |
| 7. | 250 | 150 | 400 | 0.9678 |
| 8. | 250 | 150 | 500 | 0.9678 |

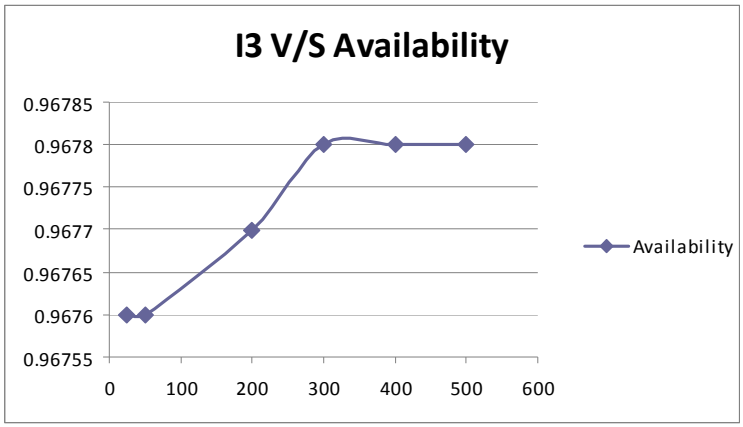



Fig.8. Inspection Interval for I3

## V. CONCLUSION

In this paper availability model considering multi stage degradation, periodic inspection, and condition based maintenance and random failure is developed. The system model is solved analytically using MARKOV approach.A sensitivity analysis is conducted to see the effect of variation in probability for various maintenance decision, variation of inspection interval and final degraded states with and without failure. As far as frequency of inspection is concerned at stage D1, less frequent inspection should be done as the health of the component is very good and unnecessary inspection will only lead to time wastage and reducing our component availability.

At stage D2, the inspection should be done frequently asthe health of the component is fairly good.

At stage D3, inspection work should be done quite frequently as the health of the component has deteriorated and frequent inspection would readily provide us information about its degradation so we can undertake necessary repair actions.

This model can be used by practicing maintenance engineers as it takes care of the cost involved.

This model cannot handle none exponential distribution and only Markov's Approach is used.

## VI. REFERENCE

1. Andrew K.S. Jardine, Daming Lin, DraganBanjevic, 2006, "A review on machinery diagnostics and prognostics implementing condition-based maintenance", Journal of Mechanical systems and signal processing 20, 1483–1510.

2. Karin S. de Smidt-Destombes, Matthieu C. van der Heijden, Aart van Harten, 2004, "On the availability of a k-out-of-N system given limited spares and repair capacity under a condition based maintenance strategy". Journal of Reliability Engineering and System Safety 83, 287–300.

3. Castanier, Grall, Be´renguer, 2005, "A condition-based maintenance policy with non-periodic inspections for a two-unit series system", Journal of Reliability Engineering and System Safety 87, 109–120.

4. Yongjin(James)Kwon, RichardChiou, LeonardStepanskı, 2009, "Remote, condition-based maintenance for web-enabled robotic system", Journal of Robotics and Computer-Integrated Manufacturing 25 , 552– 559.

5. Ling Wang, Jian Chu, Weijie Mao, 2009, "A condition-based replacement and spare provisioning policy for deteriorating systems with uncertain deterioration to failure", European Journal of Operational Research 194, 184–205.

6. Fangji Wu, TianyiWang, JayLee, 2010, "An online adaptive condition-based maintenance method for mechanical systems".Journal of Mechanical Systems and Signal Processing 24 , 2985–2995.

7. ZhigangTian, Tongdan Jin, Bairong Wu, Fangfang Ding, 2011," Condition based maintenance optimization for wind power generation systems under continuous monitoring".Journal of Renewable Energy 36, 1502-1509.

8. ZhigangTian, Haitao Liao, 2011, "Condition based maintenance optimization for multi-component systems using proportional hazards model". Journal of Reliability Engineering and System Safety 96, 581–589.

9. Rosmaini Ahmad, ShahrulKamaruddin, 2012, "An overview of time-based and condition-based maintenance in industrial application". Journal of Computers & Industrial Engineering 63 , 135–149.

10. Cui Yanbin, Cui Bo, 2012, "The Condition Based Maintenance Evaluation Model on On-post Vacuum Circuit Breaker", Journal of Systems Engineering Procedia4 , 182 – 188.

11. Qingfeng Wang, JinjiGao, 2012, "Research and application of risk and condition based maintenance task optimization technology in an oil transfer station". Journal of Loss Prevention in the Process Industries 25, 1018-1027.

12. Chiming Guoa, Wenbin Wang, Bo Guoa, Xiaosheng Si, 2012, "A Maintenance Optimization Model for Mission-Oriented Systems Based on Wiener Degradation". Journal of Reliability Engineering and System Safety 90, 1856-1874.

13. Chryssaphinou O, Liminios N, Malefaki S.,2011, "Multi-state reliability systems under discrete time Semi-Markovian hypothesis", IEEE Trans. Reliability, 60(1):80-87.

14. Endrenyi J, Anders GJ.,2006, "Aging, maintenance, and reliability", IEEE Power and Energy Magazine, 59-67.

15. Majid MAA, Nasir M.,2011, "Multi-state system availability model of electricity generation for a cogeneration district cooling plant", Asian Journal of Applied Sciences, 4(4):431-438.

16. Gorjian N, Ma L, Mittinty M, Yarlagadda P, Sun Y.,2009,"A review on degradation models in reliability analysis", Proceedings of the 4th World Congress on Engineering Asset Management, Athens, Greece, 1-16.

17. Chen A, Guo RS, Yang A, Tseng CL,1998, "An integrated approach to semiconductor equipment monitoring", Journal of Chinese Society of Mechanical Engineering, 19(6):581-591.

18. Chen A, Wu GS, 2007," Real-time health prognosis and dynamic preventive maintenance policy for equipment under aging Markovian deterioration", Internatinal Journal of Production Research ,45(15):3351-3379.

19. Wang N, Sun S, Li S, Si S.,2010,"Modelling and optimization of deteriorating equipment with predictive maintenance and inspection", IEEE 17th International conference on Industrial Engineering and Engineering Management , 942-946.

20. Jardine AKS, Lin D, Banjevic D,2006,"A review on machinery diagnostic and prognostics implementing condition-based maintenance", Mechanical Systems and Signal Processing, 20:1483-1510.

21. Martin KF,1994," A review by discussion of condition monitoring and fault diagnosis in machine tools", International Journal of Machine Tools Manufacturing, 34(4):527-551.

22. Pandian A, Ali A.,2009,"A review of recent trends in machine diagnosis and prognosis algorithm", World Congress on Nature & Biologically Inspired Computing, 1731–1736.

23. Chen D, Trivedi KS.,2005,"Optimization for condition-based maintenance with Semi Markov decision process" Reliability Engineering and System Safety, 90(1): 25-29.

24. Grall A, Berenguer C, Dieulle L.,2002, "A condition-based maintenance policy for stochastically deteriorating systems" Reliability Engineering and System Safety ,76:167-180.

25. Marquez AC, Heguedas AS,2010, "Models for maintenance optimization: A study for repairable systems and finite time periods", Reliability Engineering and System Safety ,75:367-377.

26. El-Damcese MA Temraz NS,2011, "Availability and reliability measures for multi-state system by using Markov reward model".,Reliability:Theory and Application , 2:68-85.

27. Amari SV.,2004, "Optimal design of a condition- based maintenance model. Annual Reliability and Maintainability Symposium", 528-533.

28. Fricks RM., Trivedi, KS.,1997, "Modelling failure dependencies in reliability analysis using stochastic Petri nets, in ESM", Proceedings 11th European Simulation Multiconf., Istanbul, Turkey, SCS Europe, 1-22.

29. Xie W, Hong Y, Trivedi K.,2005, "Analysis of a two-level software rejuvenation policy", Reliability Engineering and System Safety, 87(1):13-22.

30. Qiang H, Zhihua D, Xiao Z.,2010, "Application of HSMM on NC machine's state recognition", International conference on E-health Networking, Digital Ecosystems and Technologies, 189-191.

31. Taghipour S, Banjevic D, Jardine AKS,2010, "Periodic inspection optimization model for a complex repairable system", Reliability Engineering and System Safety, 95(9):944 – 952.

Optical Fiber Technology xxx (2013) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Optical Fiber Technology

www.elsevier.com/locate/yofte

# Design and analysis of a refractive index sensor based on dual-core large-mode-area fiber

Koppole Kamakshi [a,*], Vipul Rastogi [a], Ajeet Kumar [b]

[a] Department of Physics, Indian Institute of Technology Roorkee, Roorkee, India
[b] Department of Applied Physics, Delhi Technological University, New Delhi, India

ARTICLE INFO

*Article history:*
Received 28 September 2012
Revised 15 February 2013
Available online xxxx

*Keywords:*
Refractive index sensor
Large-mode-area fiber
Leakage loss
Dual-core fiber
Resonant coupling

ABSTRACT

We present a novel co-axial dual core large-mode-area (LMA) fiber design for refractive index sensing. In a dual-core fiber there is resonant coupling between the two cores, which is strongly affected by the refractive index (RI) of the outermost region. The transmittance of the fiber, therefore, varies sharply with the refractive index of surrounding medium. This characteristic of the proposed structure has been utilized to design a RI sensor. We have analyzed the structure by using the transfer matrix method. Our numerical results show that the proposed sensor is highly sensitive with the resolution of $2.0 \times 10^{-6}$ around $n_{ex} = 1.44376$. Effect of design parameters on sensitivity of the proposed sensor has also been investigated.



## 1. Introduction

Refractive index (RI) sensing is crucial for various industrial applications such as food processing, chemical and medical diagnostics. Over the past decades many fiber optic RI sensors have been exploited, aiming to measure the surrounding refractive index [1–12]. The sensing mechanisms utilized in these sensors include Fabry Perot interferometry [1,2], the index-dependent Bragg wavelength shift of a fiber Bragg grating (FBG) or long-period fiber grating (LPFG) [3–5] and variation of transmittance due to core-diameter mismatch at splices [9]. Interferometer-based sensors usually consist of two beams; one which serves as a sensing arm and while the other beam used as a reference arm. Sensing of the external RI in this type of sensors can be done by combining the two beams to generate an interference pattern. The main problem with interferometer based sensors is the complexity as they often require a mechanism to split the incoming light into two arms. In the case of fiber gratings (both FBG and LPFG), sensitivity is measured from the shifts of the transmission/reflectance spectra due to the influence of the external RI on the coupling conditions of the fiber gratings. Since LPFG couples light from the core mode to the cladding modes, it is highly sensitive when compare with the FBG-based RI sensor. Negative aspect of the fiber grating sensors is they are expensive because of the stringent grating fabrication processes and wavelength interrogation. Most of these sensors have showed the maximum performance in terms of index variation $\Delta n$ of the order of $10^{-5}$. Recently, we have proposed a low-cost core diameter mismatch sensor designed in a single single mode fiber [10]. It measures the RI values below that of fused silica with the resolution of about $10^{-4}$.

Here, as an alternative to the previously presented refractometers, we present a novel fiber optic RI sensor based on a co-axial dual-core LMA fiber. To achieve an efficient LMA design we have chosen the design parameters in such a way that only the $LP_{01}$ mode of the structure survives and higher order modes are stripped off. Variation in leakage loss of $LP_{01}$ mode of the fiber with surrounding RI has been utilized in designing the sensor. LMA helps in good coupling of light into the fiber and effective single mode operation helps in achieving sufficiently high sensitivity. We have numerically investigated the effect of design parameters on the response of the sensor. We show the maximum resolution of the sensor to be about $2.0 \times 10^{-6}$.

## 2. LMA fiber structure

The schematic of the proposed refractive index sensor set-up is shown in Fig. 1a. It uses a dual-core fiber having refractive-index profile shown in Fig. 1b. The dual core fiber consists of two cores: the inner core of width $a_{in}$ and the outer core of width $a_{out}$, two depressed cladding regions: inner cladding of thickness $b_{in}$ and outer cladding of thickness $b_{out}$. Both the cores have refractive index $n_1$ that corresponds to the RI of pure silica. The inner and outer cladding regions of the fiber have refractive index $n_2$ and $n_3$ respectively, which can be attained by using fluorine doping into pure silica using modified chemical vapor deposition (MCVD) or plasma

* Corresponding author.
*E-mail address:* kamakshikoppole@gmail.com (K. Kamakshi).

1068-5200/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.yofte.2013.03.007

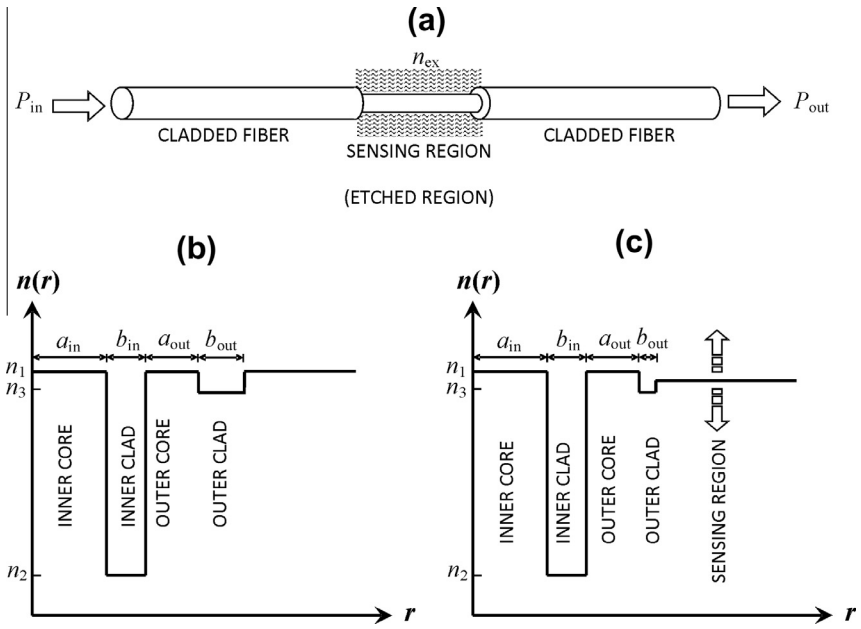



**Fig. 1.** (a) Schematic diagram of proposed refractive index sensor, (b) refractive index profile of a co-axial dual core LMA fiber, and (c) refractive index profile of an etched co-axial dual core LMA fiber.

activated chemical vapor deposition (PCVD) techniques [13]. Here, we define the relative index difference between the inner core and inner cladding as $\Delta = \frac{n_1^2 - n_2^2}{2n_1^2}$. The outer most high-index region, $n_1$ beyond outer cladding make all the modes leaky. Thus, all the modes of the structure suffer from finite leakage loss. Higher order modes can be efficiently stripped off by the suitable choice of design parameters. The sensing region is formed by etching the middle portion of the dual-core fiber. The fiber is etched up to outer clad and some portion of the outer clad is also etched to expose it to the liquid to be sensed. The refractive-index profile of the etched region is shown in Fig. 1c. We have analyzed the proposed structure using the transfer matrix method (TMM) to calculate the leakage loss of the fundamental mode [14]. In TMM an arbitrary refractive index profile is divided into a large number of homogeneous layers by using the staircase approximation. The relationship between the field coefficients in the layers can be derived by applying the boundary conditions at the interface of two consecutive layers, which was given by a $2 \times 2$ matrix. The field coefficients of the innermost and the outermost layer of the profile can then be connected by simply multiplying the transfer matrices of all the intermediate layers. By applying suitable boundary conditions in the innermost and outermost layers, a complex eigenvalue equation for propagation constant ($\beta$) is formed. Then the leakage loss of a mode can be calculated from the imaginary part of the propagation constant ($\beta_i$) by using the following relation [15]:

$$\text{Leakage loss} = 8.868 k_0 \text{Im}(n_{\text{eff}}) \quad (1)$$

We have then calculated the corresponding transmittance in 3 cm length of the fiber.

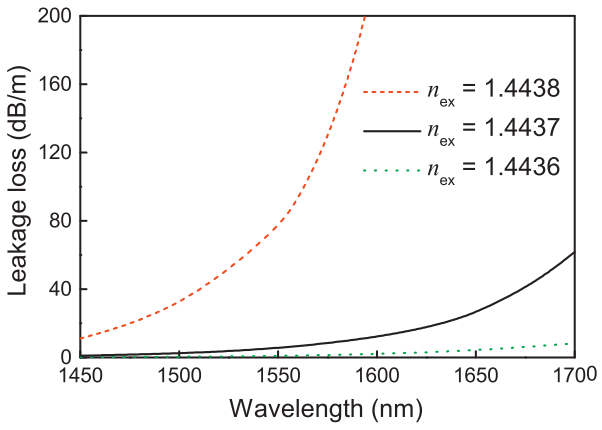



**Fig. 2.** Spectral variation of leakage loss of the $LP_{01}$ mode.

## 3. Numerical results and discussion

The cladded fiber is a five layer structure with the parameters $\Delta = 0.006$, $a_{in} = 14$ μm, $b_{in} = 3$ μm, $a_{out} = 10$ μm, $b_{out} = 26$ μm, and $n_3 = 1.4435$. A fiber with these design parameters introduces 678 dB/m leakage losses to the $LP_{11}$ mode and leaks out all the higher order modes in 3 cm of propagation length, however, the fiber also introduces a nominal loss of 0.01 dB to $LP_{01}$ mode during this propagation length. Typical length of the cladded fiber before the sensing region is few tens of cm, which is sufficient to strip-off higher order modes and the fiber in the sensing region works as single-mode fiber.

The width of outer cladding in the sensing region has been taken $b_{out} = 3.5$ μm unless stated otherwise. We have carried out the numerical simulations on the performance analysis of the proposed design by calculating the leakage loss of the fundamental mode. Spectral variation of leakage loss of the fundamental mode for three different values of $n_{ex}$ is shown in Fig. 2. The leakage loss curves shown in Fig. 2 are in fact the tails of resonance peaks and show that the resonance between the two cores is highly sensitive to $n_{ex}$. Since the higher value of $n_{ex}$ makes the outer core more leaky, one can see that the resonant wavelength shifts towards longer wavelength side as the value $n_{ex}$ decreases and lower value of $n_{ex}$ shows small spectral variation of leakage loss.

To investigate the RI sensing characteristics of the fiber we have studied the variation of leakage loss at $\lambda = 1550$ nm as a function of RI of the last layer is shown in Fig. 3. When we increase $n_{ex}$ it taps more power from the inner core and the leakage loss of $LP_{01}$ mode increases. A sharp increase in leakage loss of the mode for $n_{ex} > 1.44378$ is due to the resonant leakage of power from inner

Please cite this article in press as: K. Kamakshi et al., Design and analysis of a refractive index sensor based on dual-core large-mode-area fiber, Opt. Fiber Technol. (2013), http://dx.doi.org/10.1016/j.yofte.2013.03.007

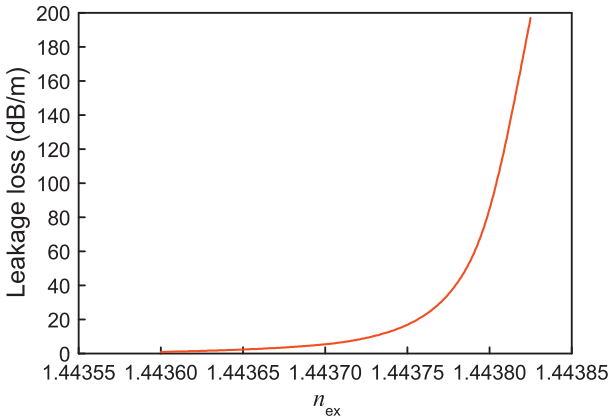



**Fig. 3.** Variation of leakage loss of fundamental mode with $n_{ex}$.

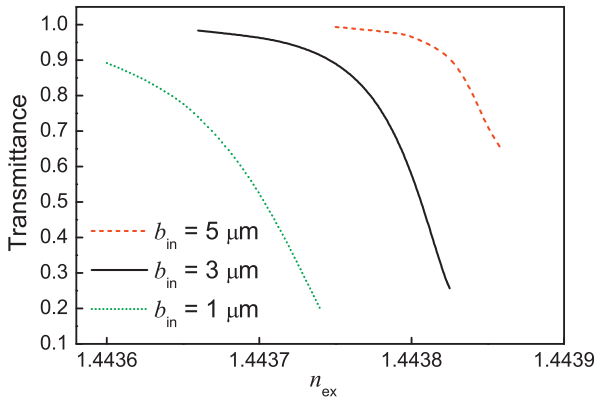



**Fig. 5.** Transmittance versus the refractive index of the external medium ($n_{ex}$) in the range $n_{ex}$ = 1.4436 to $n_{ex}$ = 1.4439 with $\lambda$ = 1550 nm for different values of $b_{in}$.

core to the outer core. This makes the device highly sensitive to the surrounding RI. Transmittance of proposed RI sensor as a function of RI of the last layer for three different lengths is shown in Fig. 4. The leakage loss of the structure is such that with 3 cm sensing length the transmittance varies from 100% to 20% in the sensing range of RI. A smaller length would reduce this variation to a smaller span as shown in Fig. 4 and 3 cm is also a practically suitable length of the sensing region for realization of the sensor. A longer length would bring down the transmittance to a much smaller value and would require more sensitive detectors while using the source of moderate power. It is also clear that the response of the sensor is nonlinear and it has different sensitivities in different regions. The resolution of the sensor, which we define as $\Delta n$ variation for 1% change in transmittance is $1.01 \times 10^{-5}$ around the RI 1.44368. The resolution improves to $2.0 \times 10^{-6}$ around $n_{ex}$ = 1.44376. Such a resolution corresponds to the one obtained in surface plasmon based sensor [16].

We have also studied the effect of design parameters on the sensitivity of the proposed sensor. Fig. 5 shows the effect of inner clad width ($b_{in}$) on the variation of transmittance with $n_{ex}$. From the figure it is obvious that the sensitivity changes slightly with $b_{in}$. For example, for 2 μm increment in $b_{in}$ from $b_{in}$ = 3 μm, the transmittance varies from 90% to 64% in the sensing range 1.44382–1.44386, while 2 μm decrement in $b_{in}$ gives the transmittance variation from 90% to 19% in the range 1.4436–1.44375. However, one can notice the change in high sensitivity RI range. On increasing $b_{in}$ the coupling between the two cores becomes narrower. This is because the resonance peak becomes sharper when the separation between the two cores increases [17].

We have also studied the impact of outer core width ($a_{out}$) and outer clad width ($b_{out}$) on the sensitivity of the sensor and the results are summarized in Figs. 6 and 7. It is clear from the figures that the outer cladding thickness and outer core width do not have significant effect on the sensitivity of the sensor. The variation of the transmittance of the fiber changes from 91% to 41% in the range 1.44380–1.44388 for $a_{out}$ = 12 μm, while $a_{out}$ = 8 μm results in change of transmittance of the fiber from 91% to 19% in the range 1.44366–1.44374 as shown in Fig. 6. Fig. 7 corresponds to the variation of the transmittance as a function of $n_{ex}$ for the changes in $b_{out}$. The high sensitivity RI range is 1.44373–1.44380 for 1 μm increment in $b_{out}$ from $b_{out}$ = 3.5 μm and 1.44376–1.44385 for 1 μm decrement in $b_{out}$. It can be clearly seen that even if the sensitivity does not vary significantly, the high sensitivity sensing range shifts with $b_{out}$ and $a_{out}$. This shifting is due to the change in resonance wavelength.

We have then worked out the tolerances with respect to $n_2$ and $n_3$ on the sensitivity of the proposed sensor. Our results show that a small change of $\pm 5 \times 10^{-4}$ in the values of $n_2$ does not cause any significant effect on the sensitivity of the sensor as this change is quite small as compared to the index difference between $n_1$ and $n_2$. However, a variation of $\pm 2 \times 10^{-4}$ in $n_3$ shifts the resonance wavelength significantly and hence the range of highly sensitive region shifts as shown in Fig. 8. The RI range of the proposed sensor corresponds to the chemical substances like W25240 – Diesel Clean-Up (RI = 1.4436 @ 20 °C) which can be exploited as diesel fuel additive and sulfuryl chloride which is a source of chlorine.

We have also designed the DCRLF based RI sensor for the range 1.4439–1.4444 which corresponds to RI of Cassava seed oil (1.444)

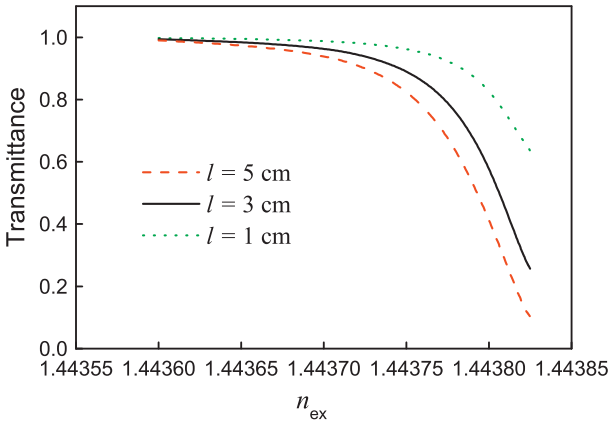



**Fig. 4.** Sensor response to the RI of external medium for different sensing lengths.

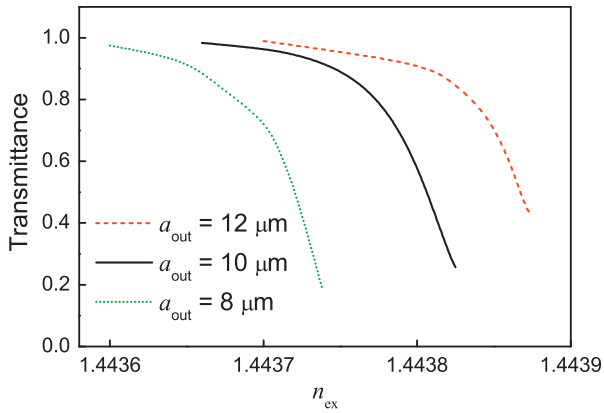



**Fig. 6.** Transmittance versus the refractive index of the external medium ($n_{ex}$) in the range $n_{ex}$ = 1.4436 to $n_{ex}$ = 1.4439 with $\lambda$ = 1550 nm for different values of $a_{out}$.

Please cite this article in press as: K. Kamakshi et al., Design and analysis of a refractive index sensor based on dual-core large-mode-area fiber, Opt. Fiber Technol. (2013), http://dx.doi.org/10.1016/j.yofte.2013.03.007

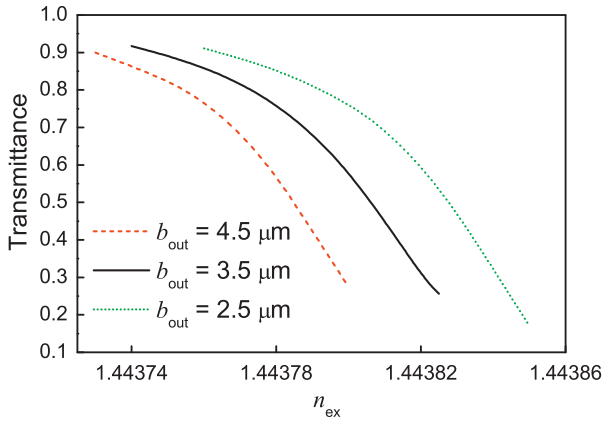



**Fig. 7.** Transmittance versus the refractive index of the external medium ($n_{ex}$) in the range $n_{ex}$ = 1.44372 to $n_{ex}$ = 1.44386 with $\lambda$ = 1550 nm for different values of $b_{out}$.

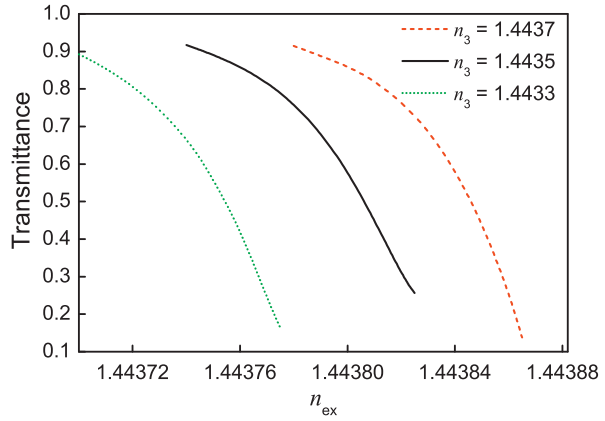



**Fig. 8.** Effect of outer cladding index on transmittance.

[18]. The various design parameters are $\Delta$ = 0.006, $a_{in}$ = 14 μm, $b_{in}$ = 1.5 μm, $a_{out}$ = 6.5 μm, $b_{out}$ = 2.5 μm, $n_3$ = 1.4425 and $l$ = 3 cm. The transmittance of the fiber is plotted in Fig. 9. We can see the variation in transmittance is from 86% to 36% and the maximum resolution of the sensor is $6.93 \times 10^{-6}$.

In practical implementation the sensing region is kept straight in an enclosure filled with the test liquid but the cladded fiber before and after the sensing region may undergo bending. However, such bends will have large bending radii. We have analyzed the bending performance of the proposed fiber using the method described in [19], which is a suitable method for calculating the bend loss of the multilayer, fiber such as the one proposed here. Bending loss $\alpha_b$ (dB/km) of the fiber is given by [19]

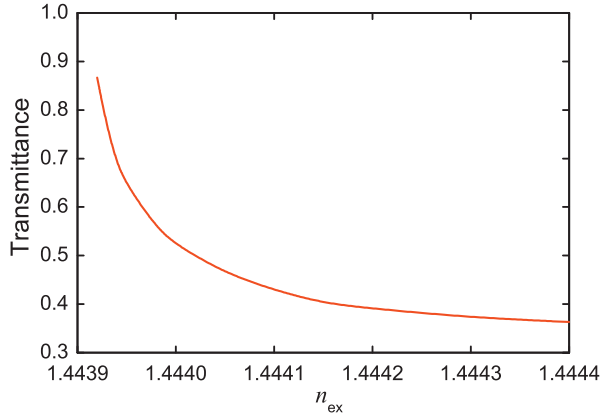



**Fig. 9.** Transmittance versus the refractive index of the external medium ($n_{ex}$) in the range $n_{ex}$ = 1.4439 to $n_{ex}$ = 1.4444 with $\lambda$ = 1550 nm.

$$\alpha_b = 1.09\sqrt{\frac{\pi}{aRW^3}}S(V,W)\exp\left[\frac{-4RW^3\Delta}{3V^2a}\right] \quad (2)$$

$$S(V,W) = \frac{a^2}{K_0^2(W)}\left[\int_0^\infty \frac{E^2(\rho)}{E^2(a)}\rho d\rho\right]^{-1}, \quad W = b^{1/2}V \quad (3)$$

where $E(\rho)$ represents the radial field distribution in the straight fiber, $a$ denotes the fiber core radius and $R$ is the bending radius. Other parameters have their usual meaning. Eq. (2) represents an approximate pure bend loss formula applicable to arbitrary refractive-index profile fibers and neglects small corrections due to field deformation at the curved fiber. Such an approximation is justified in view of high sensitivity of bend loss with radius of curvature [19]. Our results show bending loss less than 0.1 dB/m at bending radii larger than 20 cm.

## 4. Conclusions

We present a novel co-axial dual core LMA fiber based high sensitivity refractive index sensor. We have investigated the performance of the sensor with respect to the design parameters. Proposed sensor is highly sensitive in the range 1.44375–1.44382 where the transmittance of the fiber varies from 90% to 25%. We expect that the structure would have ample potential in chemical sensor applications.

## Acknowledgments

One of the authors (K. Kamakshi) acknowledges the financial support provided by Indian Institute of Technology, Roorkee, Ministry of Human Resources and Development (MHRD) Government of India. This work has been partially supported by the UK–India Education and Research Initiative (UKIERI) major award on "Application specific microstructured optical fibers". We also acknowledge help of our student Seema in carrying out some of the calculations.

## References

[1] W. Liang, Y. Huang, Y. Xu, R.K. Lee, A. Yariv, Highly sensitive fiber Bragg grating refractive index sensors, Appl. Phys. Lett. 86 (2005) 151122(1)–151122(3).
[2] Y. Wang, D.N. Wang, M. Yang, W. Hong, P. Lu, Refractive index sensor based on a microhole in single-mode fiber created by the use of femtosecond laser micromachining, Opt. Lett. 34 (2009) 3328–3330.
[3] R. Jha, J. Villatoro, G. Badenes, V. Pruneri, Refractometry based on a photonic crystal fiber interferometer, Opt. Lett. 34 (2009) 617–619.
[4] N. Ni, C.C. Chan, L. Xia, P. Shum, Fiber cavity ring-down refractive index sensor, IEEE Photon. Technol. Lett. 20 (2008) 1351–1353.
[5] M. Han, F. Guo, Y. Lu, Optical fiber refractometer based on cladding-mode Bragg grating, Opt. Lett. 35 (2010) 399–401.
[6] L. Rindorf, O. Bang, Highly sensitive refractometer with a photonic crystal-fiber long-period grating, Opt. Lett. 33 (2008) 563–565.
[7] D.K.C. Wu, B.T. Kuhlmey, B.J. Eggleton, Ultrasensitive photonic crystal fiber refractive index sensor, Opt. Lett. 34 (2009) 322–324.
[8] Y. Jung, S. Kim, D. Lee, K. Oh, Compact three segmented multimode fiber modal interferometer for high sensitivity refractive-index measurement, Meas. Sci. Technol. 17 (2006) 1129–1133.
[9] J. Villatoro, D. Monzon, Low-cost optical fiber refractive-index sensor based on core diameter mismatch, J. Lightw. Technol. 24 (2006) 1409–1413.
[10] K. Kamakshi, V. Rastogi, A. Kumar, J. Rai, Design and fabrication of a single mode optical fiber based refractive index sensor, Microw. Opt. Technol. Lett. 52 (2010) 1408–1411.
[11] M. Iga, A. Seki, K. Watanabe, Hetero-core structured fiber optic surface plasmon resonance sensor with silver film, Sensors Actuators B 101 (2004) 368–372.

[12] H.J. Patrick, A.D. Kersey, F. Bucholtz, Analysis of the response of long period fiber gratings to external index of refraction, J. Lightw. Technol. 16 (1998) 1606–1612.
[13] P.K. Bachmann, Method of Manufacturing Fluorine-doped Optical Fibers, US Patent Specification 4468413, 1984.
[14] K. Thyagarajan, S. Diggavi, A. Taneja, A.K. Ghatak, Simple numerical technique for the analysis of cylindrically symmetric refractive-index profile optical fibers, Appl. Opt. 30 (1991) 3877–3879.
[15] K. Saitoh, M. Koshiba, T. Hasegawa, E. Sasaoka, Chromatic dispersion control in photonic crystal fibers: application to ultra-flattened dispersion, Opt. Express 11 (2003) 843–852.
[16] E.K. Akowuah, T. Gorman, S. Haxha, Design and optimization of a novel surface plasmon resonance biosensor based on Otto configuration, Opt. Express 17 (2009) 23511–23521.
[17] U. Peschel, T. Peschel, F. Lederer, A compact device for highly efficient dispersion compensation in fiber transmission, Appl. Phys. Lett. 67 (1995) 2111–2113.
[18] T.O.S. Popoola, O.D. Yangomodou, Extraction, properties and utilization potentials of cassava seed oil, Biotechnology 5 (2006) 38–41.
[19] E.G. Neumann, Single-mode Fibers: Fundamentals, Springer-Verlag, 1988 (Chapter 5).

Please cite this article in press as: K. Kamakshi et al., Design and analysis of a refractive index sensor based on dual-core large-mode-area fiber, Opt. Fiber Technol. (2013), http://dx.doi.org/10.1016/j.yofte.2013.03.007

*I. J. Computer Network and Information Security*, 2013, 5, 64-72
Published Online April 2013 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijcnis.2013.05.08

# Effect of Black Hole Attack on MANET Routing Protocols

Jaspal Kumar, M. Kulkarni, Daya Gupta
Panipat Institute of Engineering & Technology, India
National Institute of Technology, Karnataka, India
Delhi College of Engineering, University of Delhi, India

***Abstract*** — Due to the massive existing vulnerabilities in mobile ad-hoc networks, they may be insecure against attacks by the malicious nodes. In this paper we have analyzed the effects of Black hole attack on mobile ad hoc routing protocols. Mainly two protocols AODV and Improved AODV have been considered. Simulation has been performed on the basis of performance parameters and effect has been analyzed after adding Black-hole nodes in the network. Finally the results have been computed and compared to stumble on which protocol is least affected by these attacks.

***Index Terms*** — MANETs, Routing Protocols, Black hole attack, AODV, Improved AODV

## I. Introduction

A Mobile Ad hoc Network (MANET) is an independent system of mobile routers attached by wireless links. The routers move freely and organize themselves randomly. The network topology may change rapidly and spontaneously. Such a network may operate in an individual fashion or may be connected to the Internet. Multi hop, mobility, large network size combined with device heterogeneity, bandwidth and battery power constrain make the design of passable routing protocols a major challenge. In recent years, a lot of routing protocols have been proposed for MANETs, out of whom two major protocols AODV and Improved AODV have been discussed in this paper.

## II. MANET Characteristics

*Autonomous and infrastructure less:* MANET is a self-organized network, independent of any established infrastructure and centralized network administration. Each node acts as a router and operates in distributed manner.

*Multi-hop routing:* Since there exists no dedicated router, so every node also acts as a router and aids in forwarding packets to the intended destination. Hence, information sharing among mobile nodes is made available.

*Dynamic network topology:* Since MANET nodes move randomly in the network, the topology of MANET changes frequently, leading to regular route changes, network partitions, and possibly packet losses.

*Variation on link and node capabilities:* Every participating node in an ad hoc network is equipped with different type of radio devices having varying transmission and receiving capabilities. They all operate on multiple frequency bands. Asymmetric links may be formed due to this heterogeneity in the radio capabilities.

*Energy-constrained operation:* The processing power of node is restricted because the batteries carried by portable mobile devices have limited power supply.

*k scalability:* A wide range of MANET applications may involve bulky networks with plenty of nodes especially that can be found in strategic networks. Scalability is crucial to the flourishing operation of MANET.

## III. MANET Applications

There are many applications of MANETs. Some of them are discussed below.

*Military Networks:* The latest digital military fields demand strong and consistent communication in different forms. Mostly devices are deployed in moving military vehicles, tanks, trucks etc which can share information randomly among them.

*Sensor Networks:* One more application of MANETs is the Sensor Networks. It is a network which consists of a large number of devices or nodes called sensors, which sense a particular incoming signal and transmit it to appropriate destination node.

*Automotive Applications:* Automotive networks are extensively discussed currently. Vehicles should be enabled to communicate on the road with each other and with traffic lights forming ad-hoc networks of diverse sizes. This network will provide drivers with information about the road conditions, traffic congestions and accident-ahead warnings which help in optimizing the traffic flow.

*Emergency services:* Ad hoc networks are broadly being used in rescue operations for disaster relief efforts during floods, earthquakes, etc.

## IV. Routing Protocols

MANET routing protocols are categorized into three main categories depending upon the criteria when the source node possesses a route to the destination, as shown in figure 1.

- Table driven/ Proactive
- Source initiated (demand driven) / Reactive
- Hybrid

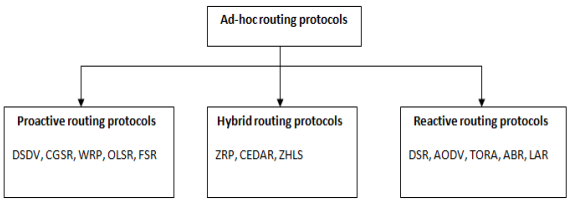



Figure 1 Classification of MANET Routing Protocols

### 4.1 Table Driven Routing Protocols

Table driven also known as proactive protocols maintain reliable and up to date routing information between all the nodes in an ad hoc network. In this each node builds its own routing table which can be used to find out a path to a destination and routing information is stored. Whenever there is any variation in the network topology, updation has to be made in the entire network [5]. Some of the main table driven protocols are:

Optimized Link State Routing protocol (OLSR)
Destination sequenced Distance vector routing (DSDV)
Wireless routing protocol (WRP)
Fish eye State Routing protocol (FSR)
Cluster Gateway switch routing protocol (CGSR)

### 4.2 Source Initiated Routing Protocols

In On-demand or Reactive routing protocols routes are formed as and when required. When a node desires to send data to any other node, it first initiates route discovery process to discover the path to that destination node. This path remains applicable till the destination is accessible or the route is not required. Different types of on demand driven protocols have been developed such as:

Ad hoc On Demand Distance Vector (AODV)
Dynamic Source routing protocol (DSR)
Temporally ordered routing algorithm (TORA)
Associativity Based routing (ABR)

### 4.3 Hybrid Routing Protocols

This type of routing protocols combines the features of both the previous categories. Nodes belonging to a particular geographical region are considered to be in same zone and are proactive in nature. Whereas the communication between nodes located in different zones is done reactively. The different types of Hybrid routing protocols are:

Zone routing protocol (ZRP)
Zone-based hierarchical link state (ZHLS)
Distributed dynamic routing (DDR)

## V. AODV Routing Protocol

Ad hoc on demand distance-vector protocol is a pure reactive protocol and it incorporates the features of both DSDV and DSR. AODV was proposed by Perkins et al. as advancement to the earlier protocol DSDV. DSDV is purely proactive protocol based on the traditional Bellman – Ford algorithm. In contrast AODV is on – demand in which route is established only when it is required. The routing in AODV is accomplished in two phases: route discovery and route maintenance as discussed below.

*Route Discovery:* Route discovery process is initiated whenever a node needs to send data packet to the destination and there is no valid route available in its routing table. The source node then broadcasts a route request (RREQ) packet to all its neighbor nodes, which then forward the request to their neighbor nodes and the process repeats as shown in figure 3. Each node is assigned a sequence no. and a broadcast ID which is incremented each time the node issues a RREQ packet. The broadcast ID together with the node's IP address, exclusively identifies a RREQ [3] which is unique in nature. The RREQ packet contains following fields:

- Sequence number of RREQ
- Broadcast ID
- The most recent sequence number of the destination

Upon receiving RREQ by a node which is either destination node or an intermediate node with a fresh route to destination, it replies by unicasting a route reply (RREP) message to the source node. As the RREP is routed back along the reverse path, intermediate nodes along this path set up forward path entries to the destination in their routing tables. When the RREP reaches source node, a route from source to destination node is established. Figure 2 indicates the path of the RREP from the destination node to the source node [9].

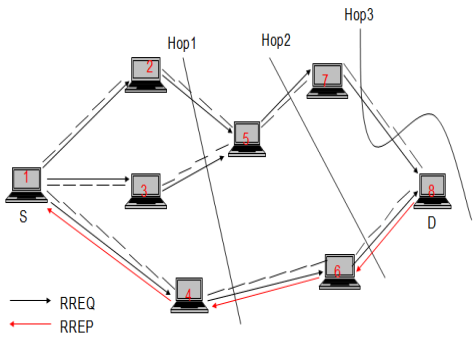



Figure 2 Propagation of Route Request packet & Route Reply packet

*Route Maintenance:* Once a route is established between source and destination, it needs maintenance usually at the source end. When any link break or failure is detected, it is declared as invalid and a route error (RERR) message is flooded to all the nodes in the network. These nodes in turn broadcast the RERR to their ancestor nodes and so on till the influenced source node. Then it is the source node who may decide whether to stop sending data or restart the route

discovery process for that particular destination by sending out a fresh RREQ message to its neighbor nodes.

### 5.1 Limitations of AODV

AODV besides being an efficient routing algorithm possesses some limitations due to which it is easily attacked by the external intruders. Following are a few limitation of AODV protocol.

1. If the sequence number of source node is lower than that of intermediate nodes, it may lead to inconsistent routes.
2. Multiple route reply packets and periodic beaconing may result in heavy routing overhead.
3. The overall performance starts degrading as network grows.

## VI. Improved AODV Routing Protocol

It is an enhanced version of AODV and is hybrid in nature. IAODV mainly integrates two features: Multipath and Path accumulation as explained below [20].

*Multipath:* Multipath AODV reduces the route discovery frequency as compared to single path AODV. It finds multiple paths between a source and a destination in a route discovery process. Single path AODV initiates a new route discovery when it detects one path failure to the destination, whereas in multipath it creates a fresh route in case all the existing routes fail or expire. It also reduces the number of similar routes between source and destination nodes. A path with most similar nodes has a higher probability to create common links.

*Path accumulation:* Path accumulation feature enables us to append all discovered paths between source and destination nodes to the control messages as shown in figure 3(a). Hence, at any intermediate node the route request (RREQ) packet contains a list of all nodes traversed. Each node receiving these control messages updates its routing table. It adds paths to each node contained in these messages.

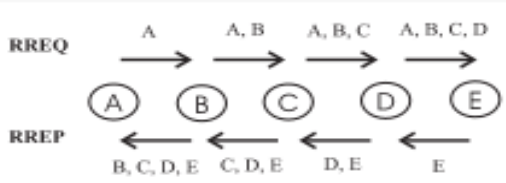



Figure 3(a) Path accumulation

### 6.1 Types of IAODV operations

*Route discovery:* Route discovery as shown in figure 3(b) includes a route request message (RREQ) and route reply message (RREP). Suppose Node 2 wants to communicate with Node 9. Each node forwarding the RREQ creates a reverse route to 2 used when sending back the RREP. When sending back the RREP, nodes on the reverse route create routes to node 9.

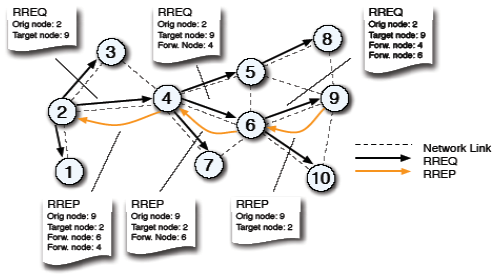



Figure 3(b) Route discovery

*Route maintenance:* It includes a Route Error message (RERR). Route maintenance is a process of responding to topology update which can happen after a route has been initially created. To maintain these paths, the nodes continuously examine the active links and update the valid timeout field of entries in its routing table during data transfer. If a node receives a data packet for a destination it does not have a valid route for, it must reply with a RERR message. When creating the RERR message, the node makes a list containing the address and sequence number of the unapproachable node. Then the node updates all the entries in routing table. The key purpose is to notify about all the additional routes being created during discovery phase that are no longer available. The node then sends a list in the RERR packet which is broadcasted in the network. This distribution process is illustrated in figure 3(c). The link between nodes 6 and 9 breaks, and node 6 generates an RERR. Only nodes having a route table entry for node 9 propagate the RERR message further.

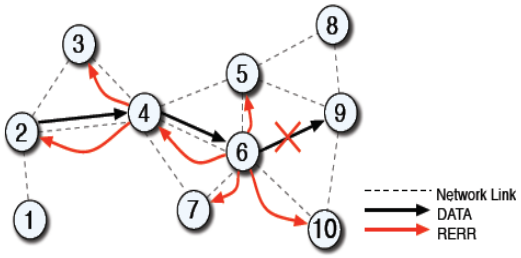



Figure 3(c) Route maintenance

### 6.2 AODV vs Improved AODV

This section briefly describes the comparison between the features possessed each by AODV and its improved version. Table 1 shows the comparison between the two protocols.

Table 1: Comparison of AODV and IAODV

| Parameters | AODV | IAODV |
|---|---|---|
| Path accumulation | No | Yes |
| Multipath | No | Yes |
| Routing type | Reactive | Hybrid |
| Security | Less | More than AODV |

## VII. Role of attacks in MANETs

MANETs often experience unusual security attacks because of their following features such as dynamically changing topology, lack of central monitoring, mutual algorithms and absence of a centralized certification authority etc. Generally mobile ad hoc networks are affected by two kinds of attacks which are classified as passive and active. Passive attacks do not affect the functionality of network, but may attempt to find out vital information by listening to traffic [16]. It is difficult to identify such attacks as under these attacks the network operates normally. These attacks basically obtain critical routing information through sniffing. Such attacks are usually complex to identify and protection against such attacks is also difficult. Moreover, it is sometimes not even possible to trace the exact location of the attacker node. Generally, such type of attacks is prevented with the help of encryption. On the other hand, active attacks aim to modify the transmitted data by adding random packets and force to interrupt the operation of network. The main purpose is to pull all packets towards the attacker for analysis or to obstruct the network communication. Such attacks can be detected and the nodes can be identified.

Passive attacks can be debarred using various encryption mechanisms. Only active attacks can be accepted out at routing level. These can either be inner outer. Inner attacks can be passive and active. Passive attacks are unauthorized disruption of the routing packets and active attack is from outside sources to degrade or damage message flow within the network nodes [17]. In order to combat these attacks a secure ad hoc environment should provide confidentially, integrity, authenticity, availability and non-repudiation. The following are few attacks based on routing mechanisms [19].

*Black Hole:* It is a network layer attack in which all the packets are dropped by sending fake packets. The attacker node advertises itself and declares having the shortest path to the destination. All the nodes start forwarding packets to this node and then the malicious node just drops all the incoming packets. Black hole attack mainly attacks AODV protocol.

*Worm Hole:* It is also a network layer attack in which two malicious nodes that is part of foreign private network record packets at one location in the network, rebroadcast them to another location through their private network and retransmits them into the network [2].

Table 2 below shows the defence against various attacks on MANETS. Every secure solution aims to resolve the network attacks by escalating the secrecy of network through encryption techniques.

Table 2: Comparison of routing attacks

| Attack | Layer | Solution |
|---|---|---|
| Blackhole | Network | SAODV |
| Wormhole | Network | Packet leashes |
| Repudiation | Application | ARAN |
| DoS | Multi layer | ARIADNE |
| Routing | Network | SEAD |

In SAODV and SEAD, hash function is used to authenticate the hop count [19]. SEAD is a Distance Vector Routing Protocol presented by Hu, Johnson & Perrig. It uses efficient one-way Hash functions to provide authentication for both the sequence number and metric field in each routing entry. It avoids asymmetric cryptography to protect against DoS attacks. ARIADNE is another On-Demand Routing Protocol presented by Hun, Johnson & Perrig based on DSR. It maintains authenticity on end-to-end basis, using symmetric key cryptography. It can authenticate routing messages using either shared secret keys or digital signatures. ARAN relies on a trusted certificate server. Every node forwarding a Route Request or Reply is required to sign the packet. It detects and protects against malicious actions carried out by 3rd party and peers. SAODV suggests using digital signatures to authenticate non-mutable data in an end-to-end manner. Hash chains are used to secure mutable fields such as hop count. It is an extension to AODV Routing Protocol. Packet Leashes have been proposed to detect and defend the wormhole attacks in ad hoc networks.

## VIII. Black hole attack

The attacker or malicious node usually exploits some routing protocols to distribute itself as having the direct and shorter route to source whose packets it wants to grab [1]. Once the attacker adds itself between the communicating nodes, it can do anything malicious with the packets passing between them. It can then choose to drop the packets thereby creating Denial of Service attacks. Security in mobile ad-hoc network is the most vital concern for basic functionality of a network [6]. Accessibility of network services, confidentiality and integrity of data can be achieved by assuring that security issues have been met. MANETs suffer from security attacks because they possess open medium, rapidly changing topology, lack of central administration and non-robust defence mechanism. These factors lead to various security threats in mobile ad hoc networks [2]. Black hole Attacks are classified into two categories. In single blackhole attack there is only one malicious node within a zone [22]. Whereas in collaborative blackhole attack multiple nodes in a group act as malicious nodes [23].

The work done in earlier years based on security issues i.e. attacks (particularly Black hole) on MANETs is mainly based on reactive routing protocols like Ad-Hoc on Demand Distance Vector (AODV) [11]. Black hole attack has been reviewed and its effects have been

analyzed by studying how these attacks disturb the performance of an ad hoc network. A very little attention has been given on the impact of Black hole attack on routing protocols and comparison of vulnerability of these protocols against the attacks [7]. The goal of this work is to study the effects of Black hole attacks on reactive routing protocols i.e. Ad-Hoc on Demand Distance Vector (AODV) and Improved Ad-Hoc on Demand Distance Vector (IAODV).

Black hole attacks mostly affect proactive protocols and with a great effect on AODV protocol [10, 4]. It is a type of denial of service attack in which the malicious node attracts all the packets by advertising the shortest path from it to destination to all the neighbours. Thus absorbs all the packets without forwarding them. Any node wants to transmit data first sends a Route Request message to all its neighbours including the malicious node. The malicious node is the first to reply to the source and therefore sends Route Reply quickly back to the source node. When source node receives RREP it immediately forwards packet to that path. On receipt of data the blackhole node start dropping all the incoming packets. The complete scenario is shown in figure 4.

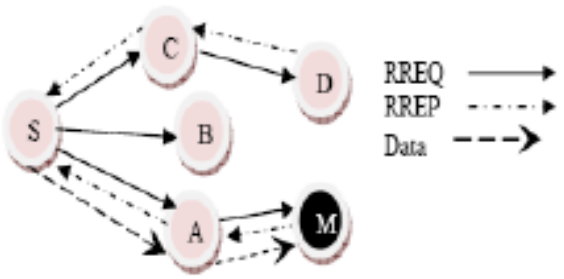



Figure 4: Malicious node in ad hoc network

Node S is supposed to be the source node desiring to correspond with destination node, D. Firstly, node S will send a RREQ message to all its neighbouring nodes which is received by nodes i.e. nodes A, B and C. Let us assume that node A has a route to the destination so it will be the first to send a RREP message back to source node S. Before this process node M being the blackhole node, will also send a false RREP message and send it to node A with a very high destination sequence number than the destination node to the source node. So node S will assume A as the shortest path to reach the destination and send data. But this is actually malicious node and not destination. Hence all the data will be trapped. This is the part of route discovery phase. In Route Maintenance phase, if any node detects any line break or node failure, it sends a Route Error (RERR) message to all the nodes that are currently using that particular route. Black hole attack in AODV protocol can be performed in two ways [21]. Black hole attacks caused by RREP and by RREQ as discussed in table 3.

Table3: Two ways of Black hole attack

| Caused by RREQ | Caused by RREP |
|---|---|
| Set the initial IP address in RREQ to the IP address of source node | Set the initial IP address in RREP to the IP address of source node |
| Set the destination IP address in RREQ to the IP address of destination node | Set the destination IP address in RREP to the IP address of destination node |
| Set the destination IP address of IP header to broadcast address | Set the destination IP address of IP header to the IP address of node that RREQ has received |
| Set the source IP address of IP header to its own IP address and put high sequence number and low hop count in the RREQ field | Set the source IP address of IP header to its own IP address |

8.1 Method to add a malicious node

The main setback of black hole attack is to hinder the communication from source to destination. To add malicious nodes in AODV the following procedure has been implemented [24].

First we need to modify aodv.cc and aodv.h files:

In aodv.h:

```
bool    malicious;
```

In aodv.cc:

```
malicious = false;
if(strcmp(arqv[1], "hacker") == 0) {
        malicious = true;
        return TCL_OK; }
```

Next we need to modify the TCL file to set a malicious node:

```
$ns at 0.0 "[$mnode_(i) set ragent_] hacker"
if  (malicious == true ) {
drop (p,DROP_RTR_ROUTE_LOOP); }
```

To protect MANETs from outside attacks, the routing protocols must fulfil certain set of requirements to guarantee the correct functioning of all the paths from source to destination. These are:

- Only the authorized nodes shall be able to execute route discovery processes
- Negligible exposure of network topology
- Early detection of distorted routing messages
- Avoiding formation of loops
- Avert redirection of data from shortest paths

8.2 Algorithm

Step1: Source node broadcasts RREQ to neighbours
Step2: Source node receives RREP from neighbours
Step3: Source node selects shortest and next shortest path based on the number of hops
Step4: Source node checks its routing table for single hop neighbouring nodes only
Step5: If the neighbour node is in its routing table then route data packet
Else
The node is malicious and sends false packets to that node
Step 6: Invoke the route discovery
Inform all the neighbouring nodes about the stranger
Step 7: Add the status of stranger to the routing table of source node

Step 8: Again send packet to neighbouring node
Step 9: If step 5 repeats then broadcast the malicious node as black hole
Step 10: Update the routing table of source node after every broadcast
Step 11: Repeat step 4 to 10 until packet reaches the destination node correctly

## IX. Simulation Environment

We have implemented Black hole attack in an ns2 simulator [15]. CBR (Constant Bit Rate) application has been implemented. The problem is investigated by means of collecting data, experiments and simulation which gives some results, these results are analyzed and decisions are made on their basis. The simulator which is used for simulation is ns2. Using ns2, we can implement your new protocol and compare its performance to TCP. To evaluate the performance of a protocol for an ad hoc network, it is necessary to analyze it under practical conditions, especially including the movement of mobile nodes. Simulation requires setting up traffic and mobility model for performance evaluation. Table 4 shows the parameters that have been used in performing simulation.

Table 4: Simulation Parameters

| Parameters | Value |
|---|---|
| Simulator | Ns-2.34 |
| Data packet size | 512 byte |
| Simulation time | 1000 sec |
| Environment size | 1000 x 1000 |
| Number of nodes | 50 |
| Transmission range | 250m |
| Pause time | 2 s |
| Observation parameters | PDF, end-to-end delay, overhead |
| No. of malicious node | 1 |
| Traffic Type | CBR |
| Mobility | 60 m/s |
| Routing Protocols | AODV and IAODV |

### 9.1 Mobility Model

There exists a variety of mobility models proposed by Sanchez and Manzoni [25], what we have implemented in our simulation is the random waypoint mobility model. A mobility model is used to describe the movement of a mobile node its location and speed variation over time while the simulation of a routing protocol. The random waypoint mobility model is the only model that is widely implemented & analyzed in simulation of routing protocols because of its simplicity and availability. It was first proposed by Johnson and Maltz [26]. At the start of the simulation each mobile node waits for a specified time called pause time, $t_p$ and randomly selects one location. A MN chooses a new random destination after staying at its previous position for a time period of $t_p$ till its expiry. A node travels across the area at a random speed distributed uniformly from $v_0$ to $v_{max}$ where $v_0$ and $v_{max}$ represent the minimum and maximum node velocities. This process of choosing random destination at random velocity is repeated again and again until the simulation is finished. We can say that a node is free to select its destination, speed and direction independent of the neighbor nodes.

### 9.2 Performance Analysis

Protocols can be compared by evaluating various performance metrics as shown below:

- ***Packet Delivery Ratio (Fraction)-*** It is calculated by dividing the number of packet received by destination through the number packet originated from source.

  $$PDF = (Pr/Ps)$$

  where Pr is total Packet received and Ps is the total Packet sent.
- ***Average end-to end delay-*** It is defined as the time taken for a data packet to be transmitted across an MANET from source to destination.

  $$D = (Tr - Ts)$$

  where Tr is receive Time and Ts is sent Time.
- ***Normalized Routing Overhead-*** It can also be defined as the ratio of routed packets to data transmissions in a single simulation. It is the routing overload per unit data delivered successfully to the destination node.

### 9.2 Experimental Setup

The simulation scenario and parameters used for performing the detailed analysis of Black hole attacks on MANET routing protocols is mentioned below. This section describes the how the performance parameters have been evaluated to simulate the routing protocols.

Following files have been used for simulation.

- *Input to Simulator:-*
  - Scenario File – Movement of nodes.
  - Traffic pattern file.
  - Simulation TCL file
- *Output File from Simulator:*
  - Trace file
  - Network Animator file
- *Output from Trace Analyzer:*
  - xgr file

Generation of Movement File:

*Traffic Pattern File:*

Ns cbrgen.tcl [-type cbr|tcp] [-nn nodes] [-seed seed] [-mc connections] [-rate rate]

Generation of Scenario File:

To generate the traffic movement file, following is example command.

./setdest -n <num_of_nodes> -p <pause_time> -s <maxspeed> -t <simtime> -x <maxx> -y <maxy> > < scenario file>

Here n – no. of nodes, p – pause time, s – speed, t - simulation time, and x, y – grid size.

9.3 NAM

NAM stands for Network Animator. It contains data for network topology. It starts with the command 'nam <nam-file>' where '<nam-file>' is the name of a nam trace file. At linux terminal command to run NAM is ./nam.

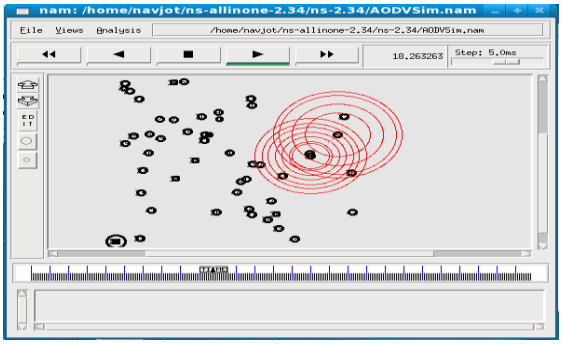



Figure 5 Network Scenario in NAM

After performing simulation as per network scenario shown in the figure 5, trace files are generated. Trace file contains following information:

- Send/Receive Packet
- Time
- Traffic Pattern
- Size of Packet
- Source Node
- Destination Node etc.

9.4 Analysis using Trace Analyzer

Awk script trace analyzer is used to analyze trace output from simulation. When files are analyzed using this trace analyzer an output xgr file is created which results in the generation of graphs.

## X. Results & Discussions

Using outputs from awk script following graphs and results are generated.

*Packet Delivery Ratio*

Simulation results of figure 6(a) show that under blackhole attack the packet delivery ratio of IAODV is more nearly similar to normal AODV, as compared to AODV under black hole attack.

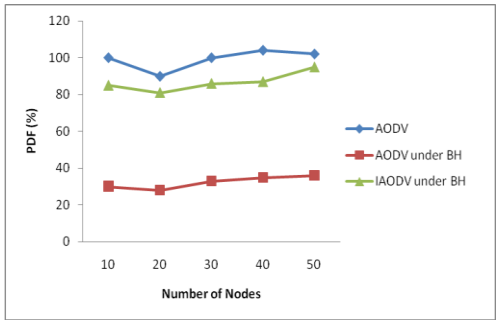



Figure 6(a): Impact of Black hole Attack on Packet Delivery Ratio.

*End To End Delay*

Simulation results in figure 6(b) show that IAODV has less end to end delay than AODV routing protocol under black hole attack.

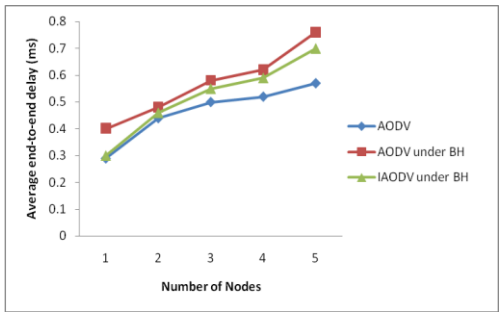



Figure 6(b): Impact of Black hole Attack on the Average End-to-End Delay

*Normalized Routing Overhead*

Simulation results in figure 6(c) show that IAODV has a high routing overhead as compared to AODV routing protocol under black hole attack.

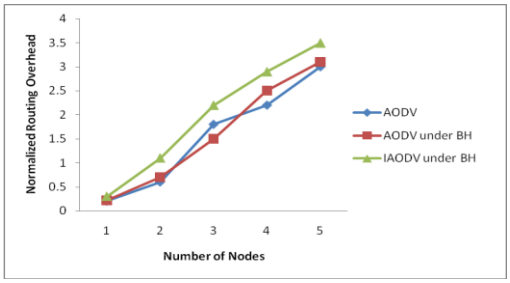



Figure 6(c): Impact of Black hole Attack on the Network overhead

Simulation results in figure 6 shows the average values for each parameter discussed above. It has been observed from the simulation scripts that when the protocols are under attack of black hole node, IAODV has a more packet delivery ratio, less average end to end delay and fewer overhead as compared to AODV routing protocol. It seems that IAODV is less effected than AODV whenever there is a black hole attack on the network. We analyzed that under black hole attack the PDF of IAODV is improved by larger amount of value than AODV. However, the values for average end-to-end delay are nearly similar in all the cases. Whereas there is a slight increase in the routing overhead this is quite negligible.

## Conclusion

In this paper, we have analyzed the Black hole attack with respect to different performance parameters such as end-to-end delay, overhead and packet delivery ratio. We have analyzed the vulnerability of two protocols AODV and Improved AODV under varying pause time. This study was conducted to evaluate the effect of Black hole attacks on the performance of these protocols. The Simulation results show that IAODV performs better

than AODV. The overhead of AODV is effected by twice as compare of IAODV. Also the effect on IAODV by the malicious node is less as compare to AODV. Based on our research and analysis of simulation result we draw the conclusion that IAODV is more vulnerable to Black hole attack than AODV. But still the detection of Black hole attacks in ad hoc networks is considered as a challenging task.

## FUTURE SCOPE

Simulation can be performed using other existing parameters. This work contains simulation based on random mobility model only. Other mobility models can also be studied and behaviour of protocols can be analyzed. Such networks are open to both the external and internal attacks due to lack of any centralized security system. Black hole attacks are needed to be analyzed on other existing MANET routing protocols such as DSDV, ZRP, DSR etc. Also attacks other than Black hole such as Wormhole, passive and active attacks shall be considered. They can be classified on the basis of how much they affect the performance of an ad hoc network. The early detection of Black hole attacks as well as the exclusion policy for such actions shall be carried out for advance research.

## REFERENCES


[1]. Tamilselvan, L. and Sankaranarayanan, V., Prevention of Black hole attack in MANET. The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, 21-21, 2007.

[2]. Dokurer, S.; Ert, Y.M.; and Acar, C.E., Performance analysis of ad hoc networks under Black hole attacks. Southeast Con, 2007, Proceedings IEEE, 148 – 153.

[3]. C. E. Perkins; E. M. Belding-Royer; and S. R. Das (2003). Ad hoc on demand distance vector (AODV) routing. RFC 3561. The Internet Engineering Task Force, Network Working Group.

[4]. Sheenu Sharma, Dr. Roopam Gupta," "Simulation Study of Black hole Attack in the Mobile Ad hoc Networks", November 2009.

[5]. M. Abolhasan, T. Wysocki, E. Dutkiewicz, " A Review of Routing Protocols for Mobile Ad-Hoc Networks", Telecommunication and Information Research Institute University of Wollongong, Australia, June, 2003.

[6]. Loay Abusalah, Ashfaq Khokhar, and Mohsen Guizani, "A Survey of Secure Mobile Ad Hoc Routing Protocols", IEEE Communications Surveys & Tutorials, Vol. 10, No. 4, Fourth Quarter 2008.

[7]. N. Shanti, Lganesan and K. Ramar, "Study of Different Attacks on Multicast Mobile Ad-Hoc Network".

[8]. M. Parsons and P. Ebinger, "Performance Evaluation of the Impact of Attacks on mobile Ad-Hoc networks"

[9]. H.L. Nguyen, U.T. Nguyen, "Study of Different Types of Attacks on Multicast in Mobile Ad-Hoc Networks," International Conference on Networking, Systems, Mobile Communications and Learning Technologies, Apr,2006

[10]. Satoshi Kurosawa, Hidehisa Nakayama, Nei Kato, Abbas Jamalipour, and Yoshiaki Nemoto. "Detecting Black hole Attack on AODV based Mobile Ad-hoc networks by Dynamic Learning Method". International Journal of Network Security, Vol.5, No.3, PP.338– 346, Nov. 2007

[11]. K. Biswas and Md. Liaqat Ali, "Security threats in Mobile Ad-Hoc Network", Master Thesis, Blekinge Institute of Technology" Sweden, 22nd March 2007

[12]. G. A. Pegueno and J. R. Rivera, "Extension to MAC 802.11 for performance Improvement in MANET", Karlstads University, Sweden, December 2006.

[13]. S. Lu, L. Li, K.Y. Lam, L. Jia, "SAODV: A MANET Routing Protocol that can Withstand Black Hole Attack.," International Conference on Computational Intelligence and Security, 2009.

[14]. S. Kurosawa et al., "Detecting Black hole Attack on AODV-Based Mobile Ad-Hoc Networks by Dynamic".

[15]. ns-2, Network simulator, http://www.isi.edu/nsnam/ns.

[16]. B. Dahill, B. N. Levine, E. Royer, and C. Shields, "A secure routing protocol for ad hoc networks," in Proceedings of the International Conference on Network Protocols (ICNP), pp. 78-87, 2002.

[17]. Y. Hu, A. Perrig and D. Johnson, Ariadne: A Secure On-demand Routing Protocol for Ad Hoc Networks, in Proceedings of ACM MOBICOM'02, 2002.

[18]. Janne Lundberg, Routing Security in Ad Hoc Networks. Tik-110.501 Seminar on Network Security.

[19]. Anuj K. Gupta, Harsh Sadawarti, "Secure Routing Techniques for MANETs", International Journal of Computer Theory and Engineering (IJCTE), ISSN: 1793-8201**,** Article No. 74, Vol.1 No. 4, pp. – 456-460, October 2009.

[20]. Nital Mistry, Devesh C Jinwala, Mukesh Zaveri, "Improving AODV Protocol against Black hole Attacks", Proceedings of the international multi conference of engineer and computer science vol. 2, 2010.

[21]. H.A. Esmaili, M.R. Khalili Shoja, Hossein gharaee, "Performance Analysis of AODV under Black Hole Attack through Use of OPNET Simulator", World of Computer

Science and Information Technology Journal (WCSIT), Vol. 1, No. 2, 49-52, 2011.

[22]. Latha Tamilselvan and Dr. V Sankaranarayanan, "Prevention of Black hole Attack in MANET", The 2nd International Conference on Wireless Broadband and Ultra Wideband Communications, 0-7695-2842-2/07, 2007.

[23]. Santhosh Krishna B V, Mrs.Vallikannu A.L , "Detecting Malicious Nodes For Secure Routing in MANETS Using Reputation Based Mechanism" International Journal of Scientific & Engineering Research, Vol. 1, Issue 3, ISSN 2229-5518, December-2010.

[24]. Harris Simaremare and Riri Fitri Sari, "Performance Evaluation of AODV variants on DDOS, Blackhole and Malicious Attacks", IJCSNS International Journal of Computer Science and Network Security, VOL.11 No.6, June 2011.

[25]. Tracy Camp, Jeff Boleng and Vanessa Davies, "A survey of Mobility Models for Ad hoc Network Research", Wireless Communications and Mobile computing: A special issue on Ad hoc network Research, vol 2, No5, pp. 483-502, 2002.

[26]. Tracy Camp, Jeff Boleng and V Davies, "A survey of Mobility Models for Ad Hoc Network Research", http://toilers.mines.edu last accessed on February 15, 2007.

**Bibliography**

**Jaspal Kumar** is currently a Ph. D. candidate in the department of Electronics and Communication Engineering at Delhi College of Engineering, Delhi (India). He received his B.E. and MTech. in 1992 and 2006. At present he is working as Asso.Prof with PIET, SAMAMKHA, and coordinating the various activities related to the electronics department. He has more than 21 years of rich experience in Industry as well as in Academics. He has been in the designing Microprocessor based Circuits in USA. He has been a visiting faculty to many institutions**.** His research interests include wireless networks, Digital electronics and communication systems

**Muralidhar Kulkarni** received his B.E.(Electronics Engineering) degree from University Visvesvaraya College of Engineering, Bangalore University, Bangalore, M. Tech (Satellite Communication and Remote Sensing) from Indian Institute of Technology, Kharagpur (IIT KGP) and PhD from JMI Central University, New Delhi in the area of Optical Communication networks. He has held the positions of Scientist in Instrumentation Division at the Central Power research Institute, Bangalore, Aeronautical Engineer in Avionics group of Design and Development team of Advanced Light Helicopter(ALH) project at Helicopter Design Bureau at Hindustan Aeronautics Limited(HAL),Bangalore, Lecturer (Electronics Engineering) at the Electrical Engineering Department of University Visvesvaraya College of Engineering, Bangalore and Assistant Professor in Electronics and Communication Engineering (ECE) Department at the Delhi College of Engineering (DCE), Delhi. He has served as Head, Department of Information Technology and Head, Computer Center at DCE , Govt. of National Capital territory of Delhi, Delhi. Currently, he is a Professor and HOD in the Department of Electronics and Communication Engineering (ECE) Department, National Institute of Technology Karnataka (NITK), Surathkal, Karnataka, India.

Dr. Kulkarni's teaching and research interests are in the areas of Digital Communications, Fuzzy Digital Image Processing, Adhoc networks, Wireless Sensor networks and Optical Communication & Networks. He has published several research papers in the above areas, in national and international journals of repute. For various contributions his Biography has been listed in the Marquis, Who's Who in Science & Engineering (2008). He has also authored four very popular books in Microwave &Radar Engineering, Communication Systems, Digital Communications and Digital Signal Processing.

**Daya Gupta** completed her Ph. D. in Computer Science. She joined Department of Computer Engineering at Delhi College of Engineering, India . where she is continuing as professor and HOD of CSE department and currently guiding BTech and MTech projects and dissertations and PhDs. She has published several research papers in referred journals and conferences. Her research interests include Computer Networks and Database Systems and ad-hoc networks.

International Journal of Current Engineering and Technology
ISSN 2277 - 4106

Available at http://inpressco.com/category/ijcet

Research Article

# Improving Consistency of Comparison Matrices in Analytical Hierarchy Process

Vandana Bagla[*a] , Anjana Gupta[b] and Aparna Mehra[c]

[a]Maharaja Agrasen Institute of Technology, Guru Gobind Singh Indraprastha University, Sector-22, Rohini, Delhi-110085, India.
[b]Delhi Technological University (DTU), Bawana Road, Delhi-110042, India.
[c]Indian Institute of Tecnology (IIT), Hauz Khas, Delhi-110016, India.

Accepted 13 March 2013, Available online 1June 2013, **Vol.3, No.2 (June 2013)**

***Abstract***

*In the field of decision-making, the concept of priority is archetypal and how priorities are derived influence the choices one makes. Priorities should not only be unique but should also reflect the dominance of the order expressed in the judgments of pair wise comparison matrix. In addition, judgments are much more sensitive and responsive to small perturbations. They are highly related to the notion of consistency of a pair wise comparison matrix simply because when dealing with intangibles, if one is able to improve inconsistency to near consistency then that could improve the validity of the priorities of a decision. This paper endeavors to accomplish nearly consistent matrices in pair wise comparisons by subsiding the effects of hypothetical decisions made by the decision makers. The proposed methodology efficiently improves group decisions by incorporating corrective measures for inconsistent judgments.*

***Key words***: *Multi Criteria Decision Making (MCDM), Eigen Value, Eigen Vector, Reciprocal Matrices, Analytical Hierarchy Process (AHP), Consistency.*

## 1. Introduction

Multi Criteria Decision Making (MCDM) is a sub-discipline of decision sciences that explicitly considers multiple criteria in decision-making environments. This concept is designed to make better choices when faced with complex decisions involving several dimensions. MCDM tactics are especially helpful when there is a need to combine hard data with subjective preferences, to make trade-offs between desired outcomes and to involve multiple decision makers. In the decision making process, there are typically multiple conflicting criteria that need to be assessed simultaneously with about same degree of precedence. This craves the search for an approach which deals the predicament with necessary sagacity to obtain a clear and unambiguous conclusion. It has been established in previous researches that the pair wise comparison methods can always be used to draw the final conclusions in a comparatively accessible and intelligible way.

The concept of pair wise comparisons is more than two hundred years old. Borda(1781) and Condorcet(1785) introduced it for voting problems in eighteenth century by using only 0 and 1 in the pair wise comparison matrices. The method was efficiently regulated by Thorndike(1920) to tackle the classical techniques of experimental psychology in early twentieth century. Thurstone(1927) also used pair wise comparisons for social values in twentieth century. Over the last three decades, a number of methods have been developed which use pair wise comparisons of the alternatives and criteria for solving MCDM problems. AHP proposed by Saaty(1980) has been a very popular approach to MCDM that involves pair wise comparisons for an objective inquisition. It has been applied during the last thirty years in many decision making situations and a wide range of applications in various fields. Advances in the decision sciences have led to the development of a number of approaches intended to minimize inconsistencies in pair wise comparisons to get closer to pragmatic scenario. Koczkodaj(1993) proposed a new definition of consistency in computational modeling. Ishizaka and Lusti(2004) designed a module to improve the consistency of AHP matrices. Antonio(2006) and Bozoki & Rapcsak(2007) introduced a new approach to gauge consistency of pair wise comparisons. Pair wise preference information is needed as input in many interactive multiple criteria decision making scenarios. The decision makers are required to make holistic binary comparisons among feasible alternatives. Information obtained from these comparisons is used to decrease the number of comparisons that have to be made when searching for the best (most preferred) alternative. Different approaches, with different assumptions for this problem have been presented. The simplest case of combinatorial consistency was analyzed by Davis(1963), other works are by Korhonen et al.(1984), and Koksalan & Sagala(1992) and many more, but only widely accepted measure of inconsistency is due to Saaty(1980).

In AHP, the calculated priorities are presumable only if the comparison matrices are consistent or near consistent. This condition is reached if (and only if) within the pair

*Corresponding Author's Email: vandana_6928@yahoo.com

wise comparison process the transitivity and reciprocity rules are respected. However it is often assumed that the decision produced by a group will always be better than that supplied by an individual. This seems plausible because multiple participants can bring differing expertise and perspectives to carry out any complex decision. Ideally, we should endeavor to curtail group imperfections and yet capitalize on inherent group advantages. In this study, it is unveiled that the consistency of a positive reciprocal matrix can be considerably improved by analyzing the stimulating factors of inconsistencies in positive reciprocal matrices. A convenient correction method based on the relative error is brought up which not only aspires to minimize the inconsistencies of such matrices but also to substantially reduce number of pair wise comparisons in decision making. This method can fully retain the effective information of original positive reciprocal matrix. It helps to solve practical problems effectively and enriches the theories and methods of decision analysis.

The paper is organized in six sections, Section 1 is introductory. Section 2 gives a brief introduction of methodology which laid the foundation of the present work. Antecedent of inconsistencies are analyzed and a corrective application based approach is introduced in section 3 which helps the decision-maker to build a consistent matrix with a controlled error by appreciably less number of pair wise comparisons. Section 4 illustrates the application part using proposed methodology to provide near consistent matrices. Section 5 elucidates the proposed methodology in group decision making. Finally, we present some findings and draw some conclusions, followed by giving recommendations for further research in section 6.

## 2. Overview of Pair wise Comparisons Approaches

MCDM is concerned with structuring and solving decision making problems involving multiple criteria (Figure1).

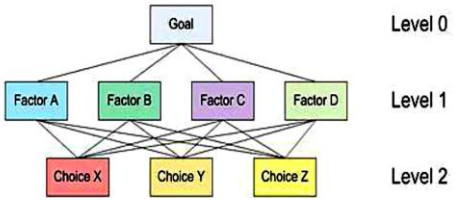



Figure1. MCDM Problem

Typically, there does not exist a unique optimal solution for such problems and it is necessary to use decision maker's preferences to get prioritized solutions. AHP is one of the most widely used methods to handle these types of problems.

### *2.1 Analytical Hierarchy (AHP)*

The analytical hierarchy process (AHP) is a decision making approach designed to aid in the solution of complex multiple criteria problems in number of application domains. The outcome of AHP is a prioritized weighting of each decision alternative. The first step in the analytical hierarchy process is to model the problem as a hierarchy. The hierarchy is a structured mean of describing the problem at hand. It consists of an overall goal at the top level, a group of options or alternatives for reaching the goal and a group of factors or criteria that relate the alternatives to the goal. In most cases the criteria are further broken down into sub criteria, sub-sub criteria and so on in many levels as per the requirement of the problem. Once the hierarchy has been constructed, the participants use the AHP to establish priorities for all its nodes. In this, the elements of a problem are compared in pairs with respect to their relative impact on a property they share in common. The pair wise comparison is quantified in a matrix form by using the scale of Relative Importance given in Saaty (1980) as shown in Table 1.
This scale has been validated for effectiveness, not only in many applications by a number of people, but also through theoretical comparison with a large number of other scales. During the elicitation process, a positive reciprocal matrix is formed in which $(i,j)^{th}$ element $a_{ij}$ is filled by the corresponding number from the Table 1.

Table 1. Analytic Hierarchy Measurement Scale

| Reciprocal Measure of Intensity of Importance | Definition | Explanation |
|---|---|---|
| 1 | Equal Importance | Two activities contribute equally to the objective |
| 3 | Weak importance of one over another | Experience and judgment slightly favor one activity over another |
| 5 | Moderate importance | Experience and judgments moderately favor one activity over another. |
| 7 | Strong Importance | An activity is strongly favored and its dominance is demonstrated in practice. |
| 9 | Absolute Importance | The evidence favoring one activity over another is of the highest possible order of affirmation |
| 2,4,6,8 | Intermediate values between two adjacent judgments. | When compromise is needed |

The number is chosen according to the following criterion.

$$\begin{cases} a_{ij}, & \text{if } x_i \text{ dominates } x_j \\ 1/a_{ij}, & \text{if } x_j \text{ dominates } x_i \\ 1, & \text{if } x_i \text{ and } x_j \text{ do not dominate over one another} \end{cases}$$

The matrix so formed is called the reciprocal matrix. This reciprocal matrix is used to calculate the local priority weight of each criterion. The local priority weight (w) is the normalized eigen vector of the priority matrix corresponding to the maximum eigen value of the matrix. For detailed reasoning of this account we refer to Forman (1990), Lunging (1992), Ball & Srinivasan (1994) and Bryson & Mobolurin (1994). An interesting property of the priority matrix is that if in addition its elements are such that

$$a_{ij}\, a_{jk} = a_{ik}, \qquad i \le j \le k \tag{1}$$

then the derived priority vector **w** satisfies

$$w_i / w_j = a_{ij}, \qquad i < j \tag{2}$$

Any reciprocal matrix satisfying (1) is called consistent. However in practice, the priority matrix seldom satisfies (1), thereby making it more important to define some relax measuring of consistency check, Saaty [8] introduced the concept of consistency index CI of a reciprocal matrix as the ratio $\frac{\lambda_{max} - n}{n-1}$ where $\lambda_{max}$ and $n$, respectively stand for the maximum eigen value and order of the reciprocal matrix. The obtained CI value is compared with the random index RI given in Table 2.

Table 2. Random consistency Index (RI)

| **N** | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| **RI** | 0 | 0 | 0.58 | 0.9 | 1.12 | 1.24 | 1.32 | 1.41 | 1.45 | 1.49 |

The Table 2 had been calculated as an average of CI's of many thousands matrices of the same order whose entries were generated randomly from the scale 1 to 9 with reciprocal force. The simulation results of RI for matrices of size 1 to 10 had been developed by Saaty (1980) and are given in Table 2. The ratio of CI and RI for the same order matrix is called the consistency ratio CR. In general, a consistency ratio of 10% or less is considered very good. If consistency is poor, inconsistency of judgments within the matrix has occurred and the evaluation process should therefore be reviewed and improved.

## 3. Research Methodology

### *3.1 Inconsistency of reciprocity*

Decision Makers are more likely to be cardinally inconsistent because they cannot estimate precisely measurement values even from a known scale and worse when they deal with intangibles (a is preferred to b twice and b to c three times, but a is preferred to c only five times). Consider the following judgment matrix for three alternatives.

$$\begin{bmatrix} 1 & 2 & 3 \\ 1/2 & 1 & a_{23} \\ 1/3 & a_{32} & 1 \end{bmatrix}$$

To fulfill the criteria of consistency i.e. $a_{ij}\, a_{jk} = a_{ik}$, $a_{23} = a_{21} \cdot a_{13} = 3/2$ and consequently $a_{32} = 2/3$. So the judgment is consistent if (and only if) $a_{23} = 3/2$ and $a_{32} = 2/3$. Here the first reason of inconsistency appears. The comparison scale of AHP (Table 1) has no such value. To overcome this difficulty, it is more appropriate to use the numbers of the form $\{a/b : a, b \in \{I^+\}\}$, where $I^+$ represents the set of positive integers excluding 0. This modified scale allows the decision maker to present a consistent judgment. This is perhaps the simplest way for composing priorities.

### *3.2 Inconsistency of Transitivity*

Whenever object a is related to b and object b is related to c, then the relation at hand is transitive provided object a is also related to c. In mathematical syntax:

$(a\ R\ b \text{ and } b\ R\ c) \Rightarrow a\ R\ c, \forall a, b, c \in A.$

The same property is respected in pair wise comparisons i.e. if A > B and B > C then A > C. Judgments are ordinally intransitive if A is preferred to B and B to C but C is preferred to A. One of the apparent reason is, large number of pairwise comparisons (n(n−1)/2) required to workout the attribute weights at a given level of hierarchy. This results in baffling responses relating to pair wise comparisons, as size of comparison matrix goes on increasing. Majority of available software entertain not more than nine attributes at a time. This is because of the fact that handling of a nine attribute matrix would need 36 pair wise comparisons at a time. It is an established fact that inconsistency of transitivity goes on increasing with the size of the matrix.

### *3.3 How to build a consistent matrix*

Proposed methodology intends to materialize an augmentation to overcome the above mentioned lacuna to some extent. Procedure outlined to implement the proposed methodology is as follows:

Step 1: A decision-maker should first rank all the n attributes to be weighed, according to their importance in the preferred domain. Reorder them in an ascending order of priorities.

Step 2: Exercise (n−1) comparisons among the consecutive criteria using the scale $\{a/b : a, b \in \{I^+\}\}$.

If any two or more criteria are equally significant, obvious priority of one over the other is 1 using the given scale.
Step 3: Priorities for remaining pairs (non-consecutive) can easily be computed logically as follows :

If B be prioritized *r* times to A and C is prioritize *s* times to B, then C is prioritized *r*×*s* times to A. Objective ratings to all potential pair wise comparisons can be provided in this manner and represented in a matrix form to provide weights to given set or criteria. It is conspicuous to mention here that priorities within a given pair of attributes are self-reciprocal, i.e. if B be prioritized *q* times to A then preference of A over B is *1/q* times.

Step 4: The procedure results in perfectly consistent comparison matrix supported by the fact $\lambda_{max} = n$ and hence
CI = 0. Eigenvector corresponding to this maximum eigenvalue provides the requisite criteria weights. Geometric mean or weighted geometric mean of individual judgments may be taken to accomplish aggregated matrices for the set of criteria at various levels of hierarchy.

Step 5: The nodes at each level are compared pair wise with respect to their contribution to the nodes above them to find their respective global weights. We rank each of the criteria in the final set by evaluating it with respect to upper level attributes separately. The evaluation process finally generates the global weights for each requisite criterion of interest. In a realistic scenario, the technique is very adaptable and can handle any number of attributes in a system. This simplification can reduce the calculation effort for the weights significantly, especially when judgment criteria are large in number and pair wise comparisons are difficult to be accomplished.

## 4. Numerical Example

We now illustrate the methodology by an independent survey conducted on referees $R_1$, $R_2$ and $R_3$. The three are catechized to rank four attributes P, Q, R and S in ascending order of priorities. Suppose the ranking awarded by $R_1$ to four attributes is (Q, R, S, P) in ascending order of priorities where R is prioritized 2 times over Q, S is prioritized 4 times over R and P is prioritized 5 times over S. Subject to $R_2$'s ranking (S, R, P, Q) where S & R are equally ranked, P is prioritized 3 times over R and Q is prioritized 7 times over P. Prioritized responses acceded by $R_3$ in ascending order is (R, Q, P, S). Here Q is prioritized 3 times over R, P is prioritized 2 times over Q and S is prioritized 5 times over P. Remaining priorities are calculated logically supervened by the methodology explained in section 3.3. Table 3 depicts the priorities procured by the four attributes accorded by the three referees.

Table 4 portrays the prioritized weights acquired by the four attributes tendered by the referees $R_1$, $R_2$ and $R_3$ subject to their rankings, guided by the methodology explained in section 3.3.

Table 3. Prioritized Weights Accorded by $R_1$, $R_2$ and $R_3$

| | | P | Q | R | S |
|---|---|---|---|---|---|
| **$A_1, A_2, A_3$ =** | P | 1 | 40, 1/7, 2 | 20, 3, 6 | 5, 3, 1/5 |
| | Q | | 1 | 1/2, 21, 3 | 1/8, 21, 1/10 |
| | R | | | 1 | 1/4, 1, 1/30 |
| | S | | | | 1 |
| $\lambda_{max}$ = 4, C.I. = 0, C.R. = 0, $\forall$ Ai, i = 1, 2, 3 | | | | | |

Table 4. Prioritized Weights Accorded by $R_1$, $R_2$ and $R_3$

| | $R_1$ | $R_2$ | $R_3$ |
|---|---|---|---|
| **P** | 0.784314 | 0.115385 | 0.15 |
| **Q** | 0.0196078 | 0.807692 | 0.075 |
| **R** | 0.0392157 | 0.0384615 | 0.025 |
| **S** | 0.025641 | 0.0384615 | 0.75 |

An aggregated comparison matrix is worked out by taking the geometric means of corresponding priorities in various components of each cell of Table 3. Table 5 shows the final rankings evolved by synthesizing the rankings provided by the referees $R_1$, $R_2$ and $R_3$.

Table 5. Aggregated Matrix Showing Final Weights of the Attributes using Proposed Methodology

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 2.2525 | 7.1138 | 1.4422 | 0.439024 |
| **Q** | 0.4439 | 1 | 3.152 | 0.6403 | 0.194812 |
| **R** | 0.1406 | 0.3172 | 1 | 0.2027 | 0.0617422 |
| **S** | 0.6934 | 1.5618 | 4.9334 | 1 | 0.304421 |
| $\lambda_{max}$ = 4, C.I. = 1.78862e − 07, C.R. = 1.99e − 07 | | | | | |

Final ranking is R, Q, S, P with CR approximately zero. Thus the efficiency of proposed methodology is substantially established in decision making scenarios.

## 5. Group Consistency

Group decision making is becoming increasingly important in decision scenarios associated with MCDM problems. AHP apparently arms the judgments which are consistent or near consistent (having CR < 0.1), whereas it discards inconsistent judgments affected by any of the above mentioned speculations. In realistic scenarios, only a handful of acceding judgments are taken into account defying the very objective of a legitimate conception. The proposed methodology provides an insight into the impediments to effective group processes and on techniques that can improve group decisions. A group decision-making methodology is being introduced as an effective approach for improving the targeted resolutions. It combines the following three components:

(1) The survey analysis
(2) The Analytic Hierarchy Process to produce judgments
(3) A logistic numerical assessment of irrational judgments.

We now illustrate the proposed methodology via an example in which eight referees say $R_1$, $R_2$, $R_3$, $R_4$, $R_5$, $R_6$, $R_7$ and $R_8$ are put to an inquisition to rank four attributes P, Q, R and S. To begin with, we first discuss the results working with the classical AHP methodology. Tables 6, 7, 8, 9, 10, 11, 12 and 13 exhibit the prioritized weights procured using AHP accorded by referees respectively. Note each decision maker $R_i$, i = 1, 2,…, 8, has to make $4(4-1)/2 = 6$ comparisons, viz. P with Q, R, S; then Q with R, S; and finally R with S, using scale from Table 1.

Table 6. Response to Inquisition by Referee $R_1$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/9 | 1/2 | 1/6 | 0489075 |
| **Q** | 9 | 1 | 7 | 3 | 0.590915 |
| **R** | 2 | 1/7 | 1 | 1/5 | 0.0771508 |
| **S** | 6 | 1/3 | 5 | 1 | 0.283027 |
| $\lambda_{max}$ = 4.09571, C.I. = 0.031905, C.R. = 0.03545 | | | | | |

Table 7. Response to Inquisition by Referee $R_2$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/5 | 3 | 6 | 0.211829 |
| **Q** | 5 | 1 | 7 | 9 | 0.657806 |
| **R** | 1/3 | 1/7 | 1 | 2 | 0.0827088 |
| **S** | 1/6 | 1/9 | 1/2 | 1 | 0.0476561 |
| $\lambda_{max}$ = 4.14228, C.I. = 0.047426, C.R. = 0.05270 | | | | | |

Table 8. Response to Inquisition by Referee $R_3$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1 | 1/2 | 1/7 | 0.0801702 |
| **Q** | 1 | 1 | 1/2 | 1/7 | 0.0801702 |
| **R** | 2 | 2 | 1 | 1/8 | 0.132531 |
| **S** | 7 | 7 | 8 | 1 | 0.707129 |
| $\lambda_{max}$ = 4.08661, C.I. = 0.0288702, C.R. = 0.032078 | | | | | |

Table 9. Response to Inquisition by Referee $R_4$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/2 | 6 | 4 | 0.368319 |
| **Q** | 2 | 1 | 1/2 | 5 | 0.327285 |
| **R** | 1/6 | 2 | 1 | 1/3 | 0.164289 |
| **S** | 1/4 | 1/5 | 3 | 1 | 0.140107 |
| $\lambda_{max}$ = 5.64218, C.I. = 0.547395, C.R. = 0.60822 | | | | | |

Table 10. Response to Inquisition by Referee $R_5$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/5 | 4 | 6 | 0.292937 |
| **Q** | 5 | 1 | 1/2 | 7 | 0.425984 |
| **R** | 1/4 | 2 | 1 | 3 | 0.238303 |
| **S** | 1/6 | 1/7 | 1/3 | 1 | 0.0427761 |
| $\lambda_{max}$ = 5.42098, C.I. = 0.473658, C.R. = 0.52629 | | | | | |

Table 11. Response to Inquisition by Referee $R_6$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/2 | 8 | 1/4 | 0.257225 |
| **Q** | 2 | 1 | 5 | 1/6 | 0.214766 |
| **R** | 1/8 | 1/5 | 1 | 7 | 0.245865 |
| **S** | 4 | 6 | 1/7 | 1 | 0.282144 |
| $\lambda_{max}$ = 9.33837, C.I. = 1.77946, C.R. = 1.97718 | | | | | |

Table 12. Response to Inquisition by Referee $R_7$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 2 | 3 | 1/4 | 0.266575 |
| **Q** | 1/2 | 1 | 1/5 | 1/6 | 0.0492226 |
| **R** | 1/3 | 5 | 1 | 7 | 0.416396 |
| **S** | 4 | 6 | 1/7 | 1 | 0.267806 |
| $\lambda_{max}$ = 6.30652, C.I. = 0.768841, C.R. = 0.85427 | | | | | |

Table 13. Response to Inquisition by Referee $R_8$

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1/2 | 1/3 | 1/4 | 0.0503552 |
| **Q** | 2 | 1 | 8 | 1/6 | 0.341091 |
| **R** | 3 | 1/8 | 1 | 5 | 0.272895 |
| **S** | 4 | 6 | 1/5 | 1 | 0.335659 |
| $\lambda_{max}$ = 7.85978, C.I. = 1.28659, C.R. = 1.42954 | | | | | |

An aggregated reciprocal matrix is developed by taking geometric mean of corresponding values of all the eight matrices for further calculations. Table 14 depicts the final weights of the attributes P, Q, R and S using classical AHP, taking into account all consistent and inconsistent judgments.

Table 14. Aggregated Matrix Showing Final Weights Using AHP

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 0.42729 | 1.86121 | 0.69361 | 0.212604 |
| **Q** | 2.34033 | 1 | 1.62658 | 0.67798 | 0.30739 |
| **R** | 0.537285 | 0.614787 | 1 | 1.36778 | 0.205152 |
| **S** | 1.44173 | 1.47497 | 0.731112 | 1 | 0.274855 |
| $\lambda_{max}$ = 4.31047, C.I. = 0.103489, C.R. = 0.11499 | | | | | |

Clearly resultant matrix shows incommensurate results as the attribute Q having awarded highest priority, yet not prioritized to S and similar other observations. Now we illustrate our proposed scheme on the same problem and simultaneously provide a comparison with the aforementioned result. We first seek the priorities for the four attributes P, Q, R, S, in ascending order from Table 14.

Following information is provided: attribute P is prioritized 1.86121 times over R, priority of S over P is 1.44173 times and that of Q over S is 0.67798 times.

Awarding the priorities to the remaining pair wise comparisons logically as explained in step 3 of the proposed methodology (section 3.3), we construct a pair wise a comparison matrix given by Table 15.

Table 15. Synthesis of Priorities Accorded by Proposed Methodology

| | P | Q | R | S | Weights |
|---|---|---|---|---|---|
| **P** | 1 | 1.02306 | 1.86121 | 0.693611 | 0.25275 |
| **Q** | 0.97746 | 1 | 1.81927 | 0.67798 | 0.247054 |
| **R** | 0.537285 | 0.549671 | 1 | 0.372667 | 0.135799 |
| **S** | 1.44173 | 1.47497 | 2.68336 | 1 | 0.364397 |
| | $\lambda_{max}$ = 4, C.Ic. = 1.43707e − 12, C.R. = 1.59678e − 12 | | | | |

Table 15 provides highly commensurate results with all the attributes having provided with prioritized weights.

## 6. Concluding Remarks

We have described the implementation of a corrective model, assisting the decision-maker in the construction of a consistent comparison matrix. It is conspicuous to mention that not only pair wise comparisons are substantially reduced, from $n(n-1)/2$ to $n-1$, but also appreciably legitimate results are shown, as evident by CR of Table 15 which is significantly lower than CR of Table 14. Another momentous advantage is that any number of attributes may be entertained in one set without any baffling responses. We hope that the proposed methodology can refocus the attention of researchers from the race of finding better judgments for inconsistent and near consistent matrices. Future work will include incorporation of the proposed approach into practical software applications, i.e. case studies will be processed and evaluated. Detailed evaluation of the present approach with other similar approaches can be obtained through these case studies.

## References


A. Ishizaka and M. Lusti (2004), An expert module to improve the consistency of AHP matrices, *International transactions in operational research (ITOR)*, Blackwell Publishing, vol. 11, no.1, pp. 97-105.

E.H. Forman (1990), Random indices for incomplete pairwise comparison matrices, *European Journal of Operational Research*, vol. 48, pp. 153-155.

E.L. Thorndike (1920), A constant error in psychological ratings, *Journal of Applied Psychology*, vol. 4, pp. 25-29.

F. Lunging (1992), Analytical hierarchy in transportation problems: an application for Istanbul, *Urban Transportation Congress of Istanbul*, vol. 2, pp. 16-18.

H.A. Davis (1963), The Method of Pair wise Comparisons, Griffin, London.

J. Antonio Alonso (2006), Consistency in the analytic hierarchy process: a new approach, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* , vol. 14, no.4, pp. 445-459.

J.C. Borda(1781), Mmoire sur les lections au scrutin, Histoire de lAcademie Royale des Sciences.

L. L. Thurstone(1927), The method of paired comparisons for social values, *Journal of Abnormal and Social Psychology*, vol. 21, pp. 384-400.

M. Condorcet (1785), Essai sur lapplication de lanalyse la probabilit des dcisions rendues la pluralit des voix, Paris.

M. Koksalan and P. Sagala (1992), An interactive approach for choosing the best of a set of alternatives, Journal of the Operational Research Society, vol. 43, pp. 259-263.

N. Bryson and A. Mobolurin (1994), An approach to using the analytic hierarchy process for solving multiple criteria decision making problems, *European Journal of Operational Research*, vol. 76, pp. 440-454

P. Korhonen, J. Wallenius and S. Zionts (1984), Solving the discrete multiple criteria problem using convex cones, *Management Science*, vol. 30, pp. 1336-1345.

S. Bozoki and T. Rapcsak (2007), On Saatys and Koczkodaj's inconsistencies of pair wise comparison matrices, Computer and Automation Institute, Hungarian Academy of Sciences.

T.L. Saaty (1980), The Analytic Hierarchy Process, McGraw Hill, New York.

V.C. Ball, J. Noel and Srinivasan (1994), Using the analytic hierarchy process in house selection, *Journal of Real Estate Finance And Economics*, vol. 9, pp. 69-85.

W. W. Koczkodaj (1993), A new definition of consistency of pair wise comparisons, *Mathematical Computational Modeling*, vol. 18. No. 7, pp. 79-84.

# Low-Voltage MOS Current Mode Logic Multiplexer

*Kirti GUPTA [1], Neeta PANDEY [1], Maneesha GUPTA [2]*

[1] Dept. of Electronics and Communication Engineering, Delhi Technological University, Delhi, India
[2] Dept. of Electronics and Communication Engineering, Netaji Subhash Institute of Technology, Delhi, India

Kirtigupta22@gmail.com, n66pandey@rediffmail.com, maneesha_gupta60@yahoo.co.in

***Abstract:*** *In this paper, a new low-voltage MOS current mode logic (MCML) multiplexer based on the triple-tail cell concept is proposed. An analytical model for static parameters is formulated and is applied to develop a design approach for the proposed low-voltage MCML multiplexer. The delay of the proposed low-voltage MCML multiplexer is expressed in terms of the bias current and the voltage swing so that it can be traded off with the power consumption. The proposed low-voltage MCML multiplexer is analyzed for the three design cases namely high-speed, power-efficient, and low-power. Finally, a comparison in performance of the proposed low-voltage MCML multiplexer with the traditional MCML multiplexer is carried out for all the cases.*

## Keywords

MOS current mode logic, low-voltage, triple-tail cell.

## 1. Introduction

The rapid advances in the VLSI technology have led to the development of high-resolution mixed-signal applications [1]-[2]. These applications demand high performance digital circuits to be integrated with analog circuitry on the same chip. The traditional CMOS logic style is not suitable as it generates a large amount of switching noise [3]-[4]. Many alternative logic styles have been suggested in literature [5]-[12]. Among them, MOS current mode logic (MCML) style is the most preferred option for high-resolution mixed-signal integrated circuits due to the reduced switching noise [12]-[13]. Also, MCML style exhibits better power-delay than CMOS at high frequencies [14]-[15]. Hence, MCML is suitable for designing high-speed communication systems [15]-[21] wherein a multiplexer is a key element for serialization of parallel data during transmission.

The implementation of traditional MCML multiplexer is based on the series-gating approach (i.e. stacked source-coupled transistor pairs) [22]. This approach requires that all the stacked transistor pairs should operate in saturation region thereby limiting the power supply requirement. The power supply may however be lowered by reducing the number of stacked transistor pair levels with triple-tail cell concept [23]-[27]. In this paper, a new low-voltage MCML multiplexer based on the triple-tail cell concept is proposed. An analytical model for static parameters is formulated and is used to size transistors of the proposed low-voltage multiplexer. From the knowledge of the transistor sizes, the delay is expressed in terms of the bias current and the voltage swing so that it can be traded off with the power consumption. Then, the proposed low-voltage multiplexer for high-speed, power-efficient and low-power design cases is illustrated and finally its performance is compared with the traditional MCML multiplexer for each case.

In this paper, the operation of the traditional MCML multiplexer is briefly reviewed in Section 2. Then, the new low-voltage MCML multiplexer is proposed and its analytical formulations for different static parameters and delay are presented in Section 3. The analysis of the proposed multiplexer for the three design cases, namely high-speed, power-efficient, and low-power, and its performance comparison with the traditional MCML multiplexer is discussed in Section 4. Finally, the paper is concluded in Section 5.

## 2. Traditional MCML Multiplexer

A traditional 2:1 multiplexer with differential inputs, namely SEL A and B is shown in Fig. 1 [28]. It consists of two levels of source-coupled transistor pairs to implement the logic function and a constant current source $M_{TR1}$ to generate bias current $I_{SS}$. The differential SEL input drives the lower level transistor pair $M_{TR2}$-$M_{TR3}$ that alternatively activates the upper level transistor pairs $M_{TR4}$-$M_{TR5}$ and $M_{TR6}$-$M_{TR7}$. When differential input SEL is high, $M_{TR3}$ is off, the bias current $I_{SS}$ flows through $M_{TR2}$ and is steered either to $M_{TR4}$ or $M_{TR5}$ according to the differential input A. Conversely, when differential input SEL is low, the bias current $I_{SS}$ flows through $M_{TR3}$ and is steered to one of the two transistors, i.e. either $M_{TR6}$ or $M_{TR7}$ depending on the differential input B. The bias current $I_{SS}$ is converted to the differential output voltage ($V_Q - \overline{V_Q}$) through the transistors $M_{TR8}$ and $M_{TR9}$ [28]. The load capacitance $C_L$ includes the effect of fanout, and the interconnect capacitances.

The minimum supply voltage, $V_{DD\_MIN\_TR}$ for the traditional multiplexer is defined as the lowest voltage at which all the transistors in the two levels and the current source operate in the saturation region [29] and has been computed as

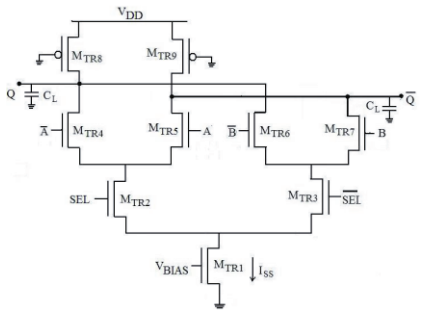



**Fig. 1.** Traditional MCML 2:1 multiplexer.

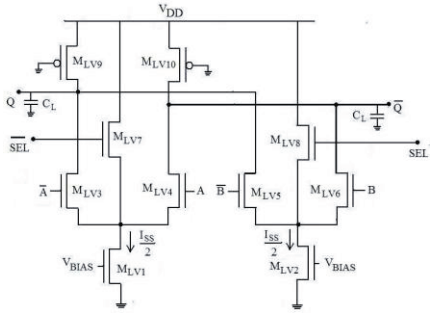



**Fig. 2.** Proposed low-voltage 2:1 multiplexer.

$$V_{DD\_MIN\_TR} = 3V_{BIAS} - 3V_{T\_TR1} + V_{T\_TR} \quad (1)$$

where $V_{T\_TR}$ is the threshold voltage of the transistors $M_{TR4,5,6,7}$, $V_{T\_TR1}$ is the threshold voltage of $M_{TR1}$, $V_{BIAS}$ is the biasing voltage of $M_{TR1}$.

# 3. Proposed Low-voltage MCML Multiplexer

The proposed low-voltage 2:1 multiplexer with differential inputs, namely SEL A and B, is shown in Fig. 2. It consists of two triple-tail cells ($M_{LV3}$, $M_{LV4}$, $M_{LV7}$) and ($M_{LV5}$, $M_{LV6}$, $M_{LV8}$) biased by separate current sources of $I_{SS}/2$ value. The transistors $M_{LV7}$ and $M_{LV8}$ are driven by the differential SEL input and are connected between the supply terminal and the common source terminal of transistor pairs $M_{LV3}$-$M_{LV4}$ and $M_{LV5}$-$M_{LV6}$ respectively. A high differential SEL voltage turns on the transistor $M_{LV8}$, and deactivates the transistor pair $M_{LV5}$-$M_{LV6}$. At the same time, the transistor $M_{LV7}$ turns off so that the transistor pair $M_{LV3}$-$M_{LV4}$ generates the output according to the differential input A. Similarly, the transistor pair $M_{LV5}$-$M_{LV6}$ gets activated for low differential SEL voltage and produces the output corresponding to the differential input B.

The minimum supply voltage, $V_{DD\_MIN\_LV}$ for the proposed multiplexer has been computed by the method outlined in [29] as

$$V_{DD\_MIN\_LV} = 2V_{BIAS} - 2V_{T\_LV1} + V_{T\_LV} \quad (2)$$

where $V_{T\_LV}$ is the threshold voltage of transistor $M_{LV3,4,5,6}$, $V_{T\_LV1}$ is the threshold voltage of $M_{LV1}$, $V_{BIAS}$ is the biasing voltage of $M_{LV1}$.

## 3.1 Static Model

The static model has been derived by modeling the load transistors $M_{LV9}$, $M_{LV10}$ by an equivalent linear resistance, $R_p$ [30]. Using the standard BSIM3v3 model, the linear resistance $R_p$ has been computed as

$$Rp = \frac{R\,\mathrm{int}}{1 - \frac{\left(R_{DSW} \cdot 1 \cdot 10^{-6}\right)/W_P}{R_{int}}} \quad (3)$$

where $R_{DSW}$ is the empirical model parameter, $W_P$ is the channel width of the load transistor and the parameter $R_{int}$ is the intrinsic resistance of the PMOS transistor in the linear region and is given as

$$R_{int} = \left[\mu_{eff,p} C_{ox} \frac{W_P}{L_P}\left(V_{DD} - \left|V_{T,p}\right|\right)\right]^{-1} \quad (4)$$

where $C_{ox}$ is the oxide capacitance per unit area. The parameters $\mu_{eff,p}$, $V_{T,p}$ and $L_P$ are the effective hole mobility, the threshold voltage and the effective channel length of the load transistor, respectively.

It may be noted that if equal aspect ratio of all transistors in the triple tail cells is considered, then the transistors $M_{LV7}$ and $M_{LV8}$ will not be able to completely switch off the transistor pair $M_{LV3}$-$M_{LV4}$ and $M_{LV5}$-$M_{LV6}$. Hence, for proper operation, the aspect ratio of transistors $M_{LV7}$, $M_{LV8}$ is made greater than other transistors' aspect ratio by a factor N. As an example if the value of differential inputs A and B is chosen such that the transistors $M_{LV3}$, $M_{LV5}$ are on while the transistors $M_{LV4}$, $M_{LV6}$ are off. Then, a high differential SEL voltage turns on the transistor $M_{LV8}$ and deactivates the transistor pair $M_{LV5}$-$M_{LV6}$. But since the transistors $M_{LV8}$ and $M_{LV5}$ have the same gate-source voltages, the currents flowing through $M_{LV5}$ ($i_{D,5}$) and $M_{LV8}$ ($i_{D,8}$) can be written as

$$i_{D,5} = \frac{I_{SS}}{2}\frac{1}{1+N}, \quad (5a)$$

$$i_{D,8} = \frac{I_{SS}}{2}\frac{N}{1+N}. \quad (5b)$$

The current through $M_{LV5}$ can be minimized by increasing factor $N$. This input condition produces minimum output voltage $V_{OL}$ as

$$\begin{aligned} V_{OL} &= V_Q - \overline{V}_Q = R_p\left[\left(i_{D,4} + i_{D,6}\right) - \left(i_{D,3} + i_{D,5}\right)\right] \\ &= -\frac{R_p I_{SS}}{2}\left(1 + \frac{1}{1+N}\right) \end{aligned} \quad (6)$$

where $i_{D,3}$, $i_{D,4}$, $i_{D,5}$, $i_{D,6}$ are the currents through transistors $M_{LV3}$, $M_{LV4}$, $M_{LV5}$, $M_{LV6}$ respectively. The differential output voltages for various input combinations are enlisted in Tab. 1. It can be observed from Tab. 1 that there are two values of both maximum output voltage $V_{OH}$ and minimum output voltage $V_{OL}$ for different input combinations. Consequently, the voltage swing, $V_{SWING1}$ for the same differential inputs (A and B) can be expressed as

| Differential inputs | | | Currents through the transistors | | | | | | Differential output $(V_Q - \bar{V}_Q)$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| **SEL** | **A** | **B** | $M_{LV3}$ | $M_{LV4}$ | $M_{LV5}$ | $M_{LV6}$ | $M_{LV7}$ | $M_{LV8}$ | **Level** | $R_P[(i_{D,4}+i_{D,6})-(i_{D,3}+i_{D,5})]$ |
| L | L | L | $I_3$ | 0 | $I_1$ | 0 | $I_2$ | 0 | $V_{OL1}$ | $-R_P\frac{I_{SS}}{2}\left(1+\frac{1}{1+N}\right)$ |
| | L | H | $I_3$ | 0 | 0 | $I_1$ | $I_2$ | 0 | $V_{OH2}$ | $R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| | H | L | 0 | $I_3$ | $I_1$ | 0 | $I_2$ | 0 | $V_{OL2}$ | $-R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| | H | H | 0 | $I_3$ | 0 | $I_1$ | $I_2$ | 0 | $V_{OH1}$ | $R_P\frac{I_{SS}}{2}\left(1+\frac{1}{1+N}\right)$ |
| H | L | L | $I_1$ | 0 | $I_3$ | 0 | 0 | $I_2$ | $V_{OL1}$ | $-R_P\frac{I_{SS}}{2}\left(1+\frac{1}{1+N}\right)$ |
| | L | H | $I_1$ | 0 | 0 | $I_3$ | 0 | $I_2$ | $V_{OL2}$ | $-R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| | H | L | 0 | $I_1$ | $I_3$ | 0 | 0 | $I_2$ | $V_{OH2}$ | $R_P\frac{I_{SS}}{2}\left(\frac{N}{1+N}\right)$ |
| | H | H | 0 | $I_1$ | 0 | $I_3$ | 0 | $I_2$ | $V_{OH1}$ | $R_P\frac{I_{SS}}{2}\left(1+\frac{1}{1+N}\right)$ |

**Tab. 1.** Differential output voltages for various input combinations. L/H= low/high differential input voltage. $I_1 = I_{SS}/2$, $I_2 = I_{SS}/2$ $(N/(1+N))$, $I_3 = I_{SS}/2$ $(1/(1+N))$.

$$V_{SWING1} = V_{OH1} - V_{OL1} = R_p I_{SS}\left(1+\frac{1}{1+N}\right) \tag{7a}$$

where $V_{OH1}$, $V_{OL1}$ are maximum output voltage and minimum output voltage respectively for the same differential inputs. The voltage swing, $V_{SWING2}$ for the different differential inputs (A and B) can be expressed as

$$V_{SWING2} = V_{OH2} - V_{OL2} = R_p I_{SS}\left(\frac{N}{1+N}\right) \tag{7b}$$

where $V_{OH2}$, $V_{OL2}$ are maximum output voltage and minimum output voltage respectively for different differential inputs.

As $V_{SWING2} < V_{SWING1}$, $V_{SWING2}$ has been considered as the worst case voltage swing, $V_{SWING}$ and has been further approximated as

$$V_{SWING} = R_p I_{SS} \text{ for large values of } N. \tag{8}$$

The small-signal voltage gain ($A_V$) and noise margin (NM) for the proposed multiplexer have been computed by the method outlined in [30] as

$$A_V = g_{m,n} R_p = \frac{V_{SWING}}{2}\sqrt{2\mu_{eff,n} C_{OX}\frac{W_N}{L_N}\frac{1}{I_{SS}}}, \tag{9}$$

$$NM = \frac{V_{SWING}}{2}\left[1-\frac{\sqrt{2}}{A_V}\right] \tag{10}$$

where $\mu_{eff,n}$, $g_{m,n}$, $W_N$ and $L_N$ are the effective electron mobility, the transconductance, the effective channel width and length of transistors $M_{LV3,4,5,6}$ respectively.

## 3.2 Transistor Sizing

In this section, an approach to size the transistors of the proposed multiplexer based on the static model is developed. For a specified value of NM and $A_V$ ( > 1.4 for MCML [31]), the voltage swing of the proposed multiplexer has been calculated using (10) as

$$V_{SWING} = \frac{2NM}{1-\frac{\sqrt{2}}{A_V}}. \tag{11}$$

It may be noted that $V_{SWING}$ should be lower than the maximum value of 2 $V_T$ so as to ensure that transistors $M_{LV3,4,5,6}$ operates in saturation region. The voltage swing obtained from (11) requires sizing of the load transistor with equivalent resistance $R_p$ (= $V_{SWING}/I_{SS}$). To this end, the equivalent resistance, $R_{P\_MIN}$, for the minimum sized PMOS transistor is first determined and then the bias current $I_{HIGH}$ for the required voltage swing is determined as

$$I_{HIGH} = \frac{V_{SWING}}{R_{P\_MIN}}. \tag{12}$$

If the bias current is higher than $I_{HIGH}$, then $R_P$ should be less than $R_{P\_MIN}$ and this is achieved by setting $L_P$ to its minimum value i.e. $L_{MIN}$ and $W_P$ which is calculated by solving (3) and (4) as

$$W_p = \frac{I_{SS}}{V_{SWING}}\cdot \frac{L_{MIN}}{\mu_{eff,p} C_{OX}\left(V_{DD}-|V_{T,p}|\right)\left\{1-\frac{R_{DSW}\cdot 10^{-6}}{L_{MIN}}\left[\mu_{eff,p} C_{OX}\left(V_{DD}-|V_{T,p}|\right)\right]\right\}}. \tag{13}$$

| Simulation Condition: $A_V$ = 4, $V_{SWING}$ = 0.4 V, $C_L$ = 50 fF, $I_{SS}$ = 100 µA | | | | | | |
|---|---|---|---|---|---|---|
| Parameter \ NMOS PMOS | | T T | F F | S S | F S | S F |
| $V_{SWING}$ (mV) | Proposed | 344 | 481 | 260 | 430 | 350 |
| | Traditional | 366 | 465 | 267 | 378 | 370 |
| $A_V$ | Proposed | 3.1 | 2.1 | 5.2 | 3.1 | 3.1 |
| | Traditional | 3.2 | 2.1 | 4.3 | 3.1 | 3.1 |
| NM (mV) | Proposed | 94.2 | 78.5 | 94.6 | 116.6 | 95.4 |
| | Traditional | 100.6 | 76.7 | 90 | 103.1 | 101.1 |
| Simulation Condition: $A_V$ = 4, $V_{SWING}$ = 0.4 V, $C_L$ = 50 fF, $I_{SS}$ = 10 µA | | | | | | |
| $V_{SWING}$ (mV) | Proposed | 410 | 498 | 265 | 420 | 415 |
| | Traditional | 342 | 519 | 294 | 443 | 407 |
| $A_V$ | Proposed | 3.8 | 1.9 | 5.5 | 2.9 | 3.7 |
| | Traditional | 2.98 | 1.81 | 4.39 | 2.67 | 2.81 |
| NM (mV) | Proposed | 130.2 | 63.6 | 98.9 | 110.6 | 129.4 |
| | Traditional | 89.8 | 56.7 | 99.6 | 104.2 | 101.1 |

**Tab. 2.** Effect of process variation on static parameters. Different design corners are denoted by T = Typical, F= Fast, S= Slow.

Similarly, if the bias current is lower than $I_{HIGH}$, then $R_P$ should be greater than $R_{P\_MIN}$ which is achieved by setting $W_P$ to its minimum value i.e. $W_{MIN}$, and $L_P$ which is calculated by solving (3) and (4) as

$$L_p = W_{MIN}\mu_{eff,p}C_{OX}\left(V_{DD} - \left|V_{T,p}\right|\right)\left(\frac{V_{SWING}}{I_{SS}} - \frac{R_{DSW}\cdot 10^{-6}}{W_{MIN}}\right). \quad (14)$$

The small-signal voltage gain ($A_V$) computed in (9) has been used to size transistors $M_{LV3,4,5,6}$. Assuming minimum channel length for the said transistors, the width has been computed as

$$W_N = \frac{2}{\mu_{eff,n}C_{OX}}\left(\frac{A_V}{V_{SWING}}\right)^2 I_{SS}L_{MIN}. \quad (15)$$

Sometimes (15) results in a value of $W_N$ smaller than the minimum channel width. This happens when the bias current is lower than the current of the minimum sized NMOS transistor, $I_{LOW}$ given as

$$I_{LOW} = \frac{1}{2}\frac{W_{MIN}}{L_{MIN}}\mu_{eff,n}C_{OX}\left(\frac{V_{SWING}}{A_V}\right)^2. \quad (16)$$

Therefore, in such cases, $W_N$ is also set to $W_{MIN}$.

The accuracy of the static model for the proposed multiplexer has been validated through SPICE simulations by using TSMC 0.18 µm CMOS process parameters. The proposed multiplexer is designed for wide range of operating conditions: voltage swing of 300 mV and 400 mV, small-signal voltage gain of 2 and 4, and the bias current ranging from 10 µA to 100 µA.

The designs were simulated and the error in simulated and theoretical values for voltage swing, small-signal voltage gain and noise margin using equations (8), (9) and (10) respectively are calculated and are plotted in Fig. 3. It may be noted that maximum error in voltage swing, small-signal voltage gain and noise margin are 16 %, 15 % and 19 % respectively.

The impact of parameter variation on the proposed low-voltage and traditional MCML multiplexer performance is studied at different design corners. The findings for various operating conditions are given in Tab. 2. It is found that the voltage swing, small-signal voltage gain, and noise margin of the proposed low-voltage multiplexer varies by a factor of 1.87, 2.94, and 2.28 respectively between the best and the worst cases. For the traditional MCML multiplexer, the voltage swing, small-signal voltage gain, and noise margin varies by a factor of 1.76, 2.42, and 1.8 respectively between the best and the worst cases. Thus, the proposed low-voltage multiplexer shows slightly higher variations than the traditional MCML multiplexer for different design corners which can be attributed to the smaller aspect ratio of transistors in the proposed low-voltage multiplexer [31].

The effect of temperature variation on proposed low-voltage and traditional MCML multiplexers performance is studied for a typical process corner. The results are shown in Tab. 3. It is found that the voltage swing, small-signal voltage gain, and noise margin of the proposed low-voltage multiplexer varies by about 0.025 %/°C, 0.17 %/°C and 0.122 %/°C respectively. For the traditional MCML multiplexer, the voltage swing, small-signal voltage gain, and noise margin varies by about 0.022 %/°C, 0.11 %/°C and 0.098 %/°C respectively. Thus, the proposed low-voltage multiplexer shows slightly higher variations than the traditional MCML multiplexer.

## 3.3 Delay Model

In this section, a delay model of the proposed multiplexer is formulated in terms of bias current and voltage swing. There are two delay parameters, namely select to Q (SEL-Q) and input to Q (A-Q or B-Q), described for a multiplexer. The SEL-Q delay is evaluated when SEL changes with constant inputs (A and B) whereas A-Q (B-Q) delay is evaluated when A (B) switches while SEL remains constant. However in practical cases, the SEL-Q delay is prominent and is therefore considered for further discussion.

| Simulation Condition: $A_V$ = 4, $V_{SWING}$ = 0.4 V, $C_L$ = 50 fF, $I_{SS}$ = 100 µA | | | | |
|---|---|---|---|---|
| Parameter \ Temp (°C) | | 0° | 70° | 125° |
| $V_{SWING}$ (mV) | Proposed | 387 | 394 | 399 |
| | Traditional | 386 | 392 | 396 |
| $A_V$ | Proposed | 3.6 | 4.0 | 4.3 |
| | Traditional | 3.58 | 3.9 | 4.1 |
| NM (mV) | Proposed | 117 | 127 | 134.8 |
| | Traditional | 116 | 124 | 130.21 |

**Tab. 3.** Effect of temperature variations on static parameters.

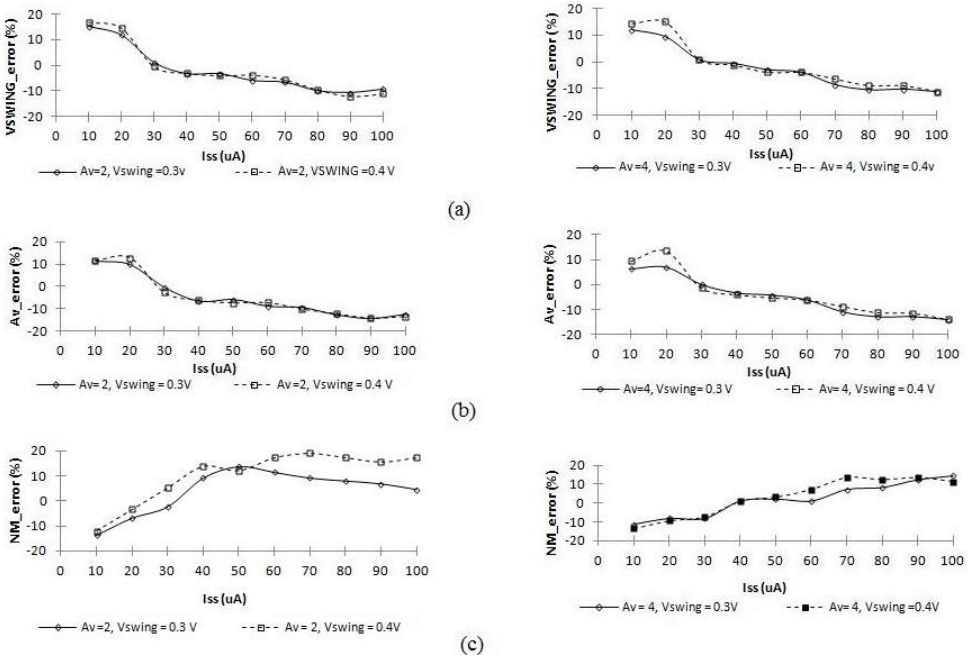



**Fig. 3.** Error in the static parameters versus $I_{SS}$ for different values of $V_{SWING}$ and $A$v, (a) $V_{SWING}$, (b) $A$v, (c) NM.

In case of a low-to-high transition on SEL input that causes output to switch by activating (deactivating) the transistor pair $M_{LV3}$-$M_{LV4}$ ($M_{LV5}$-$M_{LV6}$), the circuit reduces to a simple MCML inverter. The equivalent linear half circuit is shown in Fig. 4 where $C_{gdi}$, $C_{dbi}$ represent the gate-drain capacitance and the drain-bulk junction capacitance of the $i^{th}$ transistor. For NMOS transistors operating in saturation region, $C_{gd}$ is equal to the overlap capacitance $C_{gdo}W_n$ between the gate and the drain where $C_{gdo}$ is the drain-gate overlap capacitance per unit transistor width [30]. For the PMOS transistor operating in linear region, $C_{gd}$ is evaluated as the sum of the overlap capacitance and the intrinsic contribution associated with its channel charge [30]. The junction capacitance $C_{db}$ for the transistors has been computed as explained in [32].

The SEL-Q delay ($t_{PD\_SEL}$) of the proposed multiplexer can be expressed as

$$t_{PD\_SEL} = 0.69Rp \cdot \left(C_{db3} + C_{gd3} + C_{gd9} + C_{db9} + C_{db5} + C_{gd5} + C_L\right) \quad (17)$$

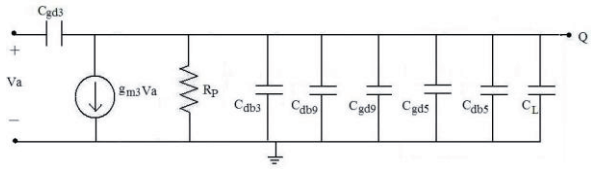



**Fig. 4.** Linear half-circuit (with low differential input A).

with

$C_{db3} = C_{db5}$, $C_{gd3} = C_{gd5}$ and $R_P = \frac{V_{SWING}}{I_{SS}}$, (17) can be rewritten as

$$t_{PD\_SEL} = 0.69\frac{V_{SWING}}{I_{SS}} \cdot \left(2C_{db3} + 2C_{gd3} + C_{gd9} + C_{db9} + C_L\right). \quad (18)$$

The capacitances may be expressed in terms of bias current and voltage swing as

$$C_{xy} = \frac{a_{xy}}{(V_{SWING})^2} I_{SS} + b_{xy} \frac{V_{SWING}}{I_{SS}} + c_{xy}$$

where $C_{xy}$ is the capacitance between the terminals x and y and $a_{xy}$, $b_{xy}$, $c_{xy}$ are the associated coefficients. Using (14) and (15), various capacitances in (18) for $I_{SS}$ ranging from $I_{LOW}$ to $I_{HIGH}$ have been expressed as

$$C_{gd3} = C_{gdo} W_3 = 2A_V^2 C_{gdo} \frac{L_{MIN}}{\mu_{eff,n} c_{OX}} \frac{I_{SS}}{(V_{SWING})^2}, \quad (20)$$

$$C_{db3} = W_3 \left(K_{jn} C_{jn} L_{dn} + 2K_{jswn} C_{jswn}\right) + 2K_{jswn} C_{jswn} L_{dn} \quad (21)$$

$$= 2A_V^2 \frac{L_{MIN}}{\mu_{eff,n} C_{OX}} \left(K_{jn} C_{jn} L_{dn} + 2K_{jswn} C_{jswn}\right) \frac{I_{SS}}{(V_{SWING})^2} + 2K_{jswn} C_{jswn} L_{dn} \quad (22)$$

where $C_{jn}$, $C_{jswn}$ are the zero-bias junction capacitance per unit area and zero-bias sidewall capacitance per unit parameter respectively. The coefficients $K_{jn}$, $K_{jswn}$ are the voltage equivalence factor for the junction and the sidewall capacitances of the NMOS transistor respectively [32]. Parameter $L_{dn}$ is extrapolated from design rules [22].

$$C_{gd9} = C_{gdo} W_{MIN} + \frac{3}{4} A_{bulk,max} W_{MIN} L_P C_{OX} \quad (23)$$

$$= C_{gdo} W_{MIN} + \frac{3}{4} A_{bulk,max} W_{MIN} C_{OX} \cdot \left\{ \mu_{eff,p} C_{OX} W_{MIN} \left(V_{DD} - |V_{T,p}|\right) \left[ \frac{V_{SWING}}{I_{SS}} - \frac{R_{DSW} 10^{-6}}{W_{MIN}} \right] \right\} \quad (24)$$

where $A_{bulk,max}$ is a parameter defined in BSIM3v3 model [28].

$$C_{db9} = W_{MIN} \left(K_{jp} C_{jp} L_{dp} + 2K_{jswp} C_{jswp}\right) + 2K_{jswp} C_{jswp} L_{dp} \quad (25)$$

where $C_{jp}$, $C_{jswp}$ are the zero-bias junction capacitance per unit area and zero-bias sidewall capacitance per unit parameter respectively. The coefficients $K_{jp}$, $K_{jswp}$ are the voltage equivalence factor for the junction and the sidewall capacitances of the PMOS transistor respectively [32]. Parameter $L_{db}$ is extrapolated from design rules [22].

The coefficients $a_{xy}$, $b_{xy}$ and $c_{xy}$ of all the capacitances in (18) are summarized in Tab. 4. Using equations (20) – (25), equation (18) can be written as

$$t_{PD\_SEL} = 0.69 V_{SWING} \left( \frac{a}{V_{SWING}^2} + b \frac{V_{SWING}}{I_{SS}^2} + \frac{c + C_L}{I_{SS}} \right) \quad (26)$$

where

$$a = 2a_{db3} + 2a_{gd3}, \quad (27a)$$

$$b = b_{gd9}, \quad (27b)$$

$$c = 2c_{db3} + 2c_{gd9} + c_{db9}. \quad (27c)$$

The delay model can also be used for $I_{SS}$ value outside the range [$I_{LOW}$, $I_{HIGH}$]. This is because for $I_{SS} > I_{HIGH}$ , the capacitance coefficients of PMOS transistor in (26) differ as explained in Section 3.2. But, since for high values of $I_{SS}$, the capacitive contribution of PMOS transistor is negligible, therefore (26) can predict the delay. Similarly, for $I_{SS} < I_{LOW}$, the capacitance coefficients of NMOS transistor in (26) differs. But, since for low values of $I_{SS}$, the delay majorly depends on the capacitances of PMOS transistor. So, the expression in (26) can estimate the delay of the proposed multiplexer.

The accuracy of the delay model for the proposed multiplexer has been validated through SPICE simulations by using TSMC 0.18 µm CMOS process parameters. The proposed multiplexer is designed for wide range of operating conditions: voltage swing of 300 mV and 400 mV, small-signal voltage gain of 2 and 4, bias current ranging from 10 µA to 100 µA, and load capacitance of 0 fF, 10 fF, 100 fF and 1 pF. It is found that there is a close agreement between the simulated and the predicted delay for all the operating conditions. The simulated and the predicted delay in particular for $V_{SWING}$ = 400 mV, $A_V$= 4 and with different load capacitances are plotted in Fig. 5.

The impact of parameter variation on proposed low-voltage and traditional multiplexers delay is studied at different design corners. The findings for various operating conditions are given in Tab. 5. It is found that the propagation delay of the proposed low-voltage multiplexer varies by a factor of 1.89 between the best and the worst cases. For the traditional MCML multiplexer, the delay varies by a factor of 1.85 between the best and the worst cases. Thus, the proposed low-voltage multiplexer shows slightly higher variation than the traditional MCML multiplexer in delay for different design corners. The process variations are more prevalent in the designs with smaller aspect ratio [31] and the results for proposed low-voltage multiplexer conform to this fact.

The effect of temperature variation on proposed low-voltage and traditional MCML multiplexers delay is studied for a typical process corner. The results are shown in Tab. 6. It is found that delay of the proposed low-voltage multiplexer varies by about 1.2 %/°C. For the traditional MCML multiplexer the delay shows a variation of 1 %/°C. Thus, the proposed low-voltage multiplexer shows slightly higher variations than the traditional MCML multiplexer.

| NMOS coefficients | |
|---|---|
| $a_{db3}$ | $\frac{2A_V^2 L_{MIN}}{\mu_{eff,n} C_{OX}}\left(K_{jn} C_{jn} L_{dn} + 2K_{jswn} C_{jswn}\right)$ |
| $a_{gd3}$ | $2A_V^2 C_{gdo} \frac{L_{MIN}}{\mu_{eff,n} C_{OX}}$ |
| $c_{db3}$ | $2K_{jswn} C_{jswn} L_{dn}$ |
| $b_{db3}$, $b_{gd3}$, $c_{gd3}$ | 0 |
| **PMOS coefficients** | |
| $b_{gd9}$ | $\frac{3}{4} A_{bulk,max} \mu_{eff,p} C_{OX}^2 W_{MIN}^2 \left(V_{DD} - \lvert V_{T,p} \rvert\right)$ |
| $c_{gd9}$ | $C_{gdo} W_{MIN} - \frac{3}{4} A_{bulk,max} \mu_{eff,p} C_{OX}^2 W_{MIN} \left(V_{DD} - \lvert V_{T,p} \rvert\right) R_{DSW} 10^{-6}$ |
| $c_{db9}$ | $K_{jp} C_{jp} L_{dp} W_{MIN} + 2K_{jswp} C_{jswp} \left(L_{dp} + W_{MIN}\right)$ |
| $a_{gd9}$, $a_{db9}$, $b_{db9}$ | 0 |

**Tab. 4.** The capacitance coefficients for the proposed multiplexer. The symbols have their usual meanings.

| Simulation Condition: $A_V$ = 4, $V_{SWING}$ = 0.4 V, $C_L$ = 50 fF, $I_{SS}$ = 100 µA | | | | | | |
|---|---|---|---|---|---|---|
| Parameter | NMOS / PMOS | T / T | F / F | S / S | F / S | S / F |
| $t_{PD}$ (ps) | Proposed | 265 | 237 | 448 | 255 | 262 |
| | Traditional | 553 | 515 | 954 | 527 | 550 |
| Simulation Condition: $A_V$ = 4, $V_{SWING}$ = 0.4 V, $C_L$ = 50 fF, $I_{SS}$ = 10 µA | | | | | | |
| $t_{PD}$ (ns) | Proposed | 2.4 | 1.7 | 3.2 | 2.1 | 2.3 |
| | Traditional | 3.7 | 3.2 | 4.6 | 3.5 | 3.6 |

**Tab. 5.** Effect of process variation on delay.

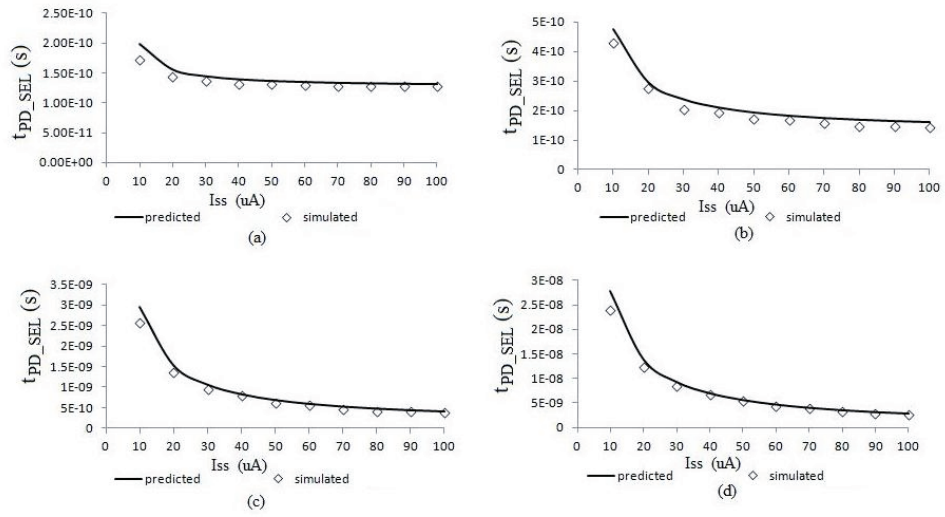



**Fig. 5.** Simulated and the predicted delay of the proposed low-voltage multiplexer versus $I_{SS}$ with NM =130 mV, $A_V$= 4 for different $C_L$ values: (a) 0 fF, (b) 10 fF, (c) 100 fF, (d) 1 pF.

| Simulation Condition: $A_V = 4$, $V_{SWING} = 0.4$ V, $C_L = 50$ fF, $I_{SS} = 100$ µA | | | | |
|---|---|---|---|---|
| Parameter \ Temp | | 0° | 70° | 125° |
| $t_{PD}$ (ps) | Proposed | 151 | 280 | 387 |
| | Traditional | 339 | 590 | 762 |

**Tab. 6.** Effect of temperature variation on delay.

# 4. Design Cases

In the previous section, the proposed multiplexer has been modeled and various parameters are expressed as a function of bias current and voltage swing. In practice, the voltage swing is set on the basis of the specified noise margin while the bias current is chosen according to power-delay considerations. Therefore, the proposed low-voltage multiplexer for high-speed, power-efficient, and low-power cases is discussed.

## 4.1 High-Speed Design

A high-speed design requires bias current that results in minimum delay. The delay in (26) decreases with the increasing $I_{SS}$ and tends to an asymptotic minimum value of $0.69\cdot(a / V_{SWING})$ for $I_{SS} \to \infty$. A substantial improvement in delay with increasing bias current may be achieved if condition

$$\frac{a}{V_{SWING}^2} \geq b\frac{V_{SWING}}{I_{SS}^2} + \frac{c + C_L}{I_{SS}} \tag{28}$$

is satisfied. However, high value of bias current results in large transistor sizes. Therefore, the bias current should be set to such a value after which the improvement in speed is not significant. If equality sign in (28) is considered then the delay is close to its minimum value and the use of high bias current is avoided. Therefore, this assumption leads to a bias current ($I_{SS\_HS}$) and delay ($t_{PD\_MIN}$) as

$$I_{SS\_HS} = \frac{c + C_L}{2a} V_{SWING}^2 \left( 1 + \sqrt{1 + 4\frac{ab}{\left(c + C_L\right)^2} \frac{1}{V_{SWING}}} \right), \tag{29}$$

$$t_{PD\_MIN} = 2 \cdot 0.69 \frac{a}{V_{SWING}}. \tag{30}$$

The proposed high-speed multiplexer designed with a noise margin of 130 mV, small-signal gain of 4, and load capacitance of 50 fF, gives $I_{SS\_HS}$ as 112 µA. A delay of 254 ps and 224 ps are obtained from (30) and simulations respectively. On the contrary, a traditional high-speed multiplexer designed using the method outlined in [28] for the same specifications results in a delay of 528 ps. This indicates that the proposed multiplexer can achieve much higher speed than the traditional one.

## 4.2 Power Efficient Design

A power efficient design requires bias current that results in minimum power-delay product (PDP). The power is calculated as the product of $V_{DD}$ and $I_{SS}$. So, the PDP of the proposed multiplexer may be expressed as:

$$\text{PDP} = 0.69 V_{DD} V_{SWING} \left( \frac{a}{V_{SWING}^2} I_{SS} + b\frac{V_{SWING}}{I_{SS}} + c + C_L \right). \tag{31}$$

Therefore, the current $I_{SS\_PDP}$ for minimum PDP may be given as

$$I_{SS\_PDP} = \sqrt{\frac{b}{a}} \left(V_{SWING}\right)^{\frac{3}{2}}. \tag{32}$$

Accordingly, the minimum PDP results to

$$\text{PDP} = 0.69 V_{DD} V_{SWING} \left( \frac{2\sqrt{ab}}{\sqrt{V_{SWING}}} + c + C_L \right). \tag{33}$$

The proposed power-efficient multiplexer designed with a noise margin of 130 mV, small signal gain of 4, and load capacitance of 50 fF, gives $I_{SS\_PSP}$ as 4.5 µA. A PDP value of 19 fJ has been obtained for the proposed multiplexer. On the other hand, a traditional power-efficient multiplexer designed using the method outlined in [28] for the same specifications results in a PDP value of 13 fJ. The result signifies that the proposed multiplexer results in higher PDP values than the traditional one.

## 4.3 Low-Power Design

In low-power designs, the bias current $I_{SS}$ is set to low values so that the term

$$b\frac{V_{SWING}}{I_{SS}^2}$$

is dominant in (26). Hence, the delay reduces to

$$t_{PD\_SEL} = 0.69b \left( \frac{V_{SWING}}{I_{SS}} \right)^2. \tag{34}$$

The proposed low-power multiplexer designed with a noise margin of 130 mV, small signal gain of 4, load capacitance of 5 fF, and with value of $I_{SS}$ as 2 µA gives a power consumption of 2.2 µW while the traditional low-power multiplexer designed using the method outlined in [28] for the same specifications results in power consumption of 2.8 µW.

# 5. Conclusions

A new low-voltage MCML multiplexer based on the triple-tail cell concept is proposed. Its static parameters are analytically modeled and are used to develop a design approach for the proposed low-voltage MCML multi-

plexer. The delay is formulated as a function of the bias current and the voltage swing and is traded off with power consumption for high-speed, power-efficient, and low-power design cases. An improvement in performance is obtained for the proposed low-voltage multiplexer in comparison to traditional MCML multiplexer for high-speed and low-power design cases.

# References


[1] JANTZI, S., MARTIN, K., SEDRA, A. Quadrature bandpass ΣΔ modulator for digital radio. *IEEE Journal of Solid-State Circuits*, 1997, vol. 32, no. 12, p. 1935 - 1949.

[2] LUSCHAS, S., SCHREIER, R., LEE, H. S. Radio frequency digital-to-analog converter. *IEEE Journal of Solid-State Circuits*, 2004, vol. 39, no. 9, p. 1462 - 1467.

[3] KUP, B., DIJKMANS, E., NAUS, P., SNEEP, J. A bit-stream digital-to-analog converter with 18-b resolution. *IEEE Journal of Solid-State Circuits*, 1991, vol. 26, no. 12, p. 1757 - 1763.

[4] TAKAAMOTO, T., HARAJIRI, S., SAWADA, M., KOBAYASHI, O., GOTOH, K. A bonded–SOI-wafer CMOS 16-bit 50-KSPS delta-sigma ADC. In *Proceedings of IEEE Custom Integrated Circuit Conference*. San Diego (CA, USA), 1991, p. 18.1.1–18.1.4.

[5] WESTE, N., ESHRAGHIAN, K., *Principles of CMOS VLSI Design: A System Perspective*. Boston (USA): Addison-Wesley, 1993.

[6] ALLSTOT, D., CHEE, S., KIAEI, S., SHRISTAWA, M. Folded source-coupled logic vs. CMOS static logic for low-noise mixed-signal ICs. *IEEE Transactions on Circuits and Systems – I*, 1993, vol. 40, no. 9, p. 553 - 563.

[7] CHOY, C., CHAN, C., KU, M., POVAZANEC, J. Design procedure of low noise high-speed adaptive output drivers. In *Proceedings of the IEEE International Symposium on Circuits and Systems*. Hong Kong (China), 1997, p. 1796 - 1799.

[8] KIAEI, S., ALLSTOT, D. Low-noise logic for mixed-mode VLSI circuits. *Microelectronics Journal*, 1992; vol. 23, no. 2, p. 103 - 114.

[9] SAEZ, R., KAYAL, M., DECLERQ, M., SCHNEIDER, M. Digital circuit techniques for mixed analog/digital circuits applications. In *Proceedings of 3rd International Conference on Electronics, Circuits, and Systems*. Rodos (Greece), 1996, p. 956 -959.

[10] NG, H., ALLSTOT, D. CMOS current steering logic for low-voltage mixed-signal integrated circuits. *IEEE Transactions on VLSI Systems*, 1997, vol. 5, no. 3, p. 301 - 308.

[11] KUNDAN, J., HASAN, S. Enhanced folded source-coupled logic technique for low-voltage mixed-signal integrated circuits. *IEEE Transactions on Circuits and Systems – II*, 2000, vol. 47, no. 8, p. 810 - 817.

[12] YAMASHINA, M., YAMADA, H. An MOS current code logic (MCML) circuit for low power sub-GHz processors. *IEICE Transactions on Electronics*, 1992, vol. E75-C, no. 10, p. 1181 -1187.

[13] BRUMA, S. Impact of on-chip process variations on MCML performance. In *Proceedings of IEEE Conference on Systems-on-Chip*. Portland (OR, USA), 2003, p. 135 - 140.

[14] MUSICER, J. M., RABAEY, J. MOS current mode logic for low power, low noise, CORDIC computation in mixed-signal environments. In *Proceedings of the International Symposium of Low Power Electronics and Design*. Rapallo (Italy), 2000, p. 102 - 107.

[15] MIZUNO, M., YAMASHINA, M., FURUTA, K., IGURA, H., ABIKO, H., OKABE, K., ONO, A., YAMADA, H. A GHz MOS adaptive pipeline technique using MOS current mode logic. *IEEE Journal of Solid-State Circuits*, 1996, vol. 31, no. 6, p. 784 - 791.

[16] GREEN, M. M., SINGH, U. Design of CMOS CML circuits for high speed broadband communications. In *Proceedings of the International Symposium on Circuits and Systems*. Bangkok (Thailand), 2003, vol. 2, p. 204 - 207.

[17] SHIN, J. K.; YOO, T. W., LEE M. Design of half-rate linear phase detector using MOS current mode logic gates for 10-Gb/s clock and data recovery circuit. In *Proceedings of IEEE International Conference on Advanced Communication Technology*. Phoenix Park (Korea), 2005, p. 205 - 210.

[18] YOUNGKYUN, J., SUNGYOUNG, J., JIN, L. A. A CMOS impulse generator for UWB wireless communication systems. In *Proceedings of the International Symposium on Circuits and Systems*. Vancouver (Canada), 2004, p. 129 - 132.

[19] TANABE, A., UMETANI, M., FUJIWARA, I., OGURA, T., KATAOKA, K., OKIARA, M., SAKURABE, H., ENDOH, T., MASUKA, F. 0.18 μm CMOS 10-Gb/s multiplexer/demultiplexer ICs using current mode logic with tolerance to threshold voltage fluctuation. *IEEE Journal of Solid-State Circuits*, 2001, vol. 36, no. 6, p. 988 - 996.

[20] ALIOTO, M., MITA, R., PALUMBO, G. Design of high-speed power efficient MOS current mode logic frequency dividers. *IEEE Transactions on Circuits and Systems – II: Express Briefs*; 2006, vol. 53, no. 11, p. 1165 - 1169.

[21] YAUN, F. *CMOS Current-mode Circuits for Data Communication*. New York (USA): Springer, 2007.

[22] ALIOTO, M., PALUMBO, G. *Model and Design of Bipolar and MOS Current-Mode logic: CML, ECL and SCL Digital Circuits*. Dordrecht (The Netherlands): Springer, 2005.

[23] KIMURA, K. Circuit design techniques for very low-voltage analog functional blocks using triple-tail cells. *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, 1995, vol. 42, p. 873 - 885.

[24] MATSUMOTO, F., NOGUCHI, Y. Linear bipolar OTAs based on a triple-tail cell employing exponential circuits. *IEEE Transactions on Circuits and Systems – II: Express Briefs*, 2004, vol. 51, no. 12, p. 670 - 674.

[25] ALIOTO, M., MITA, R., PALUMBO, G. Performance evaluation of the low-voltage CML D-latch topology. *Integration, the VLSI Journal*, 2003, vol. 36, no. 4, p. 191 - 209.

[26] KIMURA, K. *Analog Multiplier Using Multi Tail Cell.* United States Patent no. 5,986,494, 1999.

[27] KIMURA, K., *Transconductance-Variable Analog Multiplier using Triple - Tail Cells*. United states Patent no. 5,617,052, 1997.

[28] ALIOTO, M., PALUMBO, G. Power-delay optimization of D-latch/MUX source coupled logic gates. *International Journal of Circuit Theory and Applications*, 2005, vol. 33, n. 1, p. 65 - 85.

[29] HASSAN, H., ANIS, M., ELMASRY, M. Analysis and design of low-power multi-threshold MCML. In *Proceedings of the IEEE International Conference on System-on-chip*. 2004, p. 25 - 29.

[30] ALIOTO, M., PALUMBO, G., PENNISI, S. Modeling of source-coupled logic gates. *International Journal of Circuit Theory and Applications*, 2002, vol. 30, no. 4, p. 459 - 477.

[31] HASSAN, H., ANIS, M., ELMASRY, M. MOS current mode circuits: analysis, design, and variability. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2005, vol. 13, no. 8, p. 885 - 898.

[32] RABAEY, J. *Digital Integrated Circuits (A Design Perspective),* 2nd ed. Englewood Cliffs (NJ, USA): Prentice Hall, 2003.

## About Authors…

**Kirti GUPTA** received B.Tech. in Electronics & Communication Engineering from Indira Gandhi Institute of Technology, Delhi in 2002, M.Tech. in Information Technology from School of Information Technology in 2006. She held the positions of Lecturer in ECE Department at Bharati Vidyapeeth's College of Engineering from 2002 to 2007 and Assistant Professor in ECE at Bharati Vidyapeeth's College of Engineering from 2007 to 2009. She is currently working as Research Scholar in ECE Department, Delhi Technological University from 2009. Her teaching and research interests are in digital integrated circuits, and VLSI design.

**Neeta PANDEY** received her M. E. in Microelectronics from Birla Institute of Technology and Sciences, Pilani and Ph.D. from Guru Gobind Singh Indraprastha University Delhi. She has served in Central Electronics Engineering Research Institute, Pilani, Indian Institute of Technology, Delhi, Priyadarshini College of Computer Science, Noida and Bharati Vidyapeeth's College of Engineering, Delhi in various capacities. At present, she is Assistant Professor in ECE Department, Delhi Technological University. A life member of ISTE, and member of IEEE, USA, she has published papers in international, national journals of repute and conferences. Her research interests are in analog and digital VLSI Design

**Maneesha GUPTA** received B.E. in Electronics & Communication Engineering from Government Engineering College, Jabalpur in 1981, M.E. in Electronics & Communication Engineering from Government Engineering College, Jabalpur in 1983, and Ph.D. in Electronics Engg (Analysis, Synthesis & Applications of Switched Capacitor Circuits) from Indian Institute of Technology, Delhi in 1990. Dr. Gupta held the positions of Lecturer in Electronics & Communication Engineering Department at Government Engineering College, Jabalpur from 1981 to 1982, Kota Engg. College, Kota from 1986 to 1988, YMCA Institute of Engg., Faridabad in 1998 and Netaji Subhash Institute of Technology, New Delhi from 1998 to 2000. She worked as Assistant Professor in Electronics & Communication Engineering (ECE) Department of the Netaji Subhash Institute of Technology, New Delhi from 2000 to 2008. She is currently working as Professor in Electronics & Communication Engineering (ECE) Department of the Netaji Subhash Institute of Technology, New Delhi. Her teaching and research interests are Switched Capacitors Circuits and Analog Signal processing. She has co-authored over 20 research papers in the above areas in various international/national journals and conferences. She got best paper award for her paper in IETE journal of Education in 2001.

# NGFICA based Digitization of Historic Inscription Images

Indu Sreedevi*, Rishi Pandey*, N. Jayanthi* Geetanjali Bhola* and Santanu Chaudhury†
* Electronics and Communication Engg. Deptt, Delhi College of Engineering
Email:s.indu@rediffmail.com
† Electrical Engineering Department,IIT Delhi, India
Email: schaudhury@gmail.com

***Abstract*—This paper addresses the problems encountered during digitization and preservation of inscriptions such as perspective distortion and minimal distinction between foreground and background. In general inscriptions neither possess standard size and shape nor colour difference between the foreground and background. Hence the existing methods like variance based extraction and Fast-ICA based analysis fail to extract text from these inscription images. Natural gradient Flexible ICA (NGFICA) is a suitable method for separating signals from a mixture of highly correlated signals, as it minimizes the dependency among the signals by considering the slope of the signal at each point. We propose an NGFICA based enhancement of inscription images. The proposed method improves word and character recognition accuracies of the OCR system by 65.3% (from 10.1% to 75.4%) and 54.3% (from 32.4% to 86.7%) respectively.**

***Key Words*** :NGFICA, Text and non-Text region, Hampi Images

## I. Introduction

A significant amount of research has been carried out in the direction of reading inscriptions from monuments around the world. Several methods have been proposed for detection of text, localization and extraction of text from images of inscriptions [1] [2]. But, the problem of text extraction intensifies when the difference in the text (foreground) and the background is very marginal or the background is textured or the background and foreground are similar. Such is the case of camera-held images of inscriptions at the sites of historical monuments. Figure 1 shows an image of inscription found in world heritage site "Hampi". These inscriptions are generally found engraved into / projected out from, stone or other durable materials. However, due to effects of uncontrolled illuminations, wrapping, multi-lingual text, minimal difference between foreground and background images, the distortion due to perspective projection as well as the complexity of image background, extracting text from these images is a challenging problem.

The commercially available OCR's (Optical Character Recognition) have very poor recognition accuracy of images of the inscriptions on monuments. The images of English inscriptions from the monuments were passed through the commercial OCR for text extraction but the OCR failed to recognize these images. These images can be recognized by OCR only after proper enhancement. FastICA [3] based enhancement method has given good results for inscription images with a reasonable colour difference between text and the background. Most of the ancient inscriptions do not have such reasonable colour distinction between the two regions. Therefore to digitize such inscriptions we have to enhance the difference between the two regions. This paper proposes a method to enhance the minimal difference between text and non-text regions of such inscription images.

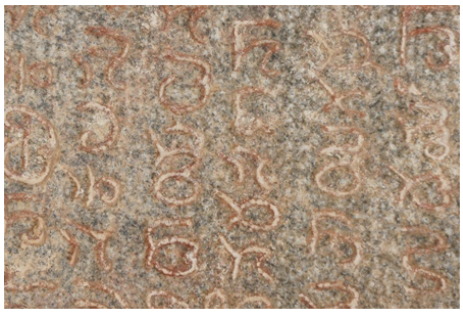


Fig. 1. Inscription found at Hampi

Natural Gradient based Flexible ICA (NGFICA) has been extensively used in separating highly correlated signals [4] as it minimizes dependency among the different signals present in the source signal using gradient descent optimization approach. For minimizing the dependency among the foreground and background of historical inscription images, we used NGFICA for obtaining independent components of the images. This paper presents a novel enhancement technique to separate the text part of the inscription image by processing NGFICA output of inscription image.

## II. Related Work

Text Extraction from document images has been of interest for the research community over a decade, but there has been very few work done in digitizing inscription images of historical monuments. High contrast edges between text and background is obtained using the red color component in the approach by Agnihotri et al. [5]. In [6], the "uniform color" blocks within the high contrast video frames are selected to correctly extract text regions. Kim et al. [7] used 64 clustered color channels for text detection where cluster colors are based

on Euclidean distance in the RGB space. The method based on variance by M. Babu et al [8] makes use of the variance in the text and non-text regions. The variance is high at the text edges and vice versa.Variance method to extract the text as in [8] did not prove successful due to blurred edges of text and minimum distinction between text and non text region. The text in inscription images does not consist of a uniform color and there is low contrast between text and background thus making the use of [6] unsuitable. Simple edge-based approaches [9] are also considered useful to identify regions with high edge density and strength. This method performs well if there is no complex background but the inscription images have complex background thus these methods cannot be used directly.

The authors of [10], estimate the intensity of non-text region (background) and do binarization in comparison with a threshold intensity. Laplacian of Gaussian filters in Sobolev space using different $\alpha$ factor for different images are used in [11] for enhancement of text images. Curvelet transform is proposed in [12] for denoising the degraded historical documents. Adaptive binarization technique for Palm Leaf Manusripts proposed in [13] , where authors used Wiener filter for noise removal and contrast adaptive binarization for segmentation of text from the back ground. [14] Proposes a wavelet based enhancement / smearing algorithm for the removal interfering strokes in archiving handwritten document images. In [15] authors proposed a hybrid approach includes both local and global thresholding technique for cleaning background noise from the ancient documents . The results in [15] shows that enhancement has been achieved but cannot be read by OCR. The above said methods were based on binarization, text extraction using variance or edge detection based methods. These methods depend upon pixel's threshold value based on difference between foreground and background part.

Garain et al [3] describe how to enhance image using FastICA algorithm which results in three independent components or layers which correspond to the contribution of text in them. The method is an enhancement method, which however is unable to enhance inscription images which are weak or highly spatially correlated sources. More recently,its convergence has been shown to slow down or even fail in the presence of saddle points, particularly for short block sizes [16]. More over it is proved in [17] that Fast ICA fails in seperating the sources for a week or closed sources.

In case of unclear and complex archaeological inscription images, there is no sharp distinction between foreground and background. Natural gradient based independent component analysis learning algorithm with flexible nonlinearity as described in [18] [19] gives better results than other algorithms as it is more efficient in minimizing dependence among correlated signals. In the proposed method we used NGFICA for minimizing the dependency between foreground, middle layer and background of such inscription images and further the characters are retrieved from the foreground.

## III. Motivation

Many of the inscriptions are couched in extravagant language, but when the information gained from inscriptions can be corroborated with information from other sources such as still existing monuments or ruins, inscriptions provide insight into world's dynastic history that otherwise lacks contemporary historical records. Digital archiving of these images are necessary for conservation and accessibility. The major challenges of digitization of such images are blurred edges of the text and minimum distinction between text and non text part. NGFICA algorithm deals with Gaussian, sub Gaussian and super Gaussian source signals as is the case with the said inscription images[18]. We are proposing a novel method to enhance degraded inscription images using NGFICA in this paper.

### A. Methodology

We separated the text (foreground), non-text (background) and noise of an image as three different components using NGFICA as it minimizes the dependency among the components. We further refined all ICA ouputs using morphological operation. The ICA output image with average threshold farthest from average threshold of original image gave good results.

1) Finding the independent components:
The images of inscriptions were complex because of the high correlation of foreground pixels with the background pixels. This merging of pixels deteriorates the clarity of the inscription images. The noise from the background due to illumination, shadows etc. added to the problem of clarity of regions(text, non-text). So, we performed Gaussian smoothing of the colored image using a 5x5 kernel. This helped to remove small scale noise and irrelevant image details. Then R, G, B components of the smoothed image were extracted. Three independent components of the colored image were obtained by performing NGFICA on the extracted R, G, B components. For reference purpose these independent components were named as text layer, non-text layer and mixed layer on the basis of their contribution to the text. The NGFICA [20] can be explained in the simplest possible way as follows. An image can be considered as a mixture of foreground, background and a common part for both. Let the mixing model be

$$X = AY \tag{1}$$

where $X[n \times T]$ be the original image
$Y[n \times T]$ be the unknown mutually independent portion of the image and $A[n \times n]$ be the mixing matrix. For a 3 channel image, Equation.1 can be written as
The de-mixing model is defined as

$$\begin{bmatrix} x_r \\ x_g \\ x_b \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ g_1 & g_2 & g_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \times \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} \tag{2}$$

$$Z = WX \tag{3}$$

where $Z[n \times T]$ is the separated sources and $W[n \times n]$ be the de-mixing matrix

$$\nabla L(W) = W^{-T} - E[g(Y)X^T] \tag{4}$$

where g(Y)is given by Equation.5

$$g(Y) = -\frac{d}{dY}(logP(Y)) \tag{5}$$

The randomized gradient $\nabla$L(W) of formula given by Equation.4 expresses the steepest drop direction of the Euclidean space cost function L(W). The natural gradient $\bar{\nabla}L(W)$ is the steepest drop direction in Riemann space of parameters W. The natural gradient $\bar{\nabla}L(W)$ can be calculated by modifying the random gradient, which is obtained by multiplying $W^TW$ in the random gradient as given by Equation-6.Thus NGFICA algorithm gives faster convergence and better performance. Faster convergence is due to the fact that decorrelation is performed together with separation and the better performance is due to the non linear function g(Y) controlled by the Gaussian exponent .

$$\bar{\nabla}L(W) = (I - E[g(Y)Y^T])W \tag{6}$$

To minimize dependency among output components we have to minimize cost function L (W)

As explained in [21] gradient adaptation is a useful technique for adjusting a set of parameters to minimize a cost function. The natural gradient is based on differential geometry and employs knowledge of the Riemannian structure of the parameter space to adjust the gradient search direction. Unlike Newton's method, natural gradient adaptation does not assume a locally-quadratic cost function. Moreover, for maximum likelihood estimation tasks, natural gradient adaptation is asymptotically Fisher-efficient. The three independent components of NGFICA are shown in Figure.3

2) Character Extraction from foreground

NGFICA output image can be considered as a foreground, back ground and noise image. We compared the average threshold of each of these output images with that of original image. The one which is farthest from original image is identified as the foreground as it had only text. The image shown in Figure.3 (b) is identified as the foreground, and the image is further enhanced using Sobel edge detection and then dilation using disc shaped structuring element to retrieve the characters.

## IV. Results and Discussion

The dataset for validating the proposed method was prepared by gathering images of inscriptions belonging to historical monuments (India Gate, Delhi) , heritage sites (Hampi, Karnataka), ancient temples (Vishnu temple, Tamil Nadu) etc. Such inscriptions are common at almost every monument and normally found engraved into / projected out from, stone or other durable materials. Some images were clicked manually using a 10 megapixel camera and some others were taken from Internet. The images demonstrated several processing difficulties like uneven illumination, wrapping, perspective distortion, multi-lingual text, text with foreground and background images, etc. The images of India Gate (English) without enhancement were tested on web based OCR [22] and the results are shown in Table.II. The enhanced outputs of India Gate images using the proposed method is shown in Figure.9 and 10. We have also compared the proposed method with Fast ICA based enhancement [3] and results are shown in Table.I in Figure.4. Other results of Hampi inscription images are shown in Figures. 5 and 11

The proposed method also worked equally well for other languages too. The results are shown in Figures.6, 7 and 8. We have tested the method on 650 images of which 550 were English word images which were passed through OCR and word accuracy of $75.4\%$ and character accuracy of $86.7\%$ was achieved. The remaining 100 images of different languages gave very good results. Natural gradient algorithm not only deals with symmetric distribution of the signal but also can

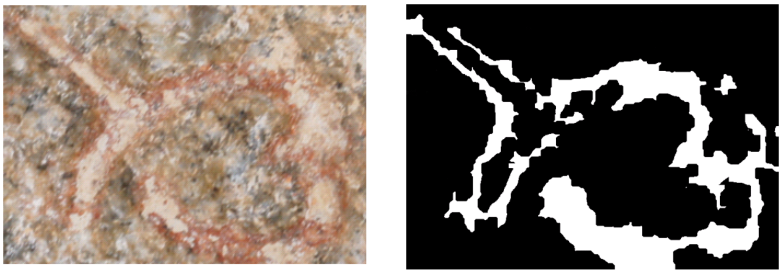
Fig. 6. Other language results

(a) (b)

Fig. 7. Other Language 1 (a) Source (b) result

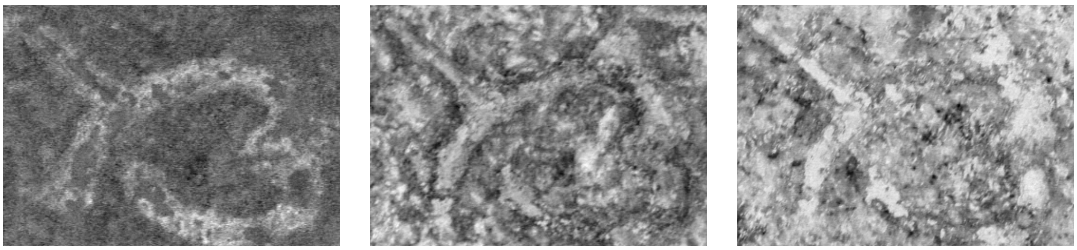
(a) (b)

Fig. 8. Other Language 2 (a) Source (b) result



(a) (b)

Fig. 2. (a) Source image (b) Final Enhanced Image



(a) (b) (c)

Fig. 3. (a) (b)and (c) NGFICA output images

(a) (b) (c)

Fig. 4. (a) Original image (b) Output image of proposed method and (c) Output after Fast ICA based enhancement

TABLE I
COMPARISON WITH FASTICA METHOD

| Input Image | Number | Fast ICA Accuracy in % | Proposed method Accuracy in % |
|---|---|---|---|
| Words | 550 | 0. 9 % | 75.4 % |
| Character | 2578 | 1 % | 86.7 % |

TABLE II
OCR ACCURACY BEFORE AND AFTER ENHANCEMENT

| Input Image | Number | Rec. by OCR Before Enhancement | Accuracy in % Before Enhancement | Rec. by OCR After Enhancement | Accuracy in % After Enhancement |
|---|---|---|---|---|---|
| Words | 550 | 56 | 10.1 % | 415 | 75.4 |
| Characters | 2578 | 835 | 32.4 % | 2235 | 86.7 % |

(a) (b) (c)

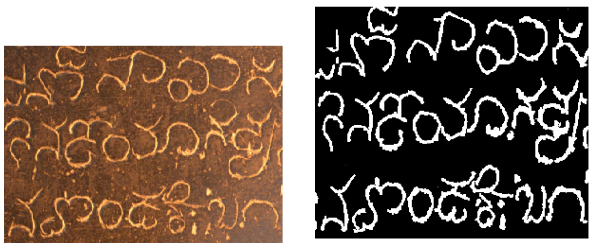
Fig. 5. (a) Original image (b) Output image of proposed method and (c) Output after Fast ICA based enhancement


(a) (b) (c)

Fig. 9. (a)Image of inscriptions (b) Output of proposed method (c) Corresponding OCR output


(a) (b) (c)

Fig. 10. (a)Image of inscriptions (b) Output of proposed method (c) Corresponding OCR output



(a) (b)

Fig. 11. (a) Source (b) result

deal with asymmetrical distribution of the signal.

## V. Conclusion

A novel method for enhancement of complex and unclear archaeological inscription images has been enhanced and validated using 650 word images. The proposed method establishes the important role of NGFICA in digitizing inscription images which has minimal distinction between text and non text region and blurred edges for the text. The method improved the word and character recognition accuracies from 10.1% to 75.4% and from 32.4% to 86.7% respectively. The method proved successful in efficiently enhancing multilingual inscription images too. This method can be further extended for digitization of ancient coins, manuscripts and archaeological sculptures.

## References

[1] S. Kuo and M. Ranganath, "Real time image enhancement for both text and color photo images," *ICIP*, vol. 1, pp. 159–162, 1995.

[2] C. Wolf, J.-M. Jolion, and F. Chassaing, "Text localization enhancement and binarization in multimedia documents," *ICPR*, vol. 4, pp. 1037–1040, Sept. 2002.

[3] U. Garain, A. Jain, A. Maity, , and B. Chanda., "Machine reading of camera-held low quality text images an ica-based image enhancement approach for improving ocr accuracy," *ICPR*, pp. 1–4, June 2008.

[4] X. Wen and D. Luo, "Performance comparison research of the fecg signal separation based on the bss algorithm," *Research Journal of Applied Sciences Engineering and Technology*, vol. 4, no. 16, pp. 2800–2804, March 2012.

[5] L. Agnihotri and N. Dimitrova, "Text detection for video analysis," *Proc. IEEE Int. Workshop on Content-Based Access of Image and Video Libraries*, pp. 109–113, June 1999.

[6] X. S. Hua, P. Yin, and H. J. Zhang, "Efficient video text recognition using multiple frame integration," *Proc. Int. Conf. Image Processing*, vol. 2, pp. 22–25, Sept. 2004.

[7] K. C. K. Kim, "Scene text extraction in natural scene images using hierarchical feature combining and verification," *in Proceedings of International Conference Pattern Recognition*, vol. 2, pp. 679–682, August 2004.

[8] G. R. M. Babu, P. Srimayee, and A. Srikrishna, "Heterogenous images using mathematical morphology," *Journal of Theoretical and Applied Information Technology.*, vol. 15, no. 5, pp. 795–825, November 2008.

[9] EhsanNadernejad, S. Sharifzadeh, and H. Hassanpour, "Edge detection techniques:evaluations and comparisons," *Applied Mathematical Sciences*, vol. 2, no. 31,

pp. 1507 – 1520, Sept. 2008.

[10] M. Seeger and C. Dance, *Binarising Camera Images for OCR*. Europe: Xerox Research Centre, 2000.

[11] S. Buzykanov, "Enhancement of poor resolution text images in the weighted sobolev space," *IWSSIP,Vienna, Austria*, pp. 536–539, April 2012.

[12] B.Gangamma, S. M. K, and A. V. Singh, "Restoration of degraded historical document image," *Journal of Emerging trends in computing and information sciences*, vol. 3, no. 5, pp. 36–39, May 2012.

[13] S. C. Pritirege, "Palm leaf manuscript color document image enhancement by using improved adaptive binarization method," *ICVGIP 2008*, pp. 536–539, December 2008.

[14] C. L. Tan, R. Cao, and P. shen, "Restoration of archival documents using a wavelet technique," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 10, pp. 536–539, October 2002.

[15] A. S. Rao, G. Sunil, N. V. Rao, T. Prabhu, L. Reddy, and A.S.C.S.Sastry, "Adaptive binarization of ancient documents," *Second International Conference on Machine Vision*, pp. 536–539, October 2009.

[16] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the fast ica algorithm and cram rao bounds for linear independent component analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1189–1203, April 2006.

[17] P.Chevalier, L. Albera, P. Comon, and Ferreol, "Comparative performance analysis of eight blind source separation methods on radio communications signals," *Proc. International Joint Conference on Neural Networks*, vol. 8, no. 2, pp. 251–276, July 2004.

[18] S. Choi, A. Cichocki, and S. Amari, "Flexible independent component analysis," *Journal of VLSI Signal Processing*, vol. 26, no. 2, pp. 25–38, August 2000.

[19] S. Choi, "Independent component analysis," *12th WSEAS International Conference on COMMUNICATIONS*, pp. 159–162, July 2008.

[20] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing System*, pp. 251–276, 1996.

[21] S. Amari and S. Douglas, *WHY NATURAL GRADIENT?* Japan: Brain Style Information Systems Group, 2001.

[22] "Optical character recognition," Website, http://www-onlineocr.net.

# Operational Trans-Resistance Amplifier Based Tunable Wave Active Filter

*Mayank BOTHRA[1], Rajeshwari PANDEY[2], Neeta PANDEY[2], Sajal K. PAUL[3]*

[1] Technology R&D, ST Microelectronics, Greater Noida, India
[2] Dept. of Electronics and Communication, Delhi Technological University, Bawana Road, Delhi, India
[3] Dept. of Electronics Engineering, Indian School of Mines, Dhanbad, India

mayankbothra@gmail.com, rajeshwaripandey@gmail.com, n66pandey@rediffmail.com, sajalkpaul@rediffmail.com

**Abstract.** *In this paper, Operational Trans-Resistance Amplifier (OTRA) based wave active filter structures are presented. They are flexible and modular, making them suitable to implement higher order filters. The passive resistors in the proposed circuit can be implemented using matched transistors, operating in linear region, making them fully integrable. They are insensitive to parasitic input capacitances and input resistances due to the internally grounded input terminals of OTRA. As an application, a doubly terminated third order Butterworth low pass filter has been implemented, by substituting OTRA based wave equivalents of passive elements. PSPICE simulations are given to verify the theoretical analysis.*

## Keywords

Wave active filter, OTRA, scattering parameters, ladder, tunable filter.

## 1. Introduction

There are many advantages of higher order filters using doubly terminated lossless ladders, like, low sensitivity to component tolerances, ample design information and design tables that can be readily applied [1]. However, inductor realization in an integrated circuit is a challenging task.

There are various techniques that circumvent these shortcomings like element replacement and operational simulation. If operational simulation is employed, Signal Flow Graphs are used to emulate the relationship between various passive elements. These are then physically realized using lossy and lossless integrators [1]. Realizing lossless integrators is difficult because of non-ideal characteristics of passive components used. Besides, floating capacitors are used in this topology, which are not very favorable in IC implementation. In the case of element replacement approach, inductors are replaced by gyrators. Although this practice leads to good results with low noise sensitivity, realizing high quality floating inductors proves to be difficult [2]. Another element replacement method using Frequency Dependent Negative Resistance (FDNR) was proposed by Bruton [1], and works well with low pass filters. LC ladder filters can also be emulated using Linear Transformation approach wherein every section of the original ladder prototype can be realized using active elements individually [3]. One drawback of this method is that it uses lossless integrators.

Apart from these approaches, the wave method [2] is also used for realizing higher order resistively terminated LC ladder filter which gives excellent results. It uses wave equivalents for different passive elements which can be readily substituted to realize a filter. In this approach, the filter realization is based on modeling the forward and reflected voltage waves. The available wave active filters [2], [4]-[10] use various active blocks such as OTA [4], current amplifier [5], CMOS cascode current mirrors [6], FPAA [7], OPAMP [2], [8], CFOA [9], and DVCCCTA [10] and operate in current [3]-[7] and voltage [2], [8]-[10] mode.

This paper presents design approach for realization of OTRA based higher order wave filter. OTRA being a current mode building block does not suffer from low slew rate and fixed gain bandwidth product [11] unlike conventional voltage mode op-amps. Additionally it has a unique feature of low impedance voltage output and is also free from the effects of parasitic capacitances and resistances at the input due to internally grounded input terminals [12], [13].

OTRA also allows the implementation of linear MOS based resistors [13], which is a huge advantage when going for IC fabrication. This property is also exploited to make the filters tunable. Although a number of tunable ladder circuits based on current-mode approach have been reported in open literature [14]-[16], they do not provide voltage output. Some of the features of the proposed work are:

- Modular structures which can be easily substituted to LC-ladder filter circuits. Provides an easy 'ready to use' method to realize ladders.
- Only lossy integrators are employed, which are easy to realize, and since OTRA inputs are virtually grounded, it is free from effect of parasitic elements.

- It doesn't employ passive resistors; instead it uses linear MOSFET based resistors which are voltage controlled thus making circuits electronically tunable.

Section 2 elaborates on the concept of wave filter. Section 3 elaborates on how OTRA can be employed for this application. Simulation results for a third order Butterworth filter are shown in Section 4 and Section 5 concludes the paper.

## 2. Wave Filter Approach

The concept of wave filter is introduced in [2], [8]. This approach talks of applying scattering parameters to ladder filters. It uses voltage waves instead of power waves. Scattering matrix of a two port network is given as:

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \cdot \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}. \tag{1}$$

In a two port network having a series branch admittance Y, as shown in Fig. 1, the scattering parameters, assuming the normalization resistance as $R_n$, are obtained as:

$$S_{11} = \frac{1}{2R_nY+1}, \tag{2}$$

$$S_{12} = \frac{2R_nY}{2R_nY+1}, \tag{3}$$

$$S_{21} = \frac{2R_nY}{2R_nY+1}, \tag{4}$$

$$S_{22} = \frac{1}{2R_nY+1}. \tag{5}$$

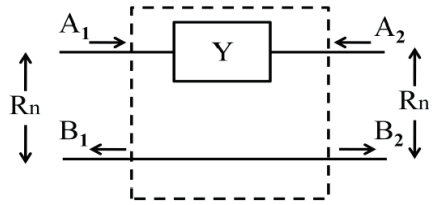



**Fig. 1.** Series branch admittance Y.

For a series branch inductance L, using (2), (3), (4) and (5), (1) reduces to

$$B_1 = \frac{1}{\left(\frac{2R_n}{sL}\right)+1}A_1 + \frac{\frac{2R_n}{sL}}{\left(\frac{2R_n}{sL}\right)+1}A_2, \tag{6}$$

$$B_2 = \frac{\frac{2R_n}{sL}}{\left(\frac{2R_n}{sL}\right)+1}A_1 + \frac{1}{\left(\frac{2R_n}{sL}\right)+1}A_2. \tag{7}$$

This can be further simplified to:

$$B_1 = A_1 - \frac{1}{(1+s\tau_L)}(A_1 - A_2), \tag{8}$$

$$B_2 = A_2 + \frac{1}{(1+s\tau_L)}(A_1 - A_2). \tag{9}$$

In (9), $\tau_L = L/2R_n$ is the time constant. Fig. 2 shows the symbolic representation of the wave equivalent of series branch inductor L.

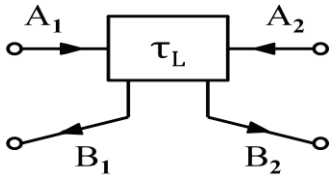



**Fig. 2.** Wave equivalent of series branch inductance L, $\tau_L = L/2R_n$.

To calculate the S-matrix for series branch capacitance C, (1) reduces to:

$$B_1 = \frac{1}{2sCR_n+1}A_1 + \frac{2sCR}{2sCR_n+1}A_2, \tag{10}$$

$$B_2 = \frac{2sCR_n}{2sCR_n+1}A_1 + \frac{1}{2sCR_n+1}A_2. \tag{11}$$

(10) and (11) can be further simplified to:

$$B_1 = A_2 + \frac{1}{1+s\tau_C}(A_1 - A_2), \tag{12}$$

$$B_2 = A_1 - \frac{1}{1+s\tau_C}(A_1 - A_2). \tag{13}$$

In (12) and (13), $\tau_C = 2CR_n$ is the time constant. It is observed that (8) and (9) are similar to (12) and (13) respectively, and can be obtained from each other by interchanging the output terminals $B_1$ and $B_2$.

This result can be generalized to show that for a series branch admittance Y, its dual admittance (Y`) can be obtained by using the following equation [2]:

$$Y' = \frac{1}{4R_n^2 Y}. \tag{14}$$

Accordingly the wave equivalent symbol of series branch capacitance C, as shown in Fig. 3, indicates this fact.

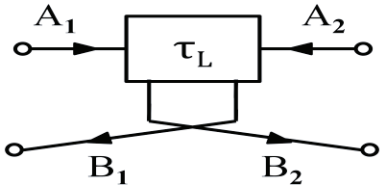
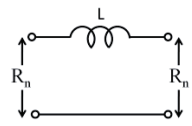
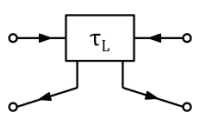
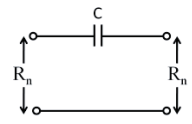
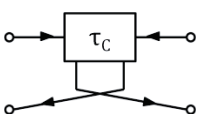
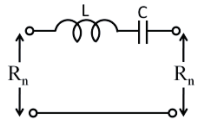
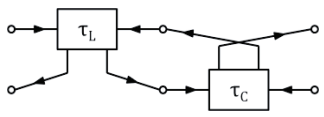
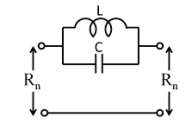
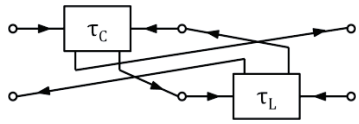



**Fig. 3.** Wave equivalent of series branch capacitance *C*, $\tau_C = 2CR_n$.

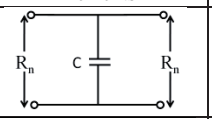
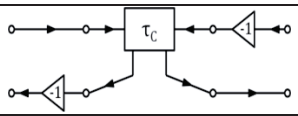
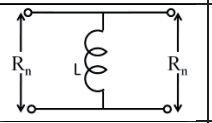
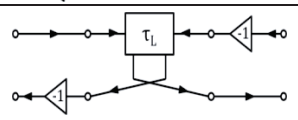
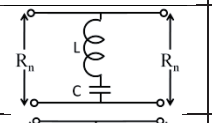
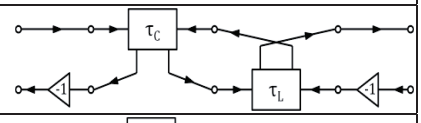
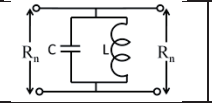
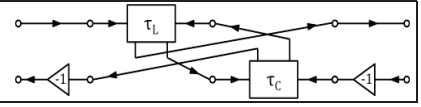
For an inductor *L* connected in series with capacitance *C* in a series arm the wave equivalent can be obtained by

cascading the wave equivalents of $L$ and $C$. If the terminals are interchanged, the wave equivalent for a tank circuit connected in series branch can be obtained. Tab. 1 [2], [8] gives wave equivalents for all the series branch elements.

Proceeding in a similar manner, wave equivalents for shunt branch elements can also be derived. Tab. 2 [2], [8] lists the results for shunt branch elements.

| Series Branch Elements | Wave Equivalents |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |

**Tab. 1.** Wave equivalents of series branch elements [2], [8].

| Shunt Branch Elements | Wave Equivalents |
|---|---|
|  |  |
|  |  |
|  |  |
|  |  |

**Tab. 2.** Wave equivalents of shunt branch elements [2], [8].

## 3. OTRA Based Wave Active Filter

Fig. 4 shows an OTRA circuit symbol. Its transfer matrix is given in (15) [12]. It has low impedance input and output terminals. Ideally the trans-resistance gain $R_m$ approaches infinity and when negative feedback is used then $I_1 = I_2$ [13].

$$\begin{bmatrix} V_1 \\ V_2 \\ V_O \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ R_m & -R_m & 0 \end{bmatrix} \begin{bmatrix} I_1 \\ I_2 \\ I_O \end{bmatrix}. \tag{15}$$

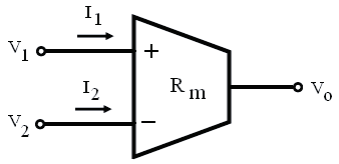



**Fig. 4.** OTRA circuit symbol.

A closer study of (8), (9), Tabs. 1 and 2 would reveal that the realization of wave equivalents would require a summer, subtractor, a subtracting lossy integrator and an inverter.

An OTRA based summer is shown in Fig. 5. The circuit makes use of three resistors and one OTRA. Equation (16) can be obtained by equating the current at the inverting and the non-inverting terminal, which can be further simplified to (17).

$$\frac{V_O}{R} = \frac{V_{IN1}}{R} + \frac{V_{IN2}}{R}, \tag{16}$$

$$V_O = V_{IN1} + V_{IN2}. \tag{17}$$

Similar analysis of a subtractor, shown in Fig. 6, gives

$$V_O = V_{IN1} - V_{IN2}. \tag{18}$$

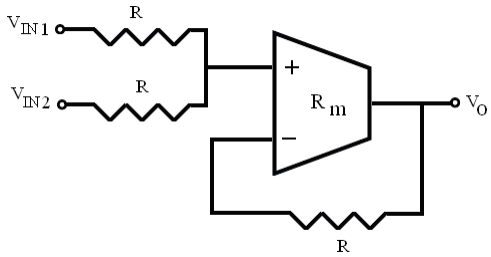



**Fig. 5.** Summer.

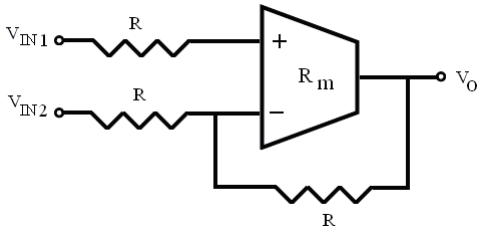



**Fig. 6.** Subtractor.

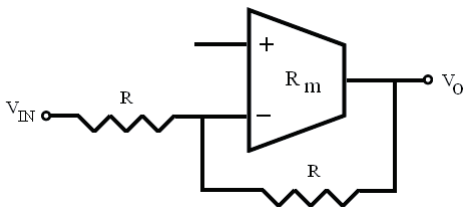



**Fig. 7.** Inverter.

Fig. 7 shows an inverter. In this case, since the non-inverting terminal has been left open, there would be no current flowing into the inverting terminal as well. It can be described by the following equation:

$$V_O = -V_{IN}. \tag{19}$$

Fig. 8 shows a subtracting lossy integrator. Its output can be described by:

$$V_O = \frac{1}{1+sCR}\left(V_{IN1} - V_{IN2}\right). \tag{20}$$

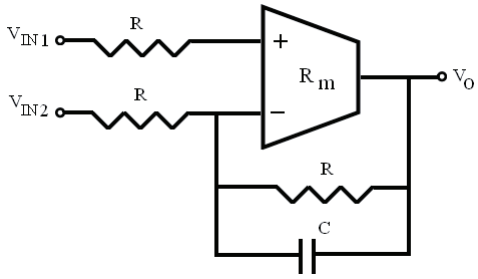



**Fig. 8.** Lossy integrator.

The current differencing property of the OTRA makes it possible to implement the resistors connected to the input terminals of OTRA, using MOS transistors with complete non linearity cancellation [13]. Fig. 9 shows the MOS based linear resistor using OTRA. Each resistor requires two matched n-MOSFETs connected in a manner as shown in Fig 9.

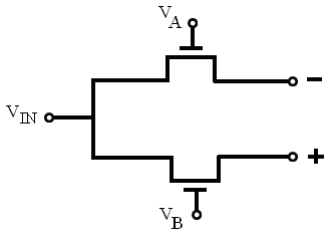



**Fig. 9.** MOS based resistor.

Symbols '+' and '-' represent the non-inverting and the inverting terminals of the OTRA. As shown in the figure, the voltages at the drain and the source terminals for both MOSFETs are equal. On taking the difference of the currents flowing in the two transistors, the non-linearity gets cancelled out. The following equation defines the resistor that has been realized.

$$R = \frac{1}{K_N\left(V_A - V_B\right)} \tag{21}$$

where $R = \dfrac{1}{K_N\left(V_A - V_B\right)}$

$K_N$ needs to be determined for the transistors being used to implement the resistors. $\mu$, $C_{OX}$ and $W/L$ represent standard transistor parameters. The choice of voltages $V_A$ and $V_B$ is important. The circuit shown in Fig. 9 realizes a resistor of value expressed in (21) at the inverting terminal. If it is desired to realize a resistor of the same value at the non-inverting terminal, then $V_A$ and $V_B$ must be interchanged.

Using the blocks defined by (17), (18), (19) and (20), the wave equivalent for a series branch inductor, defined by (8) and (9), can be drawn and is shown in Fig. 10. The dashed blocks indicate the individual blocks which constitute the entire wave equivalent. It represents the symbol shown in Fig. 2. This can now be used as the elementary block to synthesize the wave equivalents for all the elements listed in Tab. 1 and Tab. 2.

By introducing inverters and interchanging output terminals, all other wave equivalents can be obtained. The circuit in Fig. 10 can be described by the following equations:

$$B_1 = A_1 - \frac{1}{1+sCR}\left(A_1 - A_2\right), \tag{22}$$

$$B_2 = A_2 + \frac{1}{1+sCR}\left(A_1 - A_2\right). \tag{23}$$

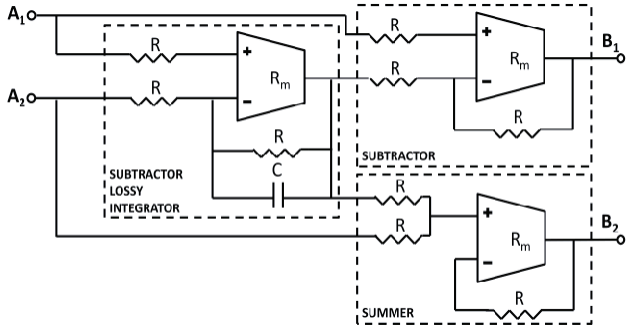



**Fig. 10.** Equivalent circuit for series branch inductance.

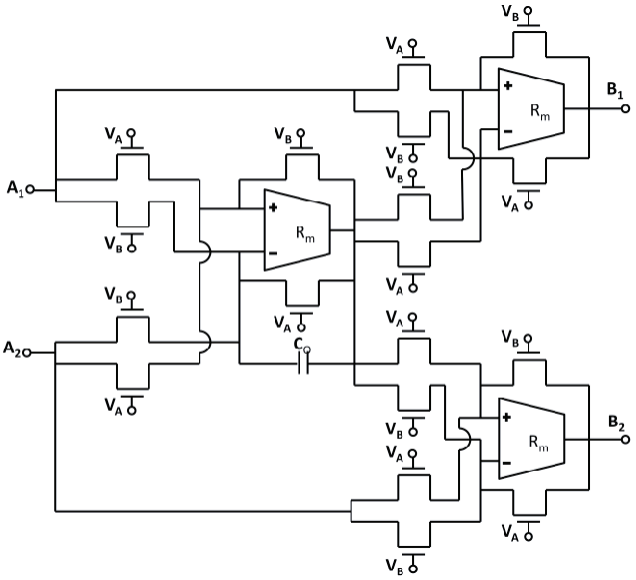



**Fig. 11.** MOS- C equivalent of circuit of Fig. 10.

Fig. 11 shows the circuit of Fig. 10 with MOS based linear resistors. The actual value of inductance $L_A$ realized by circuit of Fig. 11 would be obtained by comparing (22) and (23) with (8) and (9). The realized value is

$$L_A = 2R_nCR. \tag{24}$$

Similarly, comparing (22) and (23) with (12) and (13), the realized value of C is:

$$C_A = \frac{CR}{2R_n}. \tag{25}$$

Resistor $R$ can be controlled through voltages $V_A$ and $V_B$. If $C$ is assumed to be of some constant value, then the value of $L_A$ and $C_A$ can be controlled using $R$. This forms the basis of tunability of the circuit. Similarly, the actual values of $L$ and $C$ for wave equivalents of shunt branch elements, as presented in Tab. 2, are given by:

$$L_A = \frac{R_n CR}{2}, \tag{26}$$

$$C_A = \frac{2CR}{R_n}. \tag{27}$$

For a filter, if $L_n$ and $C_n$ are the normalized inductor and capacitor values respectively, $\omega_0$ is the normalizing pole frequency and $R_n$ is the normalizing resistance, then to de-normalize $L_n$ and $C_n$ we make use of the following expressions:

$$L_A = \frac{R_n}{\omega_0} L_n, \tag{28}$$

$$C_A = \frac{1}{R_n \omega_0} C_n. \tag{29}$$

Restating (24) and (25), such that different values of $C$ are used for $L_A$ and $C_A$, i.e. $C_L$ and $C_C$ respectively, we get:

$$L_A = 2R_n C_L R, \tag{30}$$

$$C_A = \frac{C_C R}{2R_n}. \tag{31}$$

Equating (28) and (29) with (30) and (31) respectively, we get:

$$C_L R = \frac{1}{2\omega_0} L_n, \tag{32}$$

$$C_C R = \frac{2}{\omega_0} C_n. \tag{33}$$

The expression, for controlling $\omega_0$ using $R$, needs to be worked out for each circuit. A simple algorithm can be worked out to achieve the exact expression and range of tunability. For the frequency $\omega_0$, $C_L$ and $C_C$ are calculated in terms of $R$, as per the $L_n$ and $C_n$ values. For a suitable $R$ value, once $C_L$ and $C_C$ have been fixed, either (32) or (33) can be used to describe the relationship between $R$ and $\omega_0$ as:

$$\omega_0 = \frac{K}{R}. \tag{34}$$

Using (32) and (33), $K$ can be described as follows:

$$K = \frac{2C_n}{C_C} = \frac{L_n}{2C_L}. \tag{35}$$

Combining (21) and (34) one may get:

$$\omega_0 = KK_N(V_A - V_B). \tag{36}$$

Equation (34) describes how $\omega_0$ can be controlled using $R$. $R$ in turn is controlled by (21) and the overall relationship is described by (36).

# 4. Simulation Results

To demonstrate the wave filter approach using OTRA, a doubly terminated third order Butterworth low pass filter, as shown in Fig. 12, has been implemented. The wave equivalent circuit of the same is shown in Fig. 13 in which the reflected waves are available at $V_{OL}$ and $V_{OH}$. These outputs complement each other by virtue of wave theory [8].Thus as the $V_{OL}$ represents the low pass filter response, its complementary high pass output is available at $V_{OH}$. The normalized values of components are $L_{n1} = 2$, $C_{n1} = 1$ and $C_{n2} = 1$. OTRA is realized using the CMOS circuit schematic given in Fig. 14 [17]. The filter specifications are as follows: $f_P$ = 200 kHz and maximum attenuation in pass band $\alpha_{MAX}$ is 3 dB.

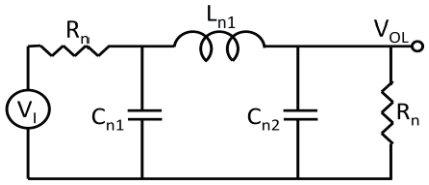



**Fig. 12.** 3rd order low pass Butterworth filter.

The value of normalizing resistance $R_n$ is chosen to be 2.5 kΩ. De-normalizing the values of $L_{n1}$, $C_{n1}$ and $C_{n2}$ we get:

$$L_A = 3.98\,\text{mH}, \tag{37}$$

$$C_{A1} = C_{A2} = 318.31\,\text{pF}. \tag{38}$$

Setting the value $R$ initially to 12 kΩ, we can calculate the value of $C_L$ for $L_{A1}$ as 66.33 pF and $C_C$ for $C_{A1}$ and $C_{A2}$ as 132.66 pF. The value of $K$, as per (35) is $1.508 \times 10^{10}$. For this simulation exercise, the value of $K_N$ was found to be $5.25 \times 10^{-4}$ $A/V^2$. The required $V_A$ and $V_B$ values for $R$ to be 12 kΩ were found to be 0.908 V and 0.75 V as per (21). Fig. 15 shows the low pass filter response of the circuit at $V_{OL}$. The complementary high pass output $V_{OH}$, as represented in Fig. 13, has been plotted in Fig.16.

The performance of the proposed circuit is compared with the previous voltage mode structures [2], [8]-[10] in terms of power consumption, THD, output noise and electronic tunability. It may be noted from Tab. 3 that the topology presented in [9] shows best THD result, however the structure is not electronically tunable. Although the most recently reported literature [10] is having better THD performance its simulated power consumption is higher as compared to the proposed one. The relevant data for structures of [2], [8] which are designed using commercially available OPAMPs, is not available in the literature.

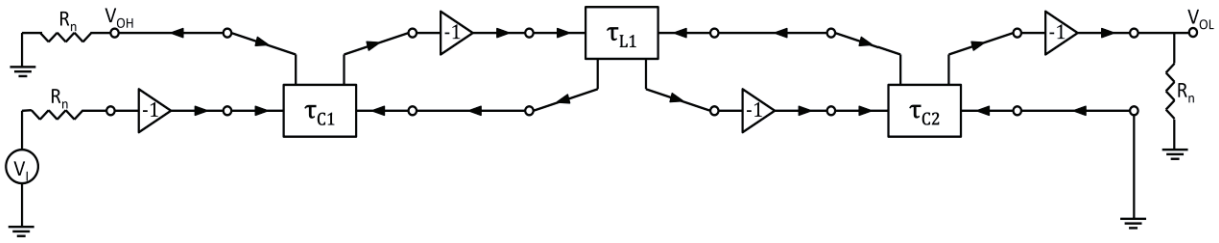



**Fig. 13.** Wave equivalent of circuit of Fig.12.

| Ref | Active block and technology used | Filter structure | % THD | Power consumption (mW) | Output noise voltage (V/ HZ $^{1/2}$) | Electronic tunability |
|---|---|---|---|---|---|---|
| [9] | CFOA (Commercially available IC AD844 with ± 5 V power supply) | 3$^{rd}$ order elliptic Low pass | 1 % for 1 Vpp signal | N/A | N/A | No |
| [10] | DVCCCTA (CMOS Technology 0.25 µm, Power supply ± 1.25 V) | 4$^{th}$ order Butterworth Low pass | Less than 5 % up to 225 mVpp signal | 59.2 | $8.36 \times 10^{-8}$ | Yes |
| proposed | OTRA CMOS Technology (0.5 µm, Power supply ±1.5 V) | 3$^{rd}$ order Butterworth Low pass | Less than 5 % up to 125 mV pp signal | 10.7 | $7.26 \times 10^{-8}$ | Yes |

**Tab.3.** Comparison with the voltage mode filter structures.

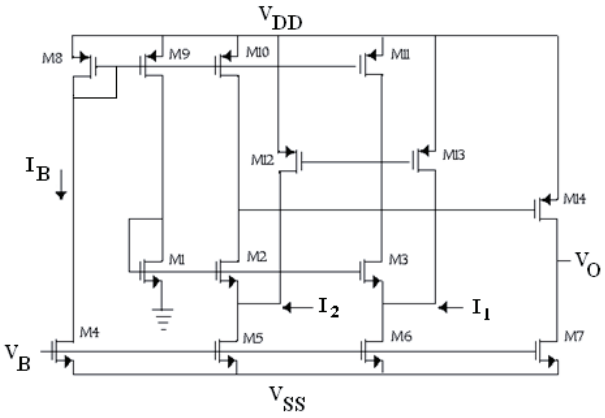



**Fig. 14.** CMOS realization of OTRA.

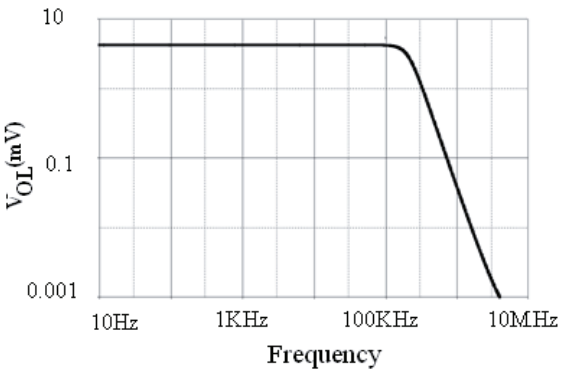



**Fig. 15.** Low pass response $V_{OL}$.

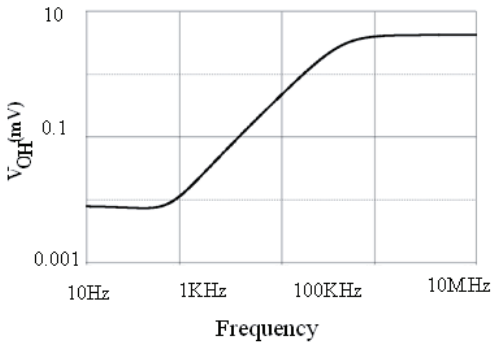



**Fig. 16.** Complementary high pass response $V_{OH}$.

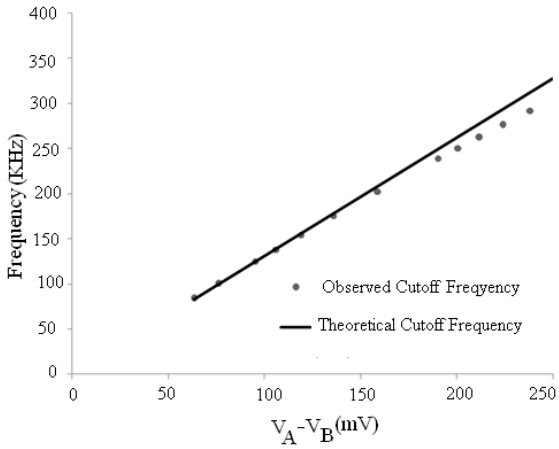



**Fig. 17.** Comparison curve between theoretical and observed frequency.

The proposed circuit can be tuned to different cut-off frequencies by controlling $V_A$ and $V_B$ as described by (36). Fig. 17 shows the comparison between theoretical and the observed frequencies, obtained by variation of control voltages $V_A$ and $V_B$. All simulations are done using PSPICE program using 0.5 μm CMOS technology parameters form MOSIS (AGILENT).

## 5. Conclusion

In this paper, the design of tunable wave active filter based on OTRA has been presented. It provides an alternative form of realizing ladders. Advantages of current mode approach have been exploited. The use of OTRA allows the simple implementation of linear resistor using only two MOSFETs. The controllability of the resistors value by a single voltage source allows the parameters of the proposed filters to be electronically tunable. On the downside, the circuit is slightly cumbersome to realize, though it is modular and can easily be implemented using the reference design tables. When compared with the previous wave active filters reported in [2], [8] and [9], the proposed one provides advantages of current mode design and is tunable as well. In comparison to the circuit presented in [10], OTRA as the basic building block is simpler to realize and also provides a low impedance voltage output, making it suitable for driving voltage input devices.

## References


[1] SHAUMANN, R., VAN VALKENBURG, M. E. *Design of Analog Filters*. Oxford (UK): Oxford University Press, 2001.

[2] WUPPER, H., MEERKOTTER, K., New active filter synthesis based on scattering parameters. *IEEE Transactions on Circuits and Systems*, 1975, vol. 22, no. 7, p. 594 - 602.

[3] HWANG, Y. S., WU, D. S., CHEN, J. J., SHIH, C. C., CHOU, W. S. Realisation of high order OTRA-MOSFET-C active filters. *Circuits, Systems Signal Processing*, 2007, vol. 26, no. 2, p. 281 - 291.

[4] TINGLEFF, J., TOUMAZOU, C. A 5th order lowpass current mode wave active filter in CMOS technology. *Analog Integrated Circuits and Signal Processing,* 1995, vol.7, p. 131 - 137.

[5] SPANIDOU, A., PSYCHALINOS, C. Current amplifier-based wave filters. *Circuits, Systems Signal Processing*, 2005, vol. 24, no. 3, p. 303 - 313.

[6] SOULIOTIS, G., HARITANTIS, I. Current-mode differential wave active filters. *IEEE Transactions on Circuits and Systems-I: Regular Papers,* 2005, vol. 52, no. 1, p. 93 - 98.

[7] FRAGOULIS, N. SOULIOTIS, G., BESIRIS, D., GIANNAKOPOULOS, K. Field-programmable analogue array design based on the wave active filter design method. *International Journal of Electronics and Communication (AEU)*, 2009, vol.63, p. 889 - 895.

[8] HARITANTIS, I., CONSTANTINIDES, A., DELIYANNIS, T. Wave active filters. *IEE Proceedings*, 1976, vol. 123, no. 7, p. 676 - 682.

[9] KOUKOU, G., PSYCHALINOS, C. Modular filter structures using current feedback operational amplifiers. *Radioengineering,* 2010, vol. 19, no. 4, p. 662 - 666.

[10] PANDEY, N., KUMAR, P. Realization of resistorless wave active filters using differential voltage current controlled conveyor transconductance amplifier. *Radioengineering*, 2011, vol. 20, no. 4, p. 911 - 916.

[11] TOUMAZOU, C., LIDGEY, F. J., HAIGH, D. G. *Analogue IC Design: The Current Mode Approach*. Stevenage (UK): *Peregrinus*, 1990.

[12] CHEN, J., TSAO, H., CHEN, C. Operational trans-resistance amplifier using CMOS technology. *Electronics Letters,* 1992, vol. 28, no. 22, p. 2087 - 2088.

[13] SALAMA, K. N., SOLIMAN, A. M. CMOS OTRA for analog signal processing applications. *Microelectronics Journal,* 1999, vol. 30, p. 235 - 245.

[14] BIOLEK, D., BIOLKOVA, V. Tunable CDTA-based ladder filters. In *Proceedings of 2nd WSEAS ICEASCS.* Singapore, 2003, p. 462 - 466.

[15] JIRASEREE-AMORNKUN, A., FUJII, N., SURAKAMPONTORN, W. Realization of electronically tunable ladder filters using multi output current controlled conveyors. In *Proceedings of the 2003 International Symposium on Circuits and Systems (ISCAS).* Bangkok (Thailand), 2003, vol. I, p. I-541 - I-544.

[16] JAIKLA, W., SITIPRUCHYANUN, M. A systematic design of electronically tunable ladder filters employing DO-OTAs. In *Proceedings of 4th International Conference on Electrical Engineering / Electronics, Computer, Telecommunications and Information Technology (ECTI-CON).* Chiang Rai (Thailand), 2007, p. 61 - 64.

[17] MOSTAFA, H., SOLIMAN, A. M. A modified CMOS realization of the operational transresistance amplifier (OTRA). *Frequenz,* 2006, vol. 60, no. 3-4, p. 70 - 77.


## About Authors ...


**Mayank BOTHRA** was born in 1987 and received his B.E. degree in Electronics and Communication from Delhi Technological University (formerly Delhi College of Engineering, Delhi University) in 2009. He worked as a development engineer in the Embedded Group at Kritikal Solutions, Noida for two years. Currently he is working as a design engineer in the Memories' team at ST Microelectronics. His research interests are in analog circuit design and microelectronics.

**Rajeshwari PANDEY** received her B.Tech. (Electronics and Telecommunication) from J. K. Institute of Applied Physics, University of Allahabad in 1988 and her M.E (Electronics and Control) from BITS, Pilani, Rajasthan, India in 1992. She has served BITS Pilani, AERF, Noida and Priyadarshini College of Computer Science, Noida in various capacities. Currently, she is assistant professor in Department of Electronics and Communication Engineering, Delhi Technological University, Delhi. Her research interests include analog integrated circuits and microelectronics.

**Neeta PANDEY** received her M.E. in Microelectronics from Birla Institute of Technology and Sciences, Pilani and Ph.D. from Guru Gobind Singh Indraprastha University, Delhi. She has served in Central Electronics Engineering Research Institute, Pilani, Indian Institute of Technology, Delhi, Priyadarshini College of Computer Science, Noida

and Bharati Vidyapeeth's College of Engineering, Delhi in various capacities. At present, she is assistant professor in ECE department, Delhi Technological University. A life member of ISTE, and member of IEEE, USA, she has published papers in international, national journals of repute and conferences. Her research interests are in analog and digital VLSI design.

**Sajal K. PAUL** received his B.Tech., M.Tech. and Ph.D. in Radio Physics and Electronics from the Institute of Radio Physics and Electronics, University of Calcutta. He has served Webel Telecommunication Industries, Kolkata; Indira Gandhi National Open University (IGNOU), Kolkata; Advanced Training Institute for Electronics & Process Instrumentation (ATI-EPI), Hyderabad; North Eastern Regional Institute of Science & Technology (NERIST), Nirjuli and Delhi College of Engineering (DCE), Delhi in various capacities. He has served the Department of Electronics Engineering, Indian School of Mines, Dhanbad as head of the department and at present is a professor of the same department. His research interest includes microelectronic devices, electronic properties of semiconductor and bipolar and MOS analog integrated circuits. Dr. Paul has more than 70 research publications in international and national journals of repute and conferences.

Advances in Physics Theories and Applications
ISSN 2224-719X (Paper) ISSN 2225-0638 (Online)
Vol.16, 2013 

# Photon-Photon Collision: Simultaneous Observation of Wave-Particle Characteristics of Light

Himanshu Chauhan
TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication,
Department of Applied Physics, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Bawana Road, Delhi-110042, India
Email: chauhan.himanshu_om@yahoo.com

Swati Rawal
TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication,
Department of Applied Physics, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Bawana Road, Delhi-110042, India
Email: swati.rawal@yahoo.com

R.K. Sinha (corresponding author)
TIFAC-Centre of Relevance and Excellence in Fiber Optics and Optical Communication,
Department of Applied Physics, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), Bawana Road, Delhi-110042, India
*Email: dr_rk_sinha@yahoo.com

**Abstract**

The proposed paper presents the analysis of electromagnetic waves meeting at a point *in terms of their particle characteristics*. The observation that light beams moves un-deviated when encountered at a point, which is commonly justified on the wave characteristics of light, is now presented as momentum and wavelength exchange phenomenon of photon collision. Theoretical and mathematical justification of photon's inter-collision, on the basis of their *quasi-point particle* behavior is offered and the observation of the non-variation of wavelength of light beams is explained. Thus, the observation of light's non-deviation at the crossing point is explained as momentum exchange phenomenon on the basis of particle characteristics of light.

**Keywords:** Basic Quantum Mechanics, Bohr's Complementary Principle, Collision Mechanics

## 1. Introduction

The proposal of light as an electro-magnetic wave (Maxwell, 1865), explained its various wavy phenomena like interference and diffraction. However, in the beginning of the 20$^{th}$ century, particle type characteristics of light also came into picture, with many experimental supports (Einstein, 1905; Compton, 1923; Raman, 1929). According to the existing literature, light can show either wave characteristic or particle characteristic (Ghatak, *Quantum Mechanics*; Beiser, *Concepts of Modern Physics*; Feynman, *The Feynman Lectures of Physics*), however, the simultaneous observation of both the characteristics of light, in a single particular experiment, is not possible (Beiser,

*Concepts of Modern Physics*), i.e. it is impossible to explain the behavior of light in an experiment, on the simultaneous basis of both particle and wave characteristics. Only one at a time can account for the situation. This is called Bohr's complementary principle. However, some of the physicists have demonstrated (even experimentally also) that it is possible to simultaneously observe both wave and particle properties of light (Afshar, 2007). The diffraction pattern of light, passing through a slit, is lost, if, the track of photons is monitored, in accordance with the limitations set by Bohr's Principle of Complementarity (Afshar, 2007). However, in Afshar's experiment, the presence of sharp interference was observed, while reliably maintaining the information about the particular pinhole through which each individual photon had passed (Afshar, 2007). Thus accounting the situation beyond the limits set by Bohr's Principle of Complementarity.

Being motivated by their work, we have presented simultaneous involvement of wave-particle characteristics of light in another well-know situation, concerning the meeting of two light beams at a common point. It is shown, theoretically, that both particle and wave type characteristics of light can explain the behavior of light in this situation. If two light beams are converged at a common point, the wave-fronts of light waves cross each other without deviating from their original direction. Moreover, no variation in the wavelength of light beams is observed. Figure 1 explains the observed situation. The two light beams originating from sources $S_1$ and $S_2$, converge at a point A making angle $\theta_1$ and $\theta_2$, with the vertical, at $S_1$ and $S_2$ respectively.

After passing un-deviated through A, the angle of light beams becomes $\phi_1 and \phi_2$, with the horizontal. Since the light beams propagate un-deviated, through each other, relation between $\theta_1$, $\theta_2$, $\phi_1 and \phi_2$ becomes:-

$$\theta_1 + \theta_2 + \phi_1 + \phi_2 = 180^0$$
$$\phi + \theta = 180^0 \qquad (1)$$

; where, $\theta=\theta_1+\theta_2$ and $\phi = \phi_1 + \phi_2$. This un-deviated propagation of light in such a situation is accountable by its wavy nature. The wave fronts of light beams passes through each other just like two ripple waves in a still pond. Thus, the wavelength of light beams remains unchanged and the waves propagate un-deviated through each other. *In the present paper, we have accounted this observation on the basis of the particle characteristics of light. This approach for explaining the non-variation of wavelength when the two light beams cross each other, on account of particle characteristics of light, is not familiar to the scientific community and literature yet, to the best of our knowledge*. It has been demonstrated that the un-deviated passage of light beams and its unchanged wavelengths are the results of light's particle-type behavior. It is shown that at the point of interaction photon collides and at a specific angle (given by equation 1) the photon's momentum and thus their wavelengths interchanges.

The idea of the simultaneous observation of the wave-particle character of light, in the same experiment, emerged on account of the following situations also:-

*Polarization of photons*: Polarization is generally accounted as a wavy phenomenon (Ghatak, *Optics*). However, photon which exhibits particle type nature undergoes polarization (Ghatak, *Optics*).

*Frequency of light remains unaltered when light moves from one medium to another*: - This observation is well accountable on the basis of the wave characteristics of light (Beiser, *Concepts of Modern Physics*), as the frequency of the oscillators of the medium remains equal to the frequency of incident radiation. However**, we have justified the same observation by taking into account the particle characteristics of light.** Suppose the energy of a photon in a medium is $h\nu_1$. On entering another medium, its energy remains invariant due to the non-availability of any energy dissipative mechanism for photon to loss some amount of its original energy. Therefore, the amount of photon's energy in the latter medium, say $h\nu_2$, is equal to its energy in the former medium $h\nu_1$, i.e.:-

$$h\nu_1 = h\nu_2$$
$$\Rightarrow \nu_1 = \nu_2$$

Therefore, on changing the medium of propagation, frequency of light remains unchanged and can be explained in

terms of both wave and particle nature of light.

Convincingly, from these two observations it appears that the simultaneous observation of wave-particle nature of light in same experiments is theoretically possible.

In the upcoming sections the observation of non-deviation and non-variation of wavelengths of light beams converging at a point, which is justified by light's wavy behavior, is now being explained on the basis of the particle characteristics of light.

**2. Photon-Collision in 2-D**

If two light beams meet at a common point, the wave front of the light waves cross each other without being deviated from their original directions and thus, the wavelength of the individual light beams remain unaltered. This unaltered parametric passage of light beams across each other suggests that the particle characteristic of light is not significant in this situation. Therefore, photon does not appear to collide at the point of meeting and no change in momentum (and thus wavelength) or deviation in the angle of incoming light is observed. The wave fronts of light simply pass through each other, similar to ripple waves in a still pond. In this section, the participation of light's particle characteristic is presented i.e. the theoretical justification of the photon collision at the converging point, satisfying the observation of non-variation of wavelengths, is presented. The notion for the occurrence of photon collision in this situation emerged on the account of the symmetry in nature i.e., if two electrons can collide mutually and a photon can collide with an electron; then a photon should also be able to collide with another photon. Thus, the particle type behavior should be visualized when two light beams are encountered at a point.

The **fundamental postulation** is: *inter-collision of photons is possible at the converging point* and *they are quasi-point particles, with volume tending to zero.* Consequently, *the collision has to be elastic*, as expected from two point particles. This elastic collision, at a specific *interchange angle*, will be shown to become perfectly elastic, ensuing photon's momentum, and therefore wavelength, interchanges. This wavelength exchange actually gives the impression that light beams cross each other un-deviated. Working with the assumption adopted, we have derived the wavelength expression for the photon scattered from its original path, due to the collision with other photon.

Figure 2 shows two light beams of wavelengths $\lambda_1$ and $\lambda_2$ originating from the sources $S_1$ and $S_2$, respectively. The beams are converged at the point A such that the angle made by them is $\theta_1$ and $\theta_2$, with the vertical, at points, $S_1$ and $S_2$ respectively. The circular figure on the path $S_1A$ shows the incoming photon from the source $S_1$ and is designated as photon-1. Similarly, the photon of the other source is designated as photon-2. Due to the collision of photons at point A, the beams get deviated from their original incident directions, such that, the angle of light beams becomes $\phi_1 and \phi_2$, with the horizontal. The photon-1 now propagates on the path AP and photon-2 on path AQ, with some varied wavelengths.

The collision of photons results to variation in their momentum and wavelengths. The wavelength of photons after collision are supposed to be $\lambda_1$' and $\lambda_2$' and the corresponding momentums as $P_1'$ and $P_2'$, respectively.

Since the collision is elastic, the momentum and energy would remain conserved during the process.

*2.1. Conservation of Energy*

The postulation of photon's elastic collision provides implementation of Energy Conservation Principle. The photons' energy ($E$) sum should be equal, before and after the collision. Mathematically:-

$$E_1 + E_2 = E'_1 + E'_2 \qquad (2)$$

The energy of a photon is related to its momentum as (Beiser, *Concepts of Modern Physics*)

$$E = Pc$$

Equation (2) therefore becomes:-

$$P_1 + P_2 = P'_1 + P'_2 \tag{3}$$

*2.2. Conservation of Momentum*

The occurrence of photon's elastic collision is achievable only when the sum of the linear momentum of photons before collision, along each individual axis, should be equal to the sum of the photon's linear momentum after collision. Applying this principle along both axes:-

*Along y-axis*:-

Taking the components of photon's momentum along x-axis, the conservation of momentum gives:

$$P_2 \cos\theta_2 - P_1 \cos\theta_1 = P'_2 \sin\phi_2 - P'_1 \sin\phi_1 \tag{4}$$

Similarly, *along x-axis:*-

$$P_1 \sin\theta_1 + P_2 \sin\theta_2 = P'_1 \cos\phi_1 + P'_2 \cos\phi_2 \tag{5}$$

Squaring and adding equation (4) and (5), gives:-

$$P_1^2 + P_2^2 + 2P_1P_2[\sin\theta_1 \sin\theta_2 - \cos\theta_1 \cos\theta_2] = (P'_1)^2 + (P'_2)^2 + 2P'_1 P'_2[\cos\phi_1 \cos\phi_2 - \sin\phi_1 \sin\phi_2] \tag{6}$$

Using the following trigonometric identity:-

$$\sin A \sin B - \cos A \cos B = -\cos(A + B)$$

Equation (6) becomes:-

$$P_1^2 + P_2^2 - 2P_1P_2 \cos(\theta_1 + \theta_2) = (P'_1)^2 + (P'_2)^2 + 2P_1'P_2'\cos(\phi_1 + \phi_2) \tag{7}$$

From equation (3):-

$$P'_1 = P_1 + P_2 - P'_2 \tag{8}$$

Substituting equation (8) in equation (7):-

$$(P'_2)^2[1 - \cos\phi] + (P'_2)[(P_1 + P_2)(\cos\phi - 1)] + P_1P_2[1 + \cos\theta] = 0 \tag{9}$$

where, $\theta = \theta_1 + \theta_2$ and $\phi = \phi_1 + \phi_2$.

Let: $Z = P_2'; (1 - \cos\phi) = l; [(P_1 + P_2)(\cos\phi - 1)] = m;\ P_1P_2(1 + \cos\theta) = n$

Therefore, equation (9) becomes quadratic in *Z*:-

$$(l)Z^2 + mZ + n = 0 \tag{10}$$

The solution of above equations is given by:-

$$Z = \frac{-m}{2l} \pm \frac{\sqrt{m^2 - 4(l)n}}{2l}$$

Substituting the values of *Z*, *l*, *m* and *n* in above equation:-

$$P'_2 = \frac{P_1 + P_2}{2} \pm \frac{\sqrt{(P_1 + P_2)^2 - 4P_1P_2\dfrac{(1 + \cos\theta)}{(1 - \cos\phi)}}}{2} \tag{11}$$

Equation (11) provides the varied momentum expression for photon-2 after collision with photon-1. Substitution of equation (11) in equation (8) provides the momentum expression for the photons-1 as:-

$$P'_1 = \frac{P_1 + P_2}{2} \mp \frac{\sqrt{(P_1 + P_2)^2 - 4P_1P_2\,\dfrac{(1+\cos\theta)}{(1-\cos\phi)}}}{2} \tag{12}$$

Thus, we obtain the expression for the momentum of photons after collision, in the form of equation (11) and equation (12). To determine the varied momentums ($P_1$' and $P_2$'), information regarding the angles is an essential requirement. Since, no equation relates $\theta$ and $\varphi$ in terms of $P_1$ and $P_2$, they are declared as purely variables. The substitution of the particular values of $\theta$ and $\varphi$ would provide the corresponding momentum of photons, after collision. It is observed that the rays of light, as shown in fig. 1, passes through each other without any deviation from the original incident direction and the wavelengths of light beams remains unaltered even after the passage of light beams. Therefore, we should verify that whether this observation of non-variation of wavelengths is satisfied by the momentum equation (11) and (12) or not. Therefore, substituting equation (1) in equation (11) and (12) gives:-

$$P'_2 = \frac{P_1 + P_2}{2} \pm \frac{\sqrt{(P_1 + P_2)^2 - 4P_1P_2}}{2} = \frac{P_1 + P_2}{2} \pm \frac{P_1 - P_2}{2} \quad \Rightarrow P'_2 = P_1, or, P'_2 = P_2$$

$$P'_1 = \frac{P_1 + P_2}{2} \mp \frac{\sqrt{(P_1 + P_2)^2 - 4P_1P_2}}{2} = \frac{P_1 + P_2}{2} \mp \frac{P_1 - P_2}{2} \quad \Rightarrow P'_1 = P_2, or, P'_1 = P_1$$

The second solutions of $P_1$' and $P_2$' are not acceptable, because they dictates non-variation in photon's momentum, which is against the postulation of the occurrence of photons collision. Therefore, only the first solutions of $P_1$' and $P_2$' are admissible. Thus, the momentum of photons after collision becomes:-

$$P'_2 = P_1 \text{ and } P'_1 = P_2$$

The above expressions, for the momentum of photons after collision, shows that photons due to collision exchanges their momentums, which implies that the nature of collision is perfectly Elastic Collision (Verma, *Concepts Of Physics*).

The angle $\varphi=\pi-\theta$ (equation 1) is termed as *interchange angle,* because only at this particular angle, the momentum of photons get interchanged and is indeed the observed angle between the light beams (fig. 1). Since, for photons (Ghatak, *Quantum Mechanics*; Beiser, *Concepts of Modern Physics*; Feynman, *The Feynman Lectures of Physics*)

$$P = h/\lambda$$

Thus, the wavelength of light beams after passage from the intersection point '*A*' becomes:-

$$\lambda'_2 = \lambda_1;\ \lambda'_1 = \lambda_2$$

The above expressions show that the wavelength of light beams after collision gets interchanged. The photon-1, due to collision, acquires the wavelength of photon-2, and vice versa. Therefore, the photon collision results in the wavelength interchange of the light beams. Figure 3 illustrates how the wavelengths of the light beams get interchanged on meeting at point A. The red photon (of wavelength $\lambda_1$), due to collision with the green photon (of wavelength $\lambda_2$) at point A, becomes a green photon and follows path AP and vice versa for photon-2. Therefore, the red photon coming from $S_1$ follows path AP (after becoming a green photon) and gives an impression, as if, it has been originated from the source $S_2$ and moved un-deviatedly. Similarly, the photon-2 follows path *AQ* and appears, as if originated from the source $S_2$. The wavelengths of two light beams thus appear unaltered. This is in accordance with the observed wavelengths of light (figure 1) beams after crossing point 'A'. Thus fig. 3 (demonstrating the varied wavelengths) turns out to be exactly like fig. 1. *The observed (fig.1) unchanged wavelength is thus shown to be a result of photon's collision with each other.*

Consequently, the observation of non-variation of light beam's wavelengths after passing from the converging point is explained on the basis of particle characteristics of light; by the photon collision. Moreover, it has been observed, that although the photons after collision could have passed through any arbitrary angle $\varphi$ between them, they have chosen only a specific angle termed interchange angle for which their collision is perfectly elastic; which is indeed the observed angle (equation 1).

## 3. Conclusion

The observation of light beam's unaltered wavelengths after passage from the crossing point is explained on the account of its particle characteristics. When two photons collide at the intersection point of two light beams, their momentum gets exchanged. Thus, the observation of non-variation of wavelengths and un-deviated passage of light beams meeting at a point is explained by taking into account the particle characteristics of light. Thus, the behavior of light, in this situation is explainable on the basis of both the particle and wave aspects of light. Conclusively, the explanation of behavior of light on account of its both wave and particle characteristics suggests that the behavior of light in other known situations may also be explainable on the basis of both wave type and particle type personality of light.

## Acknowledgment

Authors gratefully acknowledges the initiatives and support from TIFAC-Center of Relevance and Excellence in Fiber Optics and Optical Communication at Delhi College of Engineering, now DTU, Delhi under Mission REACH program of Technology Vision-2020, Government of India. Authors also like to thank SPIE DCE Chapter, Delhi for all help and facility provided to complete this research work in its current form.

## References

Afshar, S. S. et al. (2007). Paradox in Wave-Particle Duality. *Found. Phys.* 37, 295, http://arxiv.org/abs/quant-ph/0702188

Beiser, A. (2007). *Concepts of Modern Physics*. New Delhi: Tata McGraw-Hill, (chapter 2)

Compton, A. H. (1923). A Quantum Theory of The Scattering of X-rays by Light Elements. *Phys. Rev.* 21, 483

Einstein, A. (1905). On a Heuristic Viewpoint Concerning the Production and Transformation of Light. *Ann. Phys.* 17, 132-48

Feynman, R. P. (2003). *The Feynman Lectures on Physics*. New Delhi: Addison-Wesley, (Chapter-17).

Ghatak, A. (2010). *Optics*. New Delhi: McGraw-Hill Companies (Chapter-2).

Ghatak, A. & Lokanathan, S. (2008). *Quantum Mechanics*. Delhi: Macmillan India Ltd. (chapter 2).

Maxwell, J. C (1865). A Dynamical Theory of the Electromagnetic Field. *Philos. Trans. R. Soc. London.* 155, 459-512

Raman, C. V. & Krishnan, K. S. (1929). The Production of New Radiations by Light Scattering. *Proc. Roy. Soc.* 122, 23

Verma, H. C. (2006). *Concepts of physics*. Patna: Bharti Bhawan Publishers (chapter-9).

**Author's Summary**

**Himanshu Chauhan** was born at Delhi, India in Nov., 1992 and completed his 10+2 studies in Non-Medical Science from C.B.S.E, Delhi. He is currently pursuing his Bachelor's degree in Technology in the field of Mechanical Engineering from Delhi Technological University (formerly DCE), Delhi. Past since 2.5 years he is pursuing research at the Applied Physics Department, TIFAC-CORE, Mission REACH program of Technology Vision-2020, Delhi Technological University, Delhi, and has authored two research articles for international conferences and one research article in Journal. He is also a member of Optical Society of America (OSA) and SPIE-The International Society for Optical Engineering. His current research areas include Quantum Mechanics, Tachyon Particles and Photonic Crystal Waveguides for MEOMS and NEOMS.

Swati Rawal has been awarded the PhD degree in 2011 from Delhi College of Engineering, Faculty of Technology, University of Delhi, India. Her current research interests include photonic crystal waveguides, devices and slow light in photonic crystals. She is recipient of "SPIE Best Research Paper Award" for her research paper related to slow light in photonic crystals during the Internal Conference on Optics and Photonics- 2009 held at CSIO Chandigarh, India. She is also recipient of student leadership award from Optical Society of America for participating in the international conference "Frontiers in Optics" held in San Jose, CA, USA during October 10 to 15, 2009.

**Ravindra K. Sinha** received his MSc degree in physics from the Indian Institute of Technology (IIT), Kharagpur, India, in 1984, and his PhD degree in fiber optics and optical communication from the IIT, Delhi, India, in 1990. Later he worked on various research and academic positions during 1990 and 1998. He is currently professor and head of Applied Physics Department, and the chief coordinator of the TIFAC–CORE in Fiber Optics and Optical Communication at Delhi Technological University (formerly DCE), Delhi. He is the author or co-author of more than 180 research publications in the leading national and international journals and conference proceedings and chapters in books. He was awarded Emerging Optoelectronics Technology Award [(CEOT-IETE, India)]-2006 for outstanding research work in the area of nanophotonics, S. K. Mitra Memorial Award for in Best Research Paper in IETE Technical Review 2002. He is a fellow of the IETE (India), Faculty Adviser of SPIE-DCE Chapter and OSA-DCE Chapter at Delhi Technological University. He has been academic visitor of several leading university in USA,UK, Japan, Switzerland and Taiwan.

**Figures**

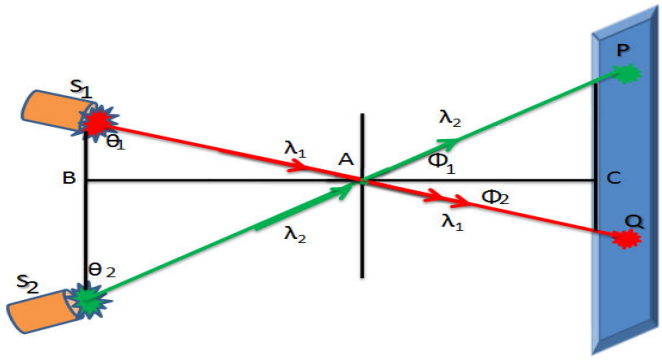



**Figure 1 The observed non-deviation of light beams from their incident directions converging at a point.**

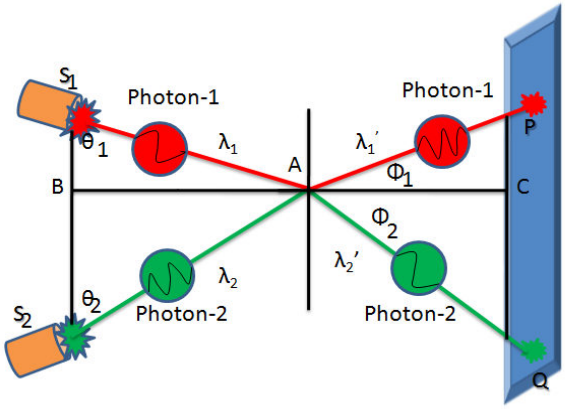



**Figure 2 Assuming the occurrence of photon collision at the converging point of two light beams, resulting in their deviation and wavelength variation.**

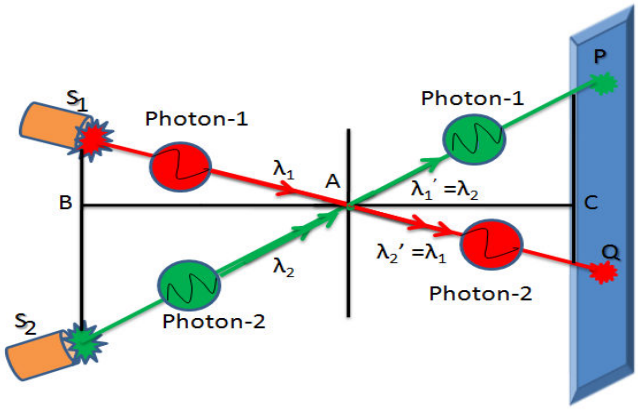



**Figure 3 Photon collision at point A, resulting in the wavelength interchange of the two light beams**

# SOFTWARE ARCHITECTURE BASED REGRESSION TESTING

Harsh Bhasin
Delhi technological university, Delhi
i_harsh_bhasin@yahoo.com

Ankush Goyal
AITM, Laboratory, Address,
Palwal, Haryana, 121102, India
Ankush49892@gmail.com

Deepika Goyal
AITM, Laboratory, Address,
Palwal, Haryana, 121102, India
Deepika.goyal256@gmail.com

**Abstract**

Software architecture plays a significant role in development of a dependable system. The purpose of regression testing is to make the system fault tolerant. The amalgamation of these two, results in the development of a robust system. The earlier works uses the conformance technique to instill confidence on implemented system with code, architecture and behavior but has not considered many parameters. The present work includes the concept of software architecture, system behavior and regression testing to propose a new framework which is sure to reduce gaps in the present frameworks and thus improve the system reliability.

***Keywords*:** Software Architecture; Dependable Systems; Regression Testing; Architecture-Based Analysis and Testing.

## 1. Introduction

The advent of object oriented languages and newer design methodology brought concepts like data abstraction to the fore-front of software development. The concepts initially labeled as obscure, later went on to become the crux of the development process. They also helped in clearly defining the components and depicting the interaction between them. This was important as many studies blamed these interactions as the major cause of failures [1]. To understand the system, we need to understand the parts of system as well; though abstraction hides the inner details, the communication between units remains essential to bring out the errors.

As many studies suggested, testing the system on the basis of test cases generated by unit testing, does not always reliably test the overall system. To install the confidence in the robustness of the system, it is essential to have test cases which are better than the test cases of unit testing. Betterment is defined in terms of testing the flow of data from one unit to another. As per the literature review, the type of tests used to test the system before 2000 were majorly based on the behavior of units rather than the system. Therefore, such methodologies were inapt to handle interaction problem. In such situations software architecture comes to our rescue. Software architecture helps us to detect problems which cannot be detected via conventional tests. As discussed earlier the designing of software plays a pivotal role in helping achieve the above. The system level and these segment level abstraction helps to untangle the knots created due to faults in communication.

As per the review carried out, software architecture helps in creating early test and hence reduces the cost of testing. The work presented intends to amalgamate the virtues of software architecture along with the concept of regression testing to propose a technique which is robust and better then the techniques proposed till now.

The paper is organized as follows. The second section of the paper presented the concepts of the literature review, the third section presents the premises of the paper and its goals, the fourth section proposes the technique. The last section presents the conclusions.

## 2. Related work

An extensive review was carried out in order to understand the intricacies of the techniques which we wish to use. The following section gives a brief overview of the techniques studied.

According to Muccini [1], testing is needed to increase the dependability of the system. The previous techniques have shown the application of conformance testing to achieve the confidence on the implemented system in

their expected behavior and architecture level. The work by H.Muccini [1] applied regression testing at software architecture level to reduce the cost of retesting the modified system. It may also be noted that, SA based analysis methods can be used in various scenarios like deadlock detection, performance analysis, component validation. SA based testing method check conformance of the implementation behavior and compare it with SA level specifications of expected behavior. Muccini [1] provides a technique to reuse previous information to get conformance of modified implementation with respect to the modified or initial architecture.

According to work by D.S. Rosenblum [11], The regression testing for analyzing the system during its life time is very expensive. The interaction between the components, however, can help us to reduce this cost. For a software system, the aim of Selective regression testing strategies to choose subset of test cases from previously run test cases, based on information about the changes made to the system to create new versions. In the paper some computationally efficient predictors of the cost-effectiveness of the two main classes of selective regression testing approaches are presented. In the work proposed by Mary Jean Harrold [2], which is based on the specification of software architecture, it was observed that the amalgamation of SA proves very effective in terms of cost and time. The software architecture testing focuses on the cost of the software while other techniques focus only on development.

Another work by M. J.Harrol [2], provides an approach that uses the integration of code level regression testing with architecture based regression testing. It uses the selective testing for the architecture based regression testing and code level regression testing. It is based on comparing nodes of the two graphs, where the first graph represents the program and the second one represents the modified version of original program. The work provides a novel approach for regression testing at architecture level by comparing both graphs.

The work by A. Bertolino [14] also affirms the use of Software Architecture in testing. The work also confirms the effect of SA on implementation. The paper proposes the extended approaches to SA bases testing. It shows how a architectural style conform the mapping among SA based and code based test cases. According to the work, Software Architecture can be used for code conformance testing and to check if implementation fulfills to its specification at the SA level. This paper extends the previous approaches to software architecture based testing and how a specific architectural style which supports implementation and facilitates the mapping among SA-based and code-based test cases can be used to deliver a completely systematic SA-based testing approach.

## 3. Background

### 3.1. *Premises of the paper*

The proposed technique is based on the concept of Muccini [1]. In the work, Software Architecture Specification (SAP) has been taken as base and the behavioral model of the software serves as a test oracle. As per the above work, topology is described in terms of components, connectors, and configurations. This is followed by the application of SA behavioral.

The above is followed by seeing SA in a way so that non relevant actions are hidden. This helps in the abstraction of state machine based model. This generates what is referred to as, Architecture Level Test Cases (ALTC), which is based on the audit sequence of events. So as to accomplish the above task the mapping function is used which maps SA level function tests to code level tests.

### 3.2. *Goals*

The goals that are to be accomplished by the proposed technique are as follows:

The first goal is to use the existing implementation-level test cases test the conformance between modified code and the architectural specifications. It is to test the conformance of a program with respect to the system, while reusing previous test information for selective regression testing, thereby reducing the test cases.

The second goal is to reuse architecture-level test cases to test the conformance of the source code with respect to the evolved software architecture.

The first goal is accomplished by generating a control graph and comparing the previous graph with the new graph. So as to facilitate the task, the information so as to how a graph is traversed is stored. Test cases selection for the new program P' with the help of test history and graph comparison, is then carried out. The concept relies on integrating code-level regression testing with architecture-based regression testing. Selective testing technique is used for code-level regression testing and for architecture-based regression testing also. The test cases are generated on the basis of software architecture. The expected behavior and the history is then mapped with the output. In the conformance testing we generate graphs for both the program and the changes in the programs. On the basis of comparisons between the old graphs and new graphs, the changes are recorded. The techniques developed earlier are not suitable because it is possible that changing the program might not result in the change in the graph. There can be many more such scenarios. If there is a change then that change may

affect the software architecture but not the behavior. Moreover, it is also possible that change in one part of the program may effect on other part of program due to coupling [15]. Previous technique has neglected the above issues.

The work intends to, merge software architecture concepts with a concept of regression testing to propose a novel technique. The technique is being tested by a set of programs. In the proposed technique, the software architecture of the program will be studied and the purposed technique will be used to find errors. The concept will be compared with a technique purposed in the work by Muccini [1]. A set of 20 programs is selected, divided into three categories small, medium and large programs.

## 4. Purposed Technique

The section proposes a new framework which clubs together the best of worlds, regression testing and software architecture. The proposed work is different from the base work since it also takes into account the behavior and not just the abstract model.

### 4.1. *Software architecture based regression testing with behavioral blent*

*Step1: SA specification.* SA based conformance testing start from behavioral specification of SA and topology of SA. In the SA structure topology describes the components, connectors and configurations. For describing the topology we use ADL (Architecture description languages) and TS (transition system) is used for describing the SA behavior.

*Step2: Testing Criterion.* There are many events in model and sequence of these events is defined as ATC (architecture-level test cases). Designing architecture level test cases calls for checking the specifications and also checking the behavior of the model. In this step the portions, which are to be tested, are identified. The interactions between the levels is also seen.

*Step 3: Test Cases.* Changing or adding an extra component can change the behavior. So, a mapping function is to be used for mapping the SA-level test case to code level test cases. This also checks the behavior of the model.

*Step 4: Test Execution.* Since many test case have been generated, now the next step is to check the result of each test case and map the corresponding results.

The model is depicted in the following figure.

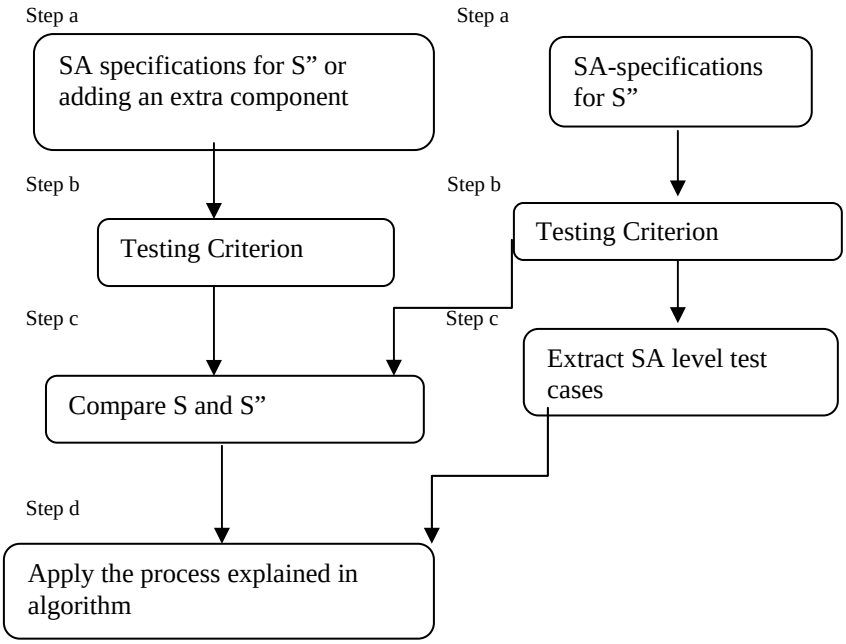



Fig. 1 Diagram of SA based regression testing

## 5. Conclusions

The work presented above adds a sprinkle of behavioral model to the Software Architecture based Regression Testing. The technique is being applied on selected set of programs. The programs have been selected in such a way that the technique can be verified by all the categories. The programs selected therefore are professional applications and some algorithm implementation. The application of the above technique to these programmers will instill the confidence on the technique.

It may be noted that the above work re-affirms the fact that the exclusion of behavior from the model, is not justified. The complete verification and validation will instill the confidence in the technique.

## References


[1] Henry Muccini, Marcio Dias, Debra J. Richardson. Software Architecture-Based Regression Testing, 10 February 2006. In: The Journal of Systems and Software 79 (2006) 1379–1396.

[2] M. J. Harrold. Architecture-Based Regression Testing of Evolving Systems. In Proc. Int. Workshop on the Role of Software Architecture in Testing and Analysis (ROSATEA), CNR-NSF, pp. 73-77, July 1998.

[3] An Andreas Johnson. Architecture –Based Verification Of Software- Intensive Systems. Mälardalen University School of Innovation, Design and Engineering 2010 March 8,Västerås Sweden.

[4] Henry Muccini, Marcio S. Dias, and Debra J. Richardson. Towards software architecture-based regression testing. SIGSOFT Software. Eng. Notes, 30(4):1_7, 2005.

[5] Badri H.S. Systematic Software Architecture Based Testing Approach – A Case Study. In International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 9, September 2012.

[6] B. Beizer. Software Testing Techniques. Van Nos-trand Reinhold, New York, NY, 1990.

[7] A. Bertolino and P. Inverardi. Architecture based software testing. In Proc. of Int'l Software Architecture. Workshop, pages 62{64, October 1996.

[8] H. Leung and L. White. A cost model to compare regression test strategies. In Proc. of Conf. on Software. Maint. Pages 201{208, Oct. 1991.

[9] D. Richardson, J. Stafford, and A. Wolf. A formal approach to architecture based software testing. Technical report, University of California, Irvine, 1998.

[10] D. Richardson and A. Wolf. Software testing at the architectural level. In Proc. of Int'l Software. Arch. Workshop, pages 6{70, October 1996.

[11] D. S. Rosenblum and E. J. Weyuker. Predicting the cost-effectiveness of regression testing strategies. In Proceedings of the ACM SIGSOFT '96 Fourth Symposium on the Foundations of Software Engineering, Oct. 1996.

[12] H. Muccini, M. Dias, and D. Richardson. Systematic Testing of Software Architectures in the C2 style. Extended version of the ETAPS 2004 publication.

[13] A. Bertolino and P. Inverardi. Architecture-based software testing. In Proc. ISAW96, October 1996.

[14] A. Bertolino, P. Inverardi, and H. Muccini. An Explorative Journey from Architectural Tests Definition down to Code Tests Execution. In IEEE Proc. Int. Conf. on Software Engineering, ICSE2001, pp. 211-220, May 2001.

[15] Harsh Bhasin, Manoj. Regression Testing Using Coupling and GeneticAlgorithms.IN (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (1), 2012, 3255 – 3259

[16] Perry, D.E., Wolf, A.L., 1992. Foundations for the Study of Software Architecture 17 (4), 40–52.

[17] Muccini, H., Bertolino, A., Inverardi, P., 2003. Using software architecture for code testing. IEEE Transactions on Software Engineering 30 (3), 160–171.

[18] Bosch, J., 2000. Design and Use of Software Architectures: Adopting and Evolving a Product-Line Approach. ACM Press/Addison-Wesley Publishing Co.

[19] Bernardo, M., Inverardi, P., 2003. Formal Methods for Software Architectures, Tutorial Book on Software Architectures and Formal Methods. SFM-03: SA Lectures, LNCS, vol. 2804.

# Spectrum Management Models for Cognitive Radios

Prabhjot Kaur, Arun Khosla, and Moin Uddin

***Abstract:*** **This paper presents an analytical framework for dynamic spectrum allocation in cognitive radio networks. We propose a distributed queuing based Markovian model each for single channel and multiple channels access for a contending user. Knowledge about spectrum mobility is one of the most challenging problems in both these setups. To solve this, we consider probabilistic channel availability in case of licensed channel detection for single channel allocation, while variable data rates are considered using channel aggregation technique in the multiple channel access model. These models are designed for a centralized architecture to enable dynamic spectrum allocation and are compared on the basis of access latency and service duration.**

***Index Terms:*** **Cognitive radio (CR), dynamic spectrum allocation (DSA), Markov models, spectrum management, queuing models.**

## I. INTRODUCTION

In an effort to increase spectrum utilization, cognitive radios (CR) came out as one of the promising approaches which also helped in mitigating spectrum scarcity problem. Development of air interface has been the biggest challenge for the success of a CR relying on dynamic spectrum allocation (DSA) technique as the key enabler [1], [2]. A medium access control (MAC) protocol enables efficient sharing of common communication channel among a group of users. In wireless communication systems, multiple users transmit different types of traffic over a shared medium. In CR networks, MAC becomes more crucial due to the coexistence of two different sets of users. The licensed set of users of a particular band is called the primary users (PU) and the unlicensed set of users who operate on the unused or underutilized frequency bands are called secondary users (SU). These unlicensed users must obey a set of rules to manage the medium access in order to obtain the maximized channel throughput while minimizing the presence of collisions and interference with PU. In past one decade, many MAC protocols have been proposed e.g., in CR networks [3]–[8]. A partially observable Markov decision process (POMDP) [3] was modeled for spectrum sensing and access in a spectrum overlay system. The authors focused to maximize the total number of transmitted bits while maintaining the collision probability below a threshold level. POMDP model was then extended to decentralised cognitive MAC protocol. This protocol has been shown to perform better sensing and access as compared to the random access techniques. However, this protocol requires global synchronization which gives a very low throughput when collision probability is high and the number of channels to be sensed per time slot is not optimized. An opportunistic cognitive MAC protocol is proposed in [4] which works for TV bands as per IEEE802.22. However, it requires a regular radio receiver in addition to a CR transreceiver. A multi channel MAC is proposed in [5] which is focused on global system for mobile communication (GSM) only. Statistical channel allocation MAC (SCA MAC) [6] is a multichannel carrier sense multiple access with collision avoidance (CSMA/CA) based protocol. SCA MAC consists of three phases: (a) Environment sensing and learning phase where the SU sense the spectrum continuously and periodically, (b) CRTS/CCTS exchange phase to determine the best channel available for communication in the control channel. Here, CRTS denotes control channel-request-to-send and CCTS denotes control-channel-clear-to-send, (c) data transmission phase. Based on channel statistics, SU select the channel with highest successful transmission probability. However, this protocol adds computational complexity. Additionally, these protocols [3]–[6] developed for CR spectrum management do not address the delay analysis, which is one of the crucial quality of service (QoS) parameters for next generation networks. There are some models and protocols [7]–[11] proposals that have been evaluated on the basis of access delay and service times. Dynamic open spectrum sharing [7] is proposed for adhoc cognitive radio environment which allows SU to adaptively select an arbitrary frequency band. An analytical model was developed to study this proposed protocol which was used to obtain the capacity of unlicensed user working on PU band. Access delay was also calculated. However, it uses three frequency bands: A busy tone band, a control channel, and a data band and thus requires three sets of transreceivers which proves to be quite costly. Also, this protocol was designed with local sensing. The performance can be improved by using distributed/collaborated sensing. Some of other works evaluating access delay [8]–[11] do not generalize the DSA model, rather are based on specifications of the developed protocol under special circumstances. Through this work, we tried to develop more generalized analytical mathematical models to closely match the DSA based spectrum management for CR network.

We designed separate models both for single channel and multiple channel access schemes where access latency and service delay are considered as evaluation parameters of these models. Spectrum mobility is also considered to incorporate the detection of PU on the allocated spectrum channel to SU. We differentiate our work from the previous state of art by modeling the spectrum management for the purpose of comparing both types of channel allocation schemes i.e., single channel versus multi channel. Thus the uniqueness of our work can be highlighted as below

- Modeling the spectrum management for a centralized CR ar-

Manuscript received March 13, 2012; approved for publication by Kai-Kit Wong, Division I Editor, November 8, 2012.

P. Kaur is with EECE Department, ITM University, Sector 23 A, Gurgaon, India 122017, email: prabhjotkaur@itmindia.edu.

A. Khosla is with the Electronics and communication Department, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, Punjab, India, email: khoslaak@nitj.ac.in.

M. Uddin is Professor and Pro-Vice Chancellor at Delhi Technological University, Bawana Road, Delhi-42, India, email: provc@dce.edu.

Digital Object Identifier 10.1109/JCN.2013.000036

chitecture to compare both the channel allocation schemes: Single channel allocation and multiple channel allocation.
- Consideration of spectrum mobility for single channel allocation
- Channel aggregation technique for multi channel allocation

These proposals have been designed and analyzed using distributed queuing based Markovian models.

## II. MEDIUM ACCESS CONTROL MODELS

We assume that for each licensed channel, primary users (PU) follow an independent and identically distributed ON/OFF renewal process as shown in Fig. 1(a). If the channel is occupied by PU, it is in ON state and if it is not occupied by PU, it is in OFF state. Thus, any channel can be occupied by SU in OFF state. Utilization for a licensed channel can be represented as an ON/OFF transition state model as shown in Fig. 1(b). Hence, the utilization of a licensed channel can be defined as the probability that the channel is in ON state as follows

$$p_i = \frac{T_{\mathrm{ON}}}{T_{\mathrm{ON}} + T_{\mathrm{OFF}}} \tag{1}$$

where $T_{\mathrm{ON}}$ and $T_{\mathrm{OFF}}$ are the mean times of ON and OFF states, respectively. We consider a centralized architecture for CR network in line with the IEEE 802.22 standard proposal [12] with a base station (BS) of this network as the central controller of PU network. To overcome the hidden terminal problem, we consider distributed spectrum sensing with SU broadcasting this channel status information (CSI) among all SU and BS in the network. We propose to use common control channel as unaided rendezvous where the radios are left to their own to find a common spectrum for multiple channel access [13]. In this paper, we assume the control channel to be always available and do not include the process of finding the control channel as part of our spectrum management models. BS receives the CSI from all surrounding CR and prepares a database of all unused frequencies and refers it when it has to allocate channels to the contending SU.

We model the functionalities of media access at BS using distributed queuing where the first queue is equivalent of sorting overlapped detections received from all SU in the network, updating its database, contention and conflict management. We call this queue as spectrum analysis queue (SAQ). This queue receives the channel requests from SU at an arrival rate of $\lambda_1$. The requests once sorted will be considered for channel allocation which is modeled as channel allocation queue (CAQ).

### A. Single Channel MAC Model

For the single channel allocation scheme, the following algorithm is proposed.

1. Every SU senses channels in its vicinity
2. SU Controller collects the sensing results and updates its log of spectrum holes
3. Channel requests reach BS and Channels allocated to SU
4. Check if PU activity is detected on the allocated channels; continue transmission if no activity
5. If activity is detected; check spectrum log and allocate new channel if spectrum holes are available

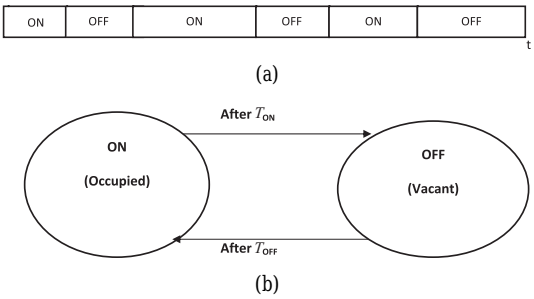



Fig. 1. Licensed channel representation: (a) Channel utilization of a licensed channel and (b) ON/OFF channel model for a licensed channel.

6. Else, terminate connection; request treated as new channel request

Due to the consideration of allocating only one channel to SU at any given time, it is referred to as single channel distributed queuing based medium access control (SCDQMAC) model. This model is shown in Fig. 2(a). Whenever PU activity is detected on the channel being used for SU transmission, two different cases are considered with their probability that (a) SU waits for the channel to be freed by PU and then resumes transmission and (b) coordinator allocates another channel to SU (handover).

Assuming that an unoccupied channel is allotted to SU and in the ongoing conversation of SU, BS detects PU activity on the same channel. PU may be detected from the results of detected spectrum holes from surrounding SU in subsequent clock cycles. This will act as new request for CAQ, arriving at a rate of $\lambda_2$. Both input traffic flows at the different queues from outside are considered Poisson distributed.

#### A.1 Latency Evaluation

As we know PU must have prioritized access for its licensed channel, we consider two cases:

*Case a:* BS has an extra channel to be allocated to SU. In this case, SU will be allocated another channel, and we consider the requests go back to CAQ. Thus, the call will be on hold during spectrum handover.

*Case b*: BS has no spectrum holes stored in its database. In this case, the ongoing call will be terminated and the request is again stacked as the new request in SAQ.

Thus, at the output of CAQ there is a random splitting with probability $p$ if the serving request is a feed back to SAQ and with probability $q$ if it is the feed back to CAQ, where, for both $p$ and $q$, $0 < p$, $q < 1$. Lastly we assume the message service times are independent for the two queues and exponentially distributed with the same mean rate as $\mu$.

As per the assumptions for this network, conditions of the Jackson theorem [14], [15] are fulfilled so that SAQ can be studied by means of the $M/M/1$ model with the arrival rate $\Lambda_1$ and CAQ can be characterized by $M/M/1$ model with mean arrival rate $\Lambda_2$. Thus, using Jackson theorem, we can write mean rate equation for each sum point in the network as

$$\Lambda_1 = \frac{\lambda_2 p + (1-p)\lambda_1}{1-p-q}, \quad \Lambda_2 = \frac{\lambda_1 + \lambda_2}{1-p-q}. \tag{2}$$

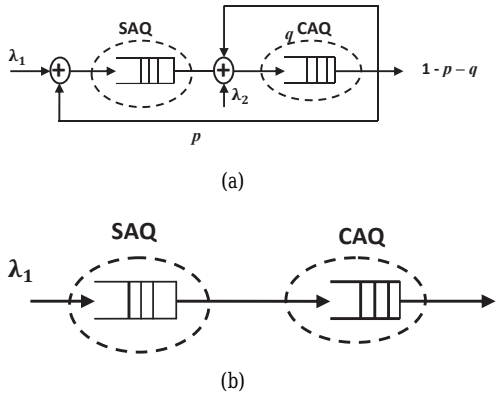



Fig. 2. Queuing models for: (a) Single channel MAC, SCDQMAC and (b) multi channel MAC, MCDQMAC.

Using classical theory for $M/M/1$ queuing, (2) and Little's theorem, the total access latency of this system, $T_{\text{single}}$ is calculated as below

$$\frac{1}{\lambda_1+\lambda_2}\left(\frac{\lambda_2+p+(1-p)\lambda_1}{(1-p-q)\mu\left(1-\frac{\lambda_2 p+(1-p)\lambda_1}{(1-p-q)\mu}\right)}\right). \tag{3}$$

A.2 Service Delay Evaluation

Service delay, $T^S_{\text{single}}$ for SCDQMAC may be given as sum of access delay, transmission time allowed to SU $T_{\text{trans}}$, and total handover time spent during the transmission.

$$T^S_{\text{single}} = T_{\text{single}} + T_{\text{trans}} + \overline{H}\cdot T_{HO} \tag{4}$$

where $\bar{H}$ is the average number of handovers and $T_{HO}$ is average handover duration for assigning a new channel on detection of PU activity on the channel. Considering $\bar{H}$ as average number of PU arriving during SU transmission ($\lambda_2\ T_{\text{trans}}$), $T_{HO}$ as the time during which channel is occupied by PU and using (1) for $p_i$, (4) is obtained as

$$T^S_{\text{single}} = T_{\text{single}} + T_{\text{trans}} + \frac{p_i}{1-p_i}T_{\text{trans}}. \tag{5}$$

*B. Multi Channel MAC Model*

Considering that $L$ channels can be sensed simultaneously by each SU and using channel aggregation technique, this model allows the BS to allocate multiple channels to contending SU for their packet transmission. Channel aggregation tends to facilitate higher transmission rates due to wider available bandwidth for transmission. If PU activity is observed on any of the channels allocated to SU, the number of channels are continuously adjusted which implies that the transmission rate cannot remain fixed and hence will vary with varying number of granted channels. Due to this feature of the model to allocate multiple channels to SU and the use of distributed queuing, we call it as multi channel distributed queuing based multiple access control (MCDQMAC) model.

B.1 Latency Evaluation

The queuing model for MCDQMAC is shown in Fig. 2(b). Due to the varying number of available spectrum holes with time, we consider the random channel utilization ($p_i$) of the licensed channels. As the data transmission rate for SUs will keep varying as per the available number of spectrum holes,the service capacity of CAQ will vary. Thus, CAQ in this case is modeled as $M/G^Y/1$ with utilization, $\rho_{CAQ}$. Due to ability of an SU to occupy multiple unused channels, the number of channels or in turn the data rate will be varied in case of PU activity detection on operating channel, rather than waiting for another channel here in this model in contrast to SCDQMAC. Support for spectrum mobility will be provided by the central allocator and waiting time for channels will not impact the performance of SU or in turn will not affect the calculations of access latency of an SU. Thus, the reason for not considering the feedbacks and arrival of PU at sum point for CAQ is obvious. For the analysis of MCDQMAC model, SAQ is analyzed using classical $M/M/1$ theory. Analysis of CAQ is based on $M/G^Y/1$ model following the works in [10]. We assume $L$ as the number of contending SU for channel allocation, ($\bar{y}$) as average number of packets that an SU can send during a service, $\varphi$ as the SU's random number of arrived packets during an ongoing service. Let $y_i$ be the pmf that an SU sends $i$ packets during a service at the equilibrium state which is given as

$$y_i = \lim_{t\to\infty} P(y_t = i) \tag{6}$$

where $y_t$ is the service capacity at any given time $t$.

The total access delay for MCDQMAC, $T_{\text{multi}}$ is given as

$$\begin{aligned} T_{\text{multi}} = \frac{1}{\lambda_1}\Bigg[&\frac{\lambda_1}{\mu-\lambda_1} + \frac{\lambda_1^2 T_{\text{trans}}^2\varphi^2(1)+\phi_0^2(1)}{2\overline{y}(1-\rho_{CAQ})} \\ &+ \frac{1-L+\rho_{CAQ}(L-\rho_{CAQ}\overline{y})}{1-\rho_{CAQ}} \\ &+ \sum_{i=1}^{L-1}(1-z_i)^1 - \overline{y}\rho_{CAQ} \\ &+ \frac{\lambda_1^2 T_{\text{trans}}^2\varphi^2(1)}{2\overline{y}(\rho_{CAQ})}\Bigg]. \end{aligned} \tag{7}$$

The functions $\varphi^2(\cdot)$ and $\phi_0^2(\cdot)$ are second derivatives of probability generating function (PGF) $\varphi$ (z) of probability P ($\varphi = j$ arrived packets) and PGF $\phi_0(z)$ for $y_i$, respectively. Please see Appendix for proof of (7).

B.2 Service Delay Evaluation

Service delay for MCDQMAC, $T^S_{\text{multi}}$ can be given as

$$T^S_{\text{multi}} = T_{\text{multi}} + \overline{S}T_{\text{trans}} \tag{8}$$

where $\bar{S}$ is average number of successful transmissions, calculated using [10]. Thus, $T^S_{\text{multi}}$ can be written as

$$T^S_{\text{multi}} = T_{\text{multi}} + (\rho_{CAQ}(N_1+1)-1)T_{\text{trans}} \tag{9}$$

where $N_1$ is the number of contending SU for spectrum access.

## III. NUMERICAL ANALYSIS

The design considerations of our proposed models are given in Table 1. For analyzing these models, we consider the number

Table 1. Considerations for design parameters of proposed models.

| Parameters considered | SCDQMAC | MCDQMAC |
|---|---|---|
| Architecture | Centralized | Centralized |
| OSA equivalence | Yes | Yes |
| Limited spectrum holes | Yes | Yes |
| PU prioritization | Yes | Yes |
| SU blocking if no spectrum available | Probabilistic | NA |
| SU treated as new user in case of spectrum mobility/no spectrum hole | Probabilistic | NA |
| Channel aggregation | No | Yes |
| Transmission rate | Fixed (Single channel) | Variable |

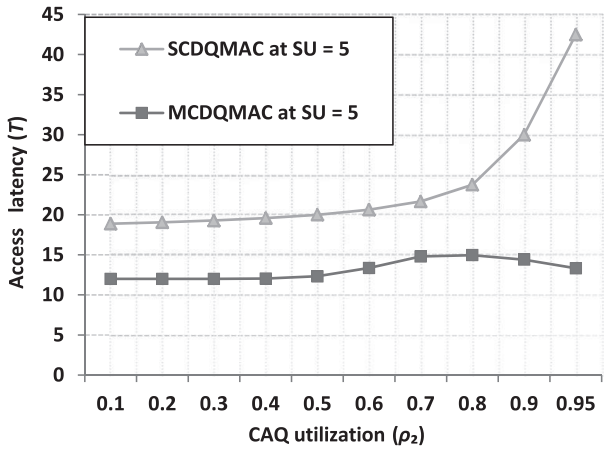



Fig. 3. Comparison of access latency for SCDQMAC and MCDQMAC for varying contending SU in the system.

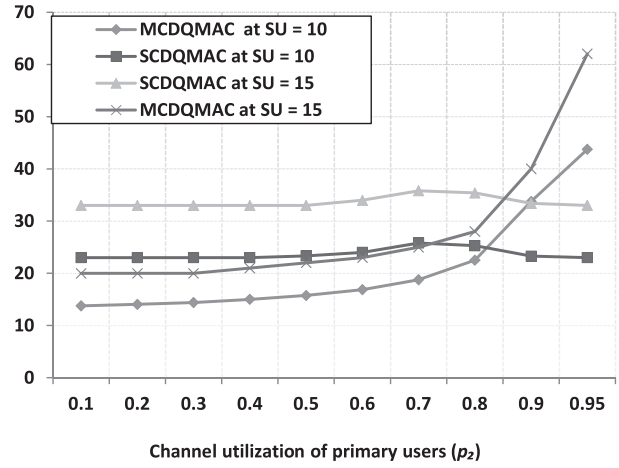



Fig. 4. Comparison of service delay for SCDQMAC and MCDQMAC for varying licensed PU channel utilization while contending SU in the system is 10 and 15.

of licensed channels, $L$ as 10, transmission time, $T_{\text{trans}}$ of SU per frame as 1 ms, and PU arrival rate as 0.5. The comparison of SCDQMAC and MCDQMAC models has been done on the basis of access delay and service duration.

### *A. Access Delay Comparison of the Models*

To compare single channel and multi channel MAC schemes, the access delay is analyzed for both the schemes using (3) and (7). For (3), the values of $\lambda_1$, $\lambda_2$ are taken as 0.3 and 0.5, respectively while the number of SU contending for spectrum access in the system, $N_1 = N_{SAQ}$ is 5. For analyzing (7), the average number of packets that SU can send is given as, $\bar{y} = L(1 - p_i)$. Also, from (13), we can write $\rho_{CAQ} = \lambda_1/(1 - p_i)$. Thus, all the terms in (7) can be represented in terms of $\lambda_1$, $p_i$, $L$, $T_{\text{trans}}$ and a value of $z_i$ for a stable system. The given distribution of this term suggests that the system is stable for all values of $z$ except at zero. Considering, $L$ as 10, $z = 0.1$, varying the PU channel utilization from 0.1 to 0.95, $\lambda_1 = 0.5$ and putting these values for a number of contending SU in the system as 5, we obtain the plots for $T_{\text{single}}$ and $T_{\text{multi}}$ as shown in Fig. 3.

It is clear that delay for a given SU in getting an access to spectrum is less for multi channel than for single channel allocation scheme. As the load on channel allocation process increases, i.e., beyond the CAQ utilization of 0.7, this difference becomes even more significant. Thus we can say that irrespective of the amount of spectrum availability with the central channel allocator (represented by CAQ here) or say utilization by PU, MCDQMAC outperforms SCDQMAC in regards to the access latency.

### *B. Service Duration Comparison of the Proposed Models*

Next, we obtain a comparison of service delay between SCDQMAC and MCDQMAC with respect to variation in utilization of PU licensed channels using (5) and (9). For analyzing the service delays encountered by these models, we consider the respective values of $T_{\text{single}}$ and $T_{\text{multi}}$ corresponding to different values of $p_i$ that varies from 0.1 to 0.95 with $T_{\text{trans}}$. These results are obtained for two different values of contending SU in the system as 10 and 15 users and are given in Fig. 4. As shown, the total end to end delay or service delay for SCDQMAC is more than MCDQMAC. Irrespective of the model, this delay does not increase sharply with increase in PU channel utilization. However, beyond the value of $p_i$ as 0.8, service delay for MCDQMAC increases rapidly. This may be because of lesser opportunities available with SU for its own packet transmission at higher licensed channel utilization. Thus, we can say that higher the number of spectrum holes with the controller, better performance will be driven for CR devices because of the controller's ability to select multiple channels (which will lead to follow MCDQMAC). However at high PU channel occupancy, single channel allocation technique proves to be more

efficient. Further, the abilities of orthogonal frequency division multiple access (OFDMA) enabling every single unused carrier frequency to be used as multiple channel allocation can be exploited to favor MCDQMAC.

## IV. CONCLUSION

We developed the analytical models for single channel and multi channel spectrum allocation schemes. The analysis shows that multichannel MAC is better than single channel spectrum allocation scheme both in regards to access delay and service delay. However, this trend deviates for higher values of system utilization or higher channel utilization of PU. MCDQMAC model considers channel aggregation technique. However the spectrum holes available with the controller may not be continuous. Thus, the use of OFDM technique is suggested. This work is further extendable to explore end-to-end delay performance of packets for two communicating nodes and validate the utilization of these models in practical systems. These models may be enhanced to study the behavior of a comprehensive MAC protocol after incorporating a back off algorithm and conflict management. A channel selector/allocator can be designed for optimized resource allocation considering different traffic types, range of spectrum holes available and interference avoiding techniques among SU and PU.

As some of the licensed channels can be used by PU more frequently than others, the work presented in MCDQMAC can further be extended to model the selection of licensed channels. SU can be made to select those channels which are less occupied. This may reduce the sensing load on SU and will also reduce the number of handovers. Thus modeling the prioritized channel usage depending on its occupancy can be included.

## APPENDIX

Considering a $p$-persistent CSMA scheme, the probability that a contending SU can be allocated the unused channels successfully $p_s$ is inversely proportional to the number of SU,

$$p_s = 1/L \tag{10}$$

CAQ utilization $\rho_{CAQ}$ is given as

$$\rho_{CAQ} = \frac{\lambda_1}{\overline{y}} E[V] \tag{11}$$

where *E[V]* is average service time for an SU while *V* follows a geometric distribution with the pmf as follows

$$P[v = u] = \rho_s(1-\rho_s)^{(u-1)} = \frac{(L-1)^{(u-1)}}{L^u} \tag{12}$$

where $u = 1, 2, 3, \cdots$. Therefore, (11) can be written as

$$\rho_{CAQ} = \frac{\lambda_1}{y}\frac{1}{\rho_s} = \frac{\lambda_1}{y} L. \tag{13}$$

Probability that the number of arrived packets is *j* is given as

$$P[\emptyset = j] = \sum_{u=1}^{\infty} \frac{e^{-\lambda_1 u}(\lambda_1 u)^j}{j!}[\rho_s(1-\rho_s)^{u-1}] \tag{14}$$

where $\rho_s$ is defined in (10).

$$\emptyset(z) = \sum_{j=0}^{\infty} P(\emptyset = j) z^j = \frac{e^{\lambda_1(z-1)}}{L-(L-1)e^{\lambda_1(z-1)}}. \tag{15}$$

Let $P_j$ be the probability that CAQ has $j$ packets of an SU in equilibrium state i.e., $P_j = P(N_q = j)$ with $N_q$ is the number of packets in CAQ in equilibrium state. The PGF for $P_j$, $P_q(z)$ is given as

$$P_q(z) = \sum_{j=0}^{\infty} P_j z^j. \tag{16}$$

Now, the average number of SU packets in CAQ, $N_{CAQ}$, is the first moment of $N_q$.

$$N_{CAQ} = \frac{d}{dz} P_q(z) = \overline{N} - \overline{y}\rho_{CAQ} + \frac{\lambda_1^2 T_{\text{trans}}^2 \Phi^2(1)}{2\overline{y}\rho_{CAQ}} \tag{17}$$

where $\overline{N}$ is the average number of packets buffered in the system in equilibrium state which is given in [16] as

$$\overline{N} = \frac{\lambda_1^2 T_{\text{trans}}^2 \varphi^2(1) + \phi_0^2(1)}{2\overline{y}(1-\rho_{CAQ})} + \frac{1-L+\rho_{CAQ}(L-\rho_{CAQ}\overline{y})}{1-\rho_{CAQ}} + \sum_{i=1}^{L-1}(1-z_i)^{-1}. \tag{18}$$

Access latency for MCDQMAC can be calculated using Little's theorem as follows

$$T_{\text{multi}} = \frac{(N_{SAQ} + N_{CAQ})}{\lambda_1} \tag{19}$$

where $N_{SAQ}$ is average number of SU packets in SAQ which is calculated using the classical theory of *M/M/1* as follows

$$N_{SAQ} = \frac{\rho_{SAQ}}{(1-\rho_{SAQ})}. \tag{20}$$

Using (17), (18) and (20), the access latency for MCDQMAC, $T_{\text{multi}}$ of (19) can be written as follows

$$T_{\text{multi}} = \frac{1}{\lambda_1}\left[\frac{\lambda_1}{\mu-\lambda_1} + \frac{\lambda_1^2 T_{\text{trans}}^2 \varphi^2(1) + \phi_0^2(1)}{2\overline{y}(1-\rho_{CAQ})} + \frac{1-L+\rho_{CAQ}(L-\rho_{CAQ}\overline{y})}{1-\rho_{CAQ}} + \sum_{i=1}^{L-1}(1-z_i)^{-1} - \overline{y}\rho_{CAQ} + \frac{\lambda_1^2 T_{\text{trans}}^2 \varphi^2(1)}{2\overline{y}(\rho_{CAQ})}\right] \tag{21}$$

which is access latency for SU as per MCDQMAC model and is same as that of (7). Hence, completes the proof.

## REFERENCES

[1] M. Devroye, P. Mitran, and V. Tarohk, "Limits on communications in a cognitive radio channel," *IEEE Commun. Mag.*, vol. 44, no. 6, pp. 44–49, June 2006.

[2] T. A. Weiss and F. K. Jondral, "Spectrum pooling: An innovative strategy for the enhancement of spectrum efficiency," *IEEE Commun. Mag.*, vol. 42, no. 3, pp. 8–14, Mar. 2004.
[3] Q. Zhao, L. Tong, A. Swami, and Y. Chen, "Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: A POMDP framework," *IEEE J. Sel. Areas Commun.*, vol. 25, no. 3, pp.589–600, Apr. 2007.
[4] Y. Yuan, P. Bahl, R. Chandra, and P. Chou, "KNOWS: Cognitive radio networks over white spaces," in *Proc. IEEE Int. Symp. DYSPAN*, 2007, pp. 416–427.
[5] A. Mishra, "A multi-channel MAC for opportunistic spectrum sharing in cognitive networks," in *Proc. IEEE MILCOM*, Oct. 2006, pp. 1–6.
[6] A. Hsu, D. Wei, and C. Kuo, "A cognitive MAC protocol using statistical channel allocation for wireless ad-hoc networks," in *Proc. IEEE WCNC*, Mar. 2007, pp. 105–110.
[7] L. Ma, X. Han, and C. Shen, "Dynamic open spectrum sharing MAC protocol for wireless ad hoc networks," in *Proc. IEEE Int. Symp. DYSPAN*, Nov. 2005, pp. 203–213.
[8] E. W. M. Wong and Chuan Foh, "Analysis of cognitive radio spectrum access with finite user population, *IEEE Commun. Lett.*, vol. 13, no. 5, pp. 294–296, May 2009.
[9] L.-C. Wang, Y.-C. Lu, C.-W. Wang, and D. S. L. Wei, "Latency analysis for dynamic spectrum access in cognitive radio: Dedicated or embedded control channel?," in *Proc. IEEE PIMRC*, 2007, pp. 1–7.
[10] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 1, pp. 118–129, Jan. 2008.
[11] M. M. Rashid, J. Hossain, E. Hossain, and V. K. Bhargava, "Opportunistic spectrum access in cognitive radio networks: A queueing analytic model and admission controller design," in *Proc. GLOBECOM*, 2007, pp. 4647–4652.
[12] IEEE 802.22. *Working group on wireless regional area networks (WRAN)*. [Online]. Available: http://www.ieee802.org/22/
[13] N. C. Theis, R. W. Thomas, and L. A. DaSilva, "Rendezvous for cognitive radios," *IEEE Trans. Mobile Comput.*, vol. 10, no. 2, pp. 216–227, Feb. 2011.
[14] J. F. H. Hayes, *Modeling and Analysis of Computer Communication Networks*. Plenum Press, New York, 1986.
[15] Giovanni Giambene, *Queuing Theory and Telecommunications: Networks and Applications*. Springer, 2005.
[16] X. Zhang and H. Su, "CREAM-MAC: Cognitive radio enabled multi channel MKAC protocol over dynamic spectrum access networks," *IEEE Special Topics Signal Process.*, vol. 5, no. 1, pp. 110–123, Feb. 2011.

**Prabhjot Kaur** did her Masters of Engineering from Punjab University, Chandigarh and has submitted her Ph.D. thesis on MAC models for Dynamic Spectrum Access in Cognitive Radio Networks at National Institute of Technology, Jalandhar, India. She is currently working as Associate Professor in Electronics and Communication department with ITM University, Gurgaon, Haryana, India. Her research interests include dynamic spectrum allocation, Ad-hoc Networks, Green Networks, MIMO, software defined radios and Cognitive Radios. She is member, IEEE and life member of IETE and ISTE societies. She has also been awarded with a Research Grant from AICTE, Govt. of India under research promotion scheme in April 2008 and an international travel grant for attending IEEE conference by Department of Science and Technology, Govt. of India. She received the 'Best Emerging Researcher Award' of the year 2012 at ITM University, India.

**Moin Uddin** is serving as Pro-vice chancellor, Delhi Technological University, Delhi. Prior to this he was Director, National Institute of Technology, Jalandhar, India, and Professor and Head in Electrical Department at Jamia Millia Islamia University, New Delhi. He has over thirty years of teaching and research experience. Twenty one research scholars have completed their Ph.D. under his guidance and six more are pursuing the same. He has designed the computer engineering curriculums of many international and national universities and is among the expert panel of various universities and institutions. He has been a Senior member, IEEE and Life Member, ISTE National Society and Member, board of studies of many institutions. He has completed several projects under Ministry of Human Resource Development and All India Council for Technical Education, Govt. of India.

**Arun Khosla** received his Ph.D. degree from Indraprastha University, Delhi in the field of Information Technology. He is an Associate Professor in Electronics and Communication Engineering department and is presently Heading this department at National Institute of Technology, Jalandhar, India. His research areas of interest are fuzzy modeling, biologically inspired computing and high performance computing. Dr. Khosla is a Member, IEEE and has been reviewer for various IEEE and other National and International conferences and journals. He also serves on the editorial board of International Journal of Swarm Intelligence Research. .

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/2040-8269.htm



# The competitiveness of SMEs in a globalized economy

## Observations from China and India

Rajesh K. Singh and Suresh K. Garg
*Mechanical Engineering Department,*
*Delhi Technological University, Delhi, India, and*
S.G. Deshmukh
*Mechanical Engineering Department,*
*Indian Institute of Technology, Delhi, India*

**Abstract**

**Purpose** – The purpose of this paper is to analyze different challenges for small and medium enterprises (SMEs) in India and China following globalization. It aims to describe the status of these enterprises and examine the roles of government policies and strategy development for competitiveness.

**Design/methodology/approach** – A questionnaire-based survey was conducted, which produced 241 valid responses. Of these, 80 percent were from SMEs. Statistical analysis of the data acquired from survey used a reliability test, *t*-test and correlation analysis. A relevant literature review pinpoints salient issues in the environment of the SMEs.

**Findings** – The governments of China and India have launched various promotional schemes for SMEs. Various challenges for SMEs in these countries are similar; however, the rate of growth is different. Indian SMEs give more attention to supplier development, total productive maintenance and the organization's culture. Chinese SMEs pay more attention to relationship management and cost reduction. Human resource development and quality improvement are also highly correlated with competitiveness.

**Research limitations/implications** – SMEs should focus on developing their human resources and improving product quality. This effort will help SMEs retain human capital as well as increase the demand for their products. Similar studies could explore Chinese SMEs in-depth for additional comparisons.

**Originality/value** – This paper will help SMEs in shaping their competitive strategies and policy formulation by respective governments.

**Keywords** Competitive strategy, Globalization, Government policy, Small to medium-sized enterprises, China, India

**Paper type** Research paper

## Introduction

Small and medium enterprises (SMEs) are considered as the backbone of economic growth in all countries because they account for 80 percent of global economic growth (Jutla *et al.*, 2002). SMEs also contribute a substantial share of the manufactured exports of East Asia (56 percent in Taiwan, over 40 percent in China and the Republic of Korea). In India, the figure is 31 percent. In the newly developing or newly industrialized countries (NICs), SMEs generally employ the largest percentage of the workforce and are responsible for income generation opportunities. These enterprises can also be described as one of the main drivers for poverty alleviation. In manufacturing sector, SMEs act as specialist suppliers of components, parts and sub-assemblies to larger companies because these items can be produced at a cheaper price compared to the price large companies must pay for in-house production of the same components. However, the input of poor quality products can adversely affect the competitiveness of these larger organizations.

Management Research Review
Vol. 33 No. 1, 2010
pp. 54-65

2040-8269
DOI 10.1108/01409171011011562

Upper estimates suggest that SMEs account for 98.9 percent of the total number of businesses in China and comprise 65.6, 63.3, 54 and 77.3 percent of gross industrial output value, sales revenues, total profits and employed people, respectively. Therefore, as Li (2004) suggests, Chinese SMEs emerge as one of the principal industrial forces for economic and social development. Liu (2004) also contends that economic development and SME development are completely interrelated. However, skills in entrepreneurship and small business development in China vary between regions and considerable gaps exist in developments across the country.

The initiation of economic reforms through industrial and trade liberalization in 1991-1992 marked the beginning of a new era for industry in India. The measures included industrial de-licensing, the removal of threshold limits on the assets of large enterprises, the implementation of a liberal policy to facilitate foreign direct investment (FDI), the expansion of the open general license (OGL) list, reductions in customs duties and similar actions. These measures made market entry easier and provided more operational freedom for enterprises. In addition, industry now had cheaper and easier access to imported inputs and capital goods (Bala Subrahmanya, 1999). Developments such as these cancel out the protection enjoyed previously by small-scale Indian industries. Because of the measures, SMEs are now exposed to the pressures from the competitive international business environment.

Because of the globalization of markets, technological advances and the changing needs and demands of consumers forced the nature of competitive paradigms to change continuously. These changes drive firms to compete along different dimensions such as designing and developing new products, adopting smart approaches to manufacturing, implementing quick-to-market distribution, purchasing cutting-edge communication and developing appropriate marketing strategies. Superior manufacturing performance leads to competitiveness with most studies reporting that an organization's competitiveness can be measured within particular financial parameters.

Vargas and Rangel (2007) observed that business performance is positively related with the development of internal capabilities such as "soft technology" (methods and processes that support the firm) and "hard technology" (externally acquired equipment, in-house development of machinery and innovation in raw materials). A strategy of continuous improvement, innovation and change is also part of the process. Singh *et al.* (2006) developed a Competitiveness Index Framework for quantifying the level of competitiveness. In this context, the present study analyzes different challenges for SMEs, pinpoints their status, describes the promotional policies related to SMEs and reviews strategy development in both India and China.

## Challenges ahead for SMEs

Global competition confronts the majority of purely domestic SMEs, whose products and sales are extremely localized and/or segmented. Trade liberalization increases the capacity of well-established foreign manufacturers and retailers to penetrate both remote and underdeveloped markets. Against this development, local SMEs find it increasingly difficult to survive or even maintain their current business position in their respective markets.

In such a demanding environment, the capacity of a firm to maintain reliable and continually improving business and manufacturing processes is critical to ensure long-term sustainability, according to Denis and Bourgault (2003). In a similar vein, Vos (2005) observed the management skills of SME managers and suggested that these managers were weak in their ability to reflect strategically on their current business

position. Moreover, SMEs are frequently oriented towards serving local niches or developing relatively narrow specializations. These enterprises often operate under the constraints of scarce resources, a flat organizational structure, a lack of technical expertise, a paucity of innovation, reduced intellectual capital and the like. The flat structure of SMEs leaves employees frustrated because they are often unable to realize either their short- or mid-term career goals. In this setting, SMEs find it difficult to employ and retain high-caliber staff.

Major constraints in the competitiveness of SMEs are access to adequate technologies (Gunasekaran *et al.*, 2001), excessive costs of product development projects (Chorda *et al.*, 2002), a lack of effective selling techniques and limited market research (Hashim and Wafa, 2002). In addition, other constraints include an inability to meet the demand for multiple technological competencies (Narula, 2004), information gaps between marketing and production functions, and lack of funds for implementing software such as ERP systems (Xiong *et al.*, 2006).

Hussain *et al.* (2006) observed that with the exception of a few top-performing businesses, the majority of SMEs in China do not possess sufficient self-accumulated capital to meet their capital requirements. As such, it appears that a finance gap exists for Chinese SMEs, which limits or constrains their potential for growth. Ernst and Young (2006) identified additional challenges that include weak intellectual property protection, which makes capitalizing on innovation difficult (Berrell and Wrathall, 2007). A shortage of management talent, underdeveloped technology transfer systems and lack of stability in the regulatory environment are also hurdles for SMEs.

Constraints on Chinese SMEs like the low level of technology, a lack of skilled workers, the low level of management expertise, the lack of access to international markets, unsupportive legislations, ineffective incentive policies and lack of financing are constant headaches for SME managers. In India, managers of SMEs face major pressures to reduce costs, improve product quality, deliver goods and services on time. Moreover, Indian SMEs operate generally in an unsupportive environment (Singh *et al.*, 2005).

## The status of SMEs in India and China

In India, 95 percent of industrial units (3.4 million) are in small-scale sector with a 40 percent value addition in the manufacturing sector. Enterprises of this type provide the second highest employment level after agriculture and account for the 40 percent of industrial production. These units contribute 35 percent to India's exports. In this setting, Indian SMEs are fundamentally important to the Indian economic system. Their potential to generate employment, bolster exports and bring flexibility into India's business environment deserves close attention and support form India's policy makers.

In 2003-2004 alone, overall production in the SME sector increased by 8.6 percent. Exports also received a boost from the growing and vibrant SMEs sector. In the 2002-2003, SMEs exports grew by 20.73 percent. However, some concerns emerged with a number of SMEs being reported as "sick units." According to the RBI criteria, perhaps 17.8 percent of the overall SMEs have problems. Deficiencies and problems such as those cited above all contribute to the labeling of these enterprises as "sick."

Although official figures are sketchy, in the early 2000s, China's 2.4 million SMEs accounted for 99 percent of all registered corporations. Certainly, since the mid-1990s, SMEs accounted for about three-quarters of the incremental industrial output value of China's economy. Today, SMEs continue to dominate most industrial sectors with over

70 percent of the gross output value of the food, papermaking and printing industries; over 80 percent of value in the garment tannery, recreation, sports outfit, plastic and metalwork industries; and over 90 percent in the wood and furniture industries. In terms of employment expansion, SMEs currently account for about three in four of all new jobs created nationwide.

Employees in SMEs account for a large proportion of the total employees nationwide – above 85 percent in the industrial sectors, 90 percent in the retailing industry and over 65 percent in the construction industry. In foreign trade and exports, the total export value of China in 2003 amounted to over USD $430 billion and China was ranked fourth in the world in the total import and export values in 2003. In science and technological innovation, SMEs in China have achieved great progress in technological innovation to become the driving force behind the spread and application of new technology and innovation.

By the end of 2003, China established over 100 high-tech enterprise incubators, over 30 university science parks, over 20 enterprise parks for returned overseas students, over 40 service centers for SMEs' technology innovation and more than 500 productivity promotion centers. All these institutions provide strong support for technological innovation within China's SMEs (Chen, 2006). A remarkable feature of these SMEs is that their founders did not possess much personal management or financial expertise and they operated with a dearth of talent within the ranks of senior management. The management systems of SMEs also face excessive government interference related to choice of markets. SMEs are often limited to a small region and tend to produce final products.

**Promotional policies for SMEs by Indian Government**

India has evolved as an extensive institutional network over time for the promotion of small scale industries (SSI). This network extends from the national to state and district levels. Different institutions are Small Industries Development Organization, Small Industries Service Institutes (SISIs), National Small Industries Corporation, National Institute of Small Industries Extension Training, Small Industries Development Corporation and State Financial Corporation and District Industries Centers. These institutions assist small firms across several functions including marketing, exporting, importing, adopting technology and the like.

To meet the challenges of international competition and to promote exports of SSI products, the following promotional schemes are being implemented:

- Small Industries Development Bank of India implements schemes for technology development and modernization of SSI units.
- SISIs organize workshops on ISO-9000 certification and awareness about quality.
- Establishment of tool rooms helps in providing tooling, dies, moulds and fixtures to small-scale units at a very low price to enable SMEs to produce quality goods to meet the requirements of the markets.
- Process-cum-Product Development Centers take up jobs from SSIs for specific product development as well process development to improve the quality of products, reduce cost of product and enhance marketability of goods.
- The government helps SMEs in marketing their products by organizing international exhibitions, sponsoring delegation from different SSI sectors to

various countries and providing pertinent information related to sales opportunities available in international markets.

- Export promotion from small-scale sector has received utmost priority of the government – every policy formulated for achieving growth in exports have a number of incentives available to small-scale exporters.
- With a view to encouraging the small-scale units to produce "quality goods", National Awards for Quality Products are given to outstanding small-scale units.
- A new scheme for technology upgrading for industrial clusters has recently commenced. The scheme includes a diagnostic study of the clusters, the identification of technological needs, types of technological interventions and the wider dissemination of information and technology within the clusters. Recently, the Indian Government raises the capital subsidy given to SMEs by 15 percent for technological upgrades.

**Promotional policies for SMEs by the Chinese Government**

In China, the focus of policies in the mid-2000s was to improve the operating environment of SMEs. The Chinese SMEs Promotion Law, which came into effect in 2003, was a milestone in policies and laws specific to SMEs. It clarified the status of SMEs in the national economy and the responsibility of the corresponding government departments. According to this law, the government would support SMEs actively, improve the quality of service for SMEs, create an environment where enterprises could compete fairly and promise to encourage the development of SMEs with more effective policies, especially in the fields of finance and taxation (Shi and Li, 2006).

Throughout this latest in the string of reforms, by 2006, China placed an emphasis on proactively supporting the development of SMEs (Hussain *et al.*, 2006). The main mission for the government in this period was to implement the SMEs Promotion Law (Chen, 2006), which seeks to improve policies and measures for development, remove institutional barriers, create a level playing field, promote scientific and technological innovations, and upgrade and optimize the industrial structure to enhance the overall quality and competitiveness of SMEs. Due to these reforms and policies, Chinese SMEs have grown quickly in size, number, financial status and profitability.

During this Promotion Law period, two factors played decisive roles. The first factor was the speedy development of the township enterprises. Most of township enterprises were small to medium in size and therefore, became a key force in driving the development of Chinese SMEs. The second factor was the rapid growth of non-public sector of the economy, notably the swift emergence of privately owned SMEs.

Some of the recent initiatives taken by the Chinese government to promote SMEs include:

- Income tax policies for small enterprises – the government lowered the tax rate from 33 percent to 18 percent for those enterprises with an annual profit of less than RMB 30,000 (approximately USD 3,600), and to 27 percent for those with an annual profit of between RMB 30,000 and RMB 100,000 (approximately USD 12,000).
- Taxation policies to promote employment – if a new urban job agency in its first year of operation find jobs for urban residents, of which more than 60 percent are unemployed workers, the agency is eligible for an exemption from business income tax for three years.

- Taxation policies for high-tech enterprises – high-tech enterprises are exempted from enterprise income tax for two years, counting from the year they begin operations.
- Taxation policies for service industries – for new enterprises engaged in transportation, post and telecommunications, consultation, information and technological services are all exempted from income tax for one year from the date of establishment.
- Fiscal policies – since 1999, the Ministry of Finance's innovation fund for technology-based SMEs supports and encourages technological innovations. As well, financial and credit policies proactively support the business initiatives of SMEs.

## Strategy development for competitiveness

Strategy development plays a significant role in improving the competitiveness of enterprises. In sections below, some of the strategies adopted by Indian and Chinese SMEs are described and discussed.

## Strategy development for competitiveness by Indian SMEs

*Background to the Indian study*

Strategy development by Indian SMEs is based on the authors' questionnaire-based survey. Around 1,200 organizations from different sectors, categories and regions were contacted and asked to participate in the survey. These organizations were selected from various directories available at Confederation of Indian Industries, Auto Component Manufacturers Association of India, Federation of Indian Chambers of Commerce and Industry, and Department of Industries, Government of India. A covering letter, which described the objectives of the research and the guidelines for completing the questionnaire, was enclosed.

A total of 241 responses received. Of these respondents, 80 percent were SMEs and remaining 20 percent were from large-scale enterprises. The majority of respondents were from the auto component, plastic and electronics sectors. In the study reported here, organizations with investments in plant and machinery of less than 1,000 million Rs are categorized as SMEs. Reliability analysis, *t*-test and correlation analyses were used to analyze data.

The major attributes of different areas of strategy development, such as cost reduction, quality improvement, competencies development, organization culture, information technology (IT) applications, supplier development, customer satisfaction, total productive maintenance (TPM) and the development of human resources (HR) were identified. In addition, the relationship of these attributes with competitiveness is discussed. The mean scores of different strategies are graphically presented in Figure 1. Cronbach's alpha for all the issues was found more than 0.5. This demonstrates the consistency of the data.

It was observed that the major areas of strategy development for SMEs are supplier development (mean = 3.84), TPM (mean = 3.77) and organization culture (3.70). Reasons for giving maximum focus on supplier development are that most SMEs operate as vendors to original equipment manufacturer (OEM) or large organizations. Therefore, to ensure the quality of the final product, quality at all steps and levels must be monitored.

To improve productivity at all levels, SMEs focus on TPM and organization culture. The results suggest that Indian SMEs give the lowest attention to IT applications.

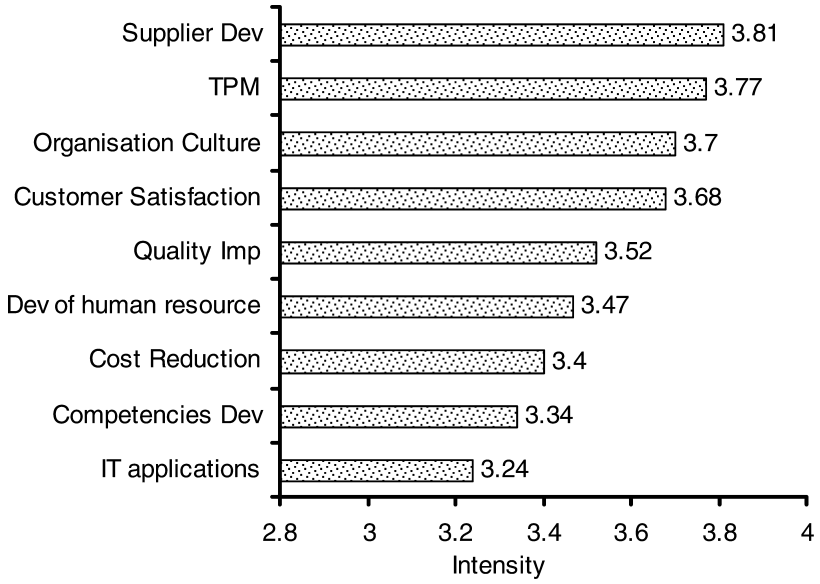



**Figure 1.**
Overall mean score of strategies of SMEs

Lal (2004) found that users of advanced e-business technology perform better than non-users in the export market. Therefore, to improve competitiveness, Indian SMEs should give serious attention to the effectiveness of IT applications at the different operational levels. The amount of money invested internally and through FDI changes depending on the operational area within a firm as well as across and within industrial sectors. Singh *et al.* (2007), for example, observed that the automation of processes, market research and welfare of employees are top ranking priorities for investments by organizations in the Indian auto components sector.

The relationship of these strategies to competitiveness is established via correlation analysis set out in Table I. All strategies are significantly correlated with competitiveness. Strategies for HR development, quality improvement and IT applications have the highest correlation with performance. This implies that Indian SMEs should continuously focus on developing their HR assets, improving product quality and effective applications of IT tools in different operational areas.

| SN | Strategies | Mean | Rank | SD | Correlation with competitiveness |
|---|---|---|---|---|---|
| 1. | Cost reduction | 3.40 | 7 | 0.69 | 0.179* |
| 2. | Quality improvement | 3.52 | 5 | 0.67 | 0.345** |
| 3. | Competencies development | 3.34 | 8 | 0.72 | 0.161* |
| 4. | Organisation culture | 3.70 | 3 | 0.60 | 0.280** |
| 5. | IT applications | 3.24 | 9 | 0.87 | 0.339** |
| 6. | Supplier development | 3.81 | 1 | 0.53 | 0.301** |
| 7. | Customer satisfaction | 3.68 | 4 | 0.45 | 0.194* |
| 8. | Total productive maintenance | 3.77 | 2 | 0.54 | 0.210** |
| 9. | Development of human resources | 3.47 | 6 | 0.62 | 0.481** |

**Notes:** *Correlation is significant at the 0.05 level (two-tailed), **correlation is significant at the 0.01 level (two-tailed), SD – standard deviation

**Table I.**
Overall mean of strategies and their correlation with competitiveness

*Strategy development for competitiveness by Chinese SMEs*

The rapidly developing Chinese SMEs, especially the privately owned enterprises, are currently the dynamic facet of the Chinese economy. By the mid-2000s, private SMEs had become the backbone of the local economy in some areas and/or regions. This development parallels the new and more relaxed approach exhibited by the government toward SMEs in the process of reforming the Chinese-style market economy. Since 1992, the Chinese government focused mainly on enhancing the overall quality and competitiveness of the domestic SME sector (Chen, 2006). Ongoing reforms and opening-up policies have created and maintained a fertile environment for the development and growth of SMEs, especially in the private sector (Li, 2004). Today's SMEs are benefactors of these developments.

One of the important reasons for the growth of Chinese SMEs is the implementation of a strategy, which encouraged SMEs to develop according to their unique nature and circumstances. Previously, SMEs adhered to outdated practices. They were also subjected to external coercion by government to concentrate their efforts on heavy industries. SMEs, however, are blessed with unique attributes like relatively little start-up capital, a fast yield on investments, flexible operating structures and systems, and the capacity to adapt and respond quickly to market changes. For Chen (2006), SMEs quickly judged the market and adjusted their development strategies and industrial structures accordingly – SMEs have an effective mechanism for self-governance.

*Guanxi* is the essence of the Chinese approach to business. One definition of *guanxi* is the "existence of direct particularistic ties between one or more individuals" (Tsui and Farh, 1997). It is evident that *guanxi* relationships play major roles in the success of the development of local SMEs in China (Clegg *et al.*, 2007). Chinese dependency on this particular form of social capital means that internal management processes tend to be more flexible and dynamic compared to similar processes in the West, where the emphasis is on formal, explicit and information-loaded procedures (Gibb, 2006).

The minimization of transaction costs via informal relationship development is characteristic of the Chinese system. Connections between firms are highly personalized and fluid (Castells, 2000). In the start-up period, most of the town and village enterprises (TVEs) under observation aimed to create capabilities to minimize costs (Li *et al.*, 2006). Their strategic intent was to develop cost minimization capabilities rather than make short-term profits. Such capabilities to minimize costs help enterprises of this type survive the competition from well-established international joint ventures and the Chinese state-owned enterprises (SOEs). In addition to the inherent low-cost advantages of SMEs generally, TVEs endeavored to reduce further their operation costs by sourcing cheap materials, simplifying production processes and duplicating Western product designs. Most TVEs under observation produced no-brand products with cheap materials and competed on price rather than quality.

## Discussion and concluding remarks

This paper reinforces the fact that SMEs are significant contributor to economic growth in both China and India. There are many similarities in the approaches adopted by the governments of both India and China to promote SMEs. So too, the major challenges of building product quality, reducing costs and upgrading technology generally are common to SMEs in both countries. Chinese and Indian SMEs also lack capacity in product design and development capabilities. For this reason, these

enterprises are highly dependent on collaborators or foreign technology. In India, major strategies for SMEs are supplier development, TPM and building an appropriate organizational culture.

Human resource development and quality improvement are highly correlated with competitiveness. Chinese SMEs give more focus to cost reduction and relationship management. Relationship management is the emerging management concept for effective resource utilization by Chinese SMEs. However, despite the promotional policies in place in China and India, the rates of economic growth differ significantly. In the late 1970s to mid-1980s, both countries had similar levels of per capita income at roughly USD 200. Following the decades of economic liberalization in China, per capita income grew to USD 8,859. India's grew to USD 3,300.

China accelerated its GDP growth rate to over 8 percent while India could manage only 5.5 percent percent since the early 1980s (*Business Economics*, 2007). This result might be due to Indian factors like the ease in starting a business, receiving credit, the lack of infrastructural growth, law and order issues and restrictive labor laws. Less focus by Indian enterprises on innovation and research and development (R&D) may explain the comparatively low growth in GDP. India spends 0.85 percent of GDP on R&D while China spends 1.5 percent of GDP on R&D. A World Bank report ranks India at 134 out of 175 on the ease of doing business in that country. The report ranks China at 93, Brazil as 121 and Russia at 96. Singapore, New Zealand and the USA are at the top of the list. However, in spite of these obstacles, India has emerged as one of the most attractive destinations worldwide for FDI. India has delivered the second highest annual rate-of-return of 38.36 percent among the BRIC (Brazil, Russia, India and China). The best rate-of-return was delivered by Brazil at 46.19 percent. The Chinese market offered up an annual rate-of-return of 31.36 percent.

Major problem with Indian SMEs is that they operate at very low scale of production and this hinders their capacity to reduce the costs of products and engage in technological upgrades. Capturing a certain scale of operations is very critical in a SMEs growth path. China, for example, produces on a mass scale while Indian SSIs do not expand because they want to retain the facilities and incentives for a longer period even if this retention inhibits growth. Until quite recently, FDI in the Indian SSIs sectors was permissible only if the company in question agreed to export 50 percent of its production. This also affected the upgrade plans of Indian SSIs. However, the Indian government recently removed this restriction to help improve the competitiveness of SMEs in the global market.

Morrison (2006) notes that abundant cheap labor provides China with a comparative advantage concentrated in low-cost, labor-intensive industries that manufacture products. According to Morrison (2006), major problems in China include the inefficient SOEs, corruption in the banking system, growing public unrest over pollution, the levels of government corruption, increasing income inequality between urban and rural areas, and an ill-defined legal system. Due to resource restrictions, China's resource processing industries that include wood pulp and wood product industries will face significant challenges from countries with abundant resources. Chinese SMEs will also face fierce competition from other NICs with low labor costs. For example, even in the late 1990s, the average hourly wage of a textile worker in China was USD 0.58, which was higher than the India rate of USD 0.56, the Indonesia rate of USD 0.43 and the rate of USD 0.44 in Pakistan.

SMEs in both the countries today face tough and challenging times in improving performance. Factors of cost, quality, product range and delivery of services are important areas for development and improvement. To sustain a fair level of competitiveness in both the domestic and global markets, SMEs must strive to utilize information and communication technologies to reach the right markets in cost-effective ways. SMEs of both countries should concentrate on developing HR initiatives and implementing quality improvement techniques.

Part of this process involves improving management talent and techniques as well as improving the level of equipment, technology and innovation capabilities within Chinese and Indian SMEs. In addition, cultivating existing relationships and building new ones with other SMEs as well as stakeholders up and down the supply chain will help improve the competitiveness of SMEs and enhance their sustainability (Gloet, 2006). Improving upon management styles, developing new sales strategies and using cutting-edge marketing methods will also improve the competitiveness of SMEs in both countries. However, the governments of India and China should continue to provide and develop further efficient administrative and legal institutions, quality infrastructure and reduce bureaucratic hurdles at every opportunity.

## References

Bala Subrahmanya, M.H. (1999), "Shifts in India's small industry policy", *National Bank News Review*, July-September, pp. 27-37.

Berrell, M. and Wrathall, J. (2007), "Between Chinese culture and the rule of law: what foreign managers in China should know about intellectual property rights", *Management Research News*, Vol. 30 No. 1, pp. 57-76.

*Business Economics* (2007), 1-15 May, pp. 27-33.

Castells, M. (2000), *The Rise of the Network Society*, Blackwell, Oxford.

Chen, J. (2006), "Development of Chinese small and medium-sized enterprises", *Journal of Small Business and Enterprise Development*, Vol. 13 No. 2, pp. 140-7.

Chorda, I.M., Gunasekaran, A. and Aramburo, B.L. (2002), "Product development process in Spanish SMEs: an empirical research", *Technovation*, Vol. 22 No. 5, pp. 301-12.

Clegg, S., Wang, K. and Berrell, M. (Eds.) (2007), *Business Networks and Strategic Alliances in China*, Edward Elgar, Cheltenham.

Denis, L. and Bourgault, M. (2003), "Linking manufacturing improvement programs to the competitive priorities of Canadian SMEs", *Technovation*, Vol. 23 No. 8, pp. 705-15.

Ernst & Young (2006), "Strategic growth markets", various reports available at: www.ey.com/US/en/Services/Strategic-Growth-Markets

Gibb, A. (2006), "Making markets in business development services for SMEs", *Journal of Small Business and Enterprise Development*, Vol. 13 No. 2, pp. 263-83.

Gloet, M. (2006), "Knowledge management and the links to HRM: developing leadership and management capabilities to support sustainability", *Management Research News*, Vol. 29 No. 7, pp. 402-13.

Gunasekaran, A., Marri, H.B., Mcgauahey, R. and Grieve, R.J. (2001), "Implications of organization and human behavior on the implementation of CIM in SMEs: an empirical analysis", *International Journal of CIM*, Vol. 14 No. 2, pp. 175-85.

Hashim, M.K. and Wafa, S.A. (2002), *Small and Medium Sized Enterprises in Malaysia – Development Issues*, Prentice-Hall, Englewood Cliffs, NJ.

Hussain, J., Millman, C. and Matlay, H. (2006), "SME financing in the UK and in China: a comparative perspective", *Journal of Small Business and Enterprise Development*, Vol. 13 No. 4, pp. 584-99.

Jutla, D., Bodorik, P. and Dhaliqal, J. (2002), "Supporting the e-business readiness of small and medium-sized enterprises: approaches and metrics", *Internet Research: Electronic Networking Applications and Policy*, Vol. 12 No. 2, pp. 139-64.

Lal, K. (2004), "E-business and export behaviour: evidence from Indian firms", *World Development*, Vol. 32 No. 3, pp. 505-17.

Li, J. (2004), *Financing China's Rural Enterprises*, Routledge (Taylor and Francis Group), London.

Li, L., Qian, G. and Ng, P. (2006), "Capability sequencing: strategies by township and village enterprises in China", *Journal of Small Business and Enterprise Development*, Vol. 13 No. 2, pp. 185-97.

Liu, D. (2004), "The relation estimate of regional economic development and growth of small businesses", *Quantitative and Technical Economics*, No. 5, p. 24.

Morrison, W.M. (2006), "China's economic conditions", *CRS Brief for Congress*, Congressional Research Service, Library of Congress, 12 January.

Narula, R. (2004), "R&D collaboration by SMEs: new opportunities and limitations in the face of globalization", *Technovation*, Vol. 24, pp. 153-61.

Shi, J. and Li, P. (2006), "An initial review of Policies for SMEs in the US, Japan and China", *IEEE International Conference on Management of Innovation and Technology*, pp. 270-4.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2005), "Development of flexible strategies by Indian SMEs in electronics sector in emerging economy", *Global Journal of Flexible Systems Management*, Vol. 6 No. 2, pp. 15-26.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2006), "Competitiveness analysis of a medium scale organisation in India: A Case", *International Journal of Global Business and Competitiveness*, Vol. 2 No. 1, pp. 27-40.

Singh, R.K., Garg, S.K. and Deshmukh, S.G. (2007), "Strategy development for competitiveness: a study on indian auto component sector", *International Journal of Productivity and Performance Management*, Vol. 56 No. 4, pp. 285-304.

Tsui, A.N. and Farh, L.J. (1997), "Where guanxi matters: relational demography and guanxi and the social context", *Work and Occupation*, Vol. 24 No. 1, pp. 36-79.

Vargas, D.M. and Rangel, R.G.T. (2007), "Development of internal resources and capabilities as sources of differentiation of SME under increased global competition: a field study in Mexico", *Technological Forecasting and Social Change*, Vol. 74 No. 1, p. 909.

Vos, J.P. (2005), "Developing strategic self descriptions of SMEs", *Technovation*, Vol. 25 No. 9, pp. 989-99.

Xiong, M.H., Tor, S.B., Bhatnagar, R., Khoo, L.P. and Venkat, S. (2006), "A DSS approach to managing customer enquiry stage", *International Journal of Production Economics*, Vol. 103 No. 1, pp. 332-46.

**About the authors**

Rajesh K. Singh is a Senior Lecturer in Mechanical Engineering Department at Delhi College of Engineering, Delhi, India. His areas of interest include competitiveness, small business management, technology management and quality management. He has published papers in journals such as *International Journal of Productivity and Performance Management*, *Singapore Management Review* and *International Journal of Services and Operations Management*. Rajesh K. Singh is the corresponding author and can be contacted: rksdce@yahoo.com

Suresh K. Garg is a Professor in Mechanical Engineering Department at Delhi College of Engineering, Delhi, India. His areas of interest include competitive strategies, JIT manufacturing

systems, quality management, technology management and supply chain management. He has published papers in journals such as *International Journal of Manufacturing Technology and Management, International Journal of Productivity & Quality Management* and the *International Journal of Services and Operations Management.*

S.G. Deshmukh is a Professor in Mechanical Engineering Department at Indian Institute of Technology in Delhi, India. His major research interests include manufacturing management, supply chain management, technology management and quality management. He has published in a range of journals, including *International Journal of Operations & Production Management*, *International Journal of Production Research* and *Competitiveness Review.* He also serves on the editorial board of several international journals.