

A 'modified' Langmuir–Blodgett technique for transfer of graphene oxide monolayer sheets on solid substrates

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2016 Mater. Res. Express 3 035002

(<http://iopscience.iop.org/2053-1591/3/3/035002>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 132.239.1.231

This content was downloaded on 28/03/2016 at 12:54

Please note that [terms and conditions apply](#).

Materials Research Express



PAPER

A 'modified' Langmuir–Blodgett technique for transfer of graphene oxide monolayer sheets on solid substrates

RECEIVED
2 November 2015REVISED
6 December 2015ACCEPTED FOR PUBLICATION
8 February 2016PUBLISHED
24 February 2016V Divakar Botcha^{1,3}, Gulbagh Singh¹, Pavan K Narayanam^{1,4}, S S Talwar¹, R S Srinivasa² and S S Major¹¹ Department of Physics, Indian Institute of Technology Bombay, Powai, Mumbai-400076, India² Department of Metallurgical Engineering and Materials Science, Indian Institute of Technology Bombay, Powai, Mumbai-400076, India³ Department of Physics, University of Petroleum and Energy studies, Bidholi, Dehradun, Uttarkhand-248007, India⁴ Materials chemistry division, Indira Gandhi Centre for Atomic Research, Kalpakkam-603102, India

E-mail: syed@iitb.ac.in

Keywords: graphene oxide, reduced graphene oxide, Langmuir–Blodgett, modified Langmuir–Blodgett**Abstract**

A 'modified' Langmuir–Blodgett (MLB) technique has been developed as a simple, low cost, energy efficient and scalable alternative to conventional Langmuir–Blodgett (LB) technique for the transfer of graphene oxide (GO) monolayer sheets from air–water interface on to solid substrates. This technique uses a simple apparatus consisting of a glass or teflon reservoir and a vertically held stationary substrate. The transfer of GO sheets from the air–water interface onto hydrophilic substrates takes place by a downward movement of the meniscus, through controlled draining of subphase. The MLB technique facilitates the transfer of well defined, uniformly distributed, morphologically flat and adherent GO sheets under a wide range of conditions, namely subphase pH in the range, 3.5–6.5, target pressure in the range, 0.5–8 mN m⁻¹ and meniscus speed in the range, 0.5–3 mm min⁻¹ on RCA-1 treated Si and SiO₂/Si. Compared to LB technique, it offers a substantially larger tolerance for subphase and transfer conditions, even on 'weakly' hydrophilic (RCA-1 treated) Si substrates, as neither crumpling of GO sheets nor transfer of GO agglomerates and multilayers are observed in a wide window of operation. Subsequent chemical reduction of the MLB transferred GO sheets with hydrazine vapor followed by thermal treatment at 400 °C results in reduced graphene oxide (RGO) sheets, which contain ~70% graphitic carbon, with O/C ratio ~0.16 and display Raman G-mode ~1590 cm⁻¹, together with I(D)/I(G) ratio ~1. The electrical conductivity and field effect mobility values of RGO sheets were measured in back gated field effect geometry and are comparable to those for RGO sheets obtained by LB as well as other approaches, followed by chemical/thermal reduction. Thus the newly developed MLB technique opens a simple, versatile and scalable route to obtain device quality GO and RGO sheets on solid substrates.

1. Introduction

Graphene oxide (GO) has attracted immense attention as a precursor to reduced graphene oxide (RGO), which is considered to be a scalable and low cost alternative to graphene for a variety of applications [1–4]. Both GO and RGO are however, becoming increasingly interesting as novel, tailorable and functional 2D nano-materials in their own right, since their physical and chemical properties can be manipulated over a wide range, depending on their O/C ratio and functionalization [1, 5]. GO sheets, obtained by the oxidative exfoliation of graphite, consist of a mixture of sp² and sp³ hybridized carbon atoms, bonded with hydroxyl and epoxide groups on the basal plane, and ionizable carboxylic acid and phenolic groups at the edges [6]. The removal of these oxygen functional groups by chemical/thermal reduction results in a substantial de-oxygenation and restoration of graphitic network, leading to the formation of RGO sheets. It is also highly desirable to obtain films of GO and RGO sheets of controlled morphology on solid substrates for several device applications. This is facilitated by the presence of oxygen functional groups, which make GO sheets dispersible in water and other organic solvents and hence highly processible. GO dispersions have thus been used for the deposition of GO films on different

substrates by several approaches, such as, drop casting [7], spin casting [8], dip coating [9], spray coating [10], knife-blading [11], vacuum spraying [11], vacuum filtration [12] and electrophoretic deposition [13]. Most of these approaches, however, often yield agglomerates of GO and overlapping sheets with folds and wrinkles, formed during solvent evaporation [1, 9, 12–15].

The poly-aromatic islands of un-oxidized benzene rings on the basal plane of GO impart hydrophobic character to GO sheets [6]. Owing to the presence of hydrophilic and ionizable carboxylic groups at the edges, GO sheets behave as amphiphilic ‘molecules’ and can thus be spread as Langmuir monolayer at the air–water interface and subsequently transferred on to solid substrates by Langmuir–Blodgett (LB) technique, as was demonstrated originally by Cote *et al* [16]. The LB technique has since been extensively used to transfer large area (20–100 μm) GO monolayer sheets of controlled morphology and surface density [16, 17]. The effect of critical LB deposition parameters, such as, subphase pH, target pressure and lifting speed on the transfer behavior and morphology of GO sheets has been extensively studied to identify the conditions under which uniformly distributed, morphologically flat and adherent monolayer sheets of GO are transferred on various substrates [16–19]. However, despite these advantages, the LB process has the limitation of scalability, which casts serious doubts on its applicability to industrial processes.

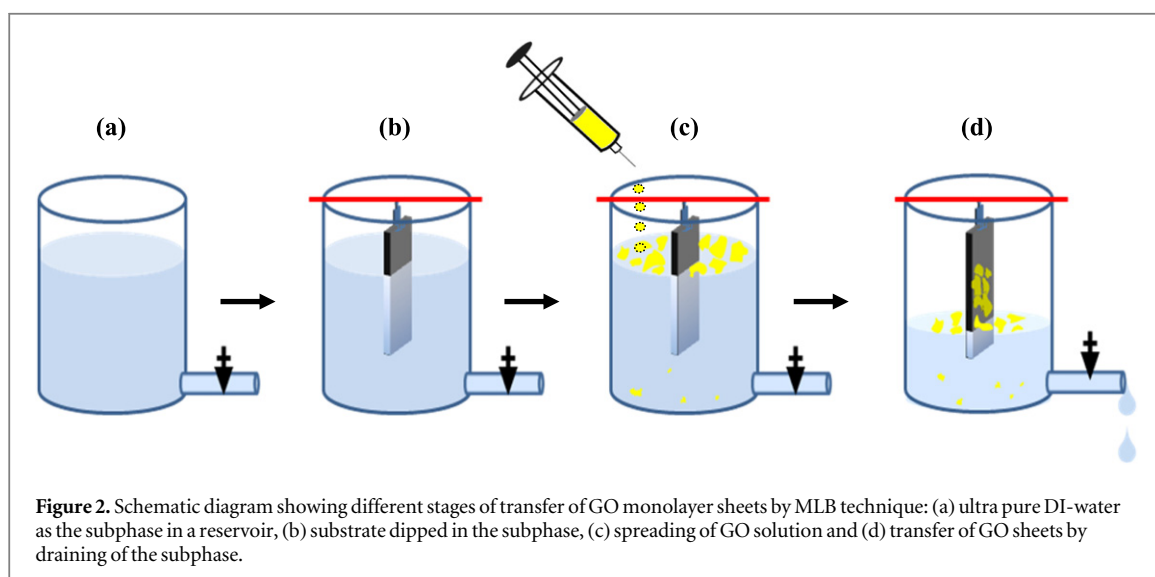
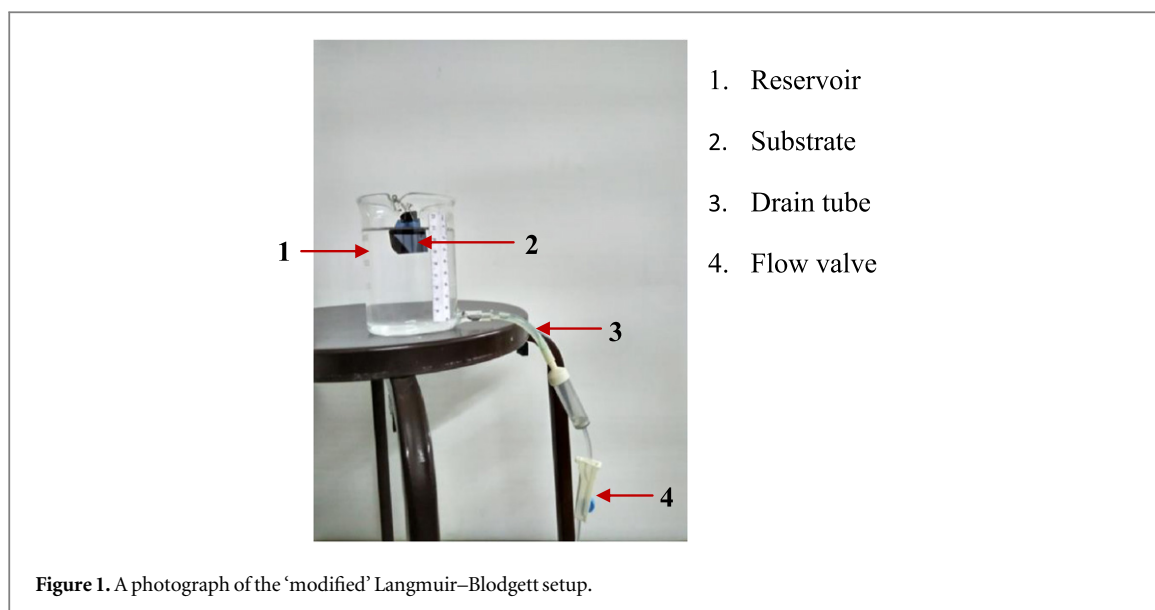
This paper describes a new and improvised deposition technique that has been developed as an alternative to ‘conventional’ LB technique, for the transfer of GO monolayer sheets from the air–water interface on to solid substrates. The new technique is named ‘modified Langmuir–Blodgett (MLB)’ technique, since it makes use of the same basic principles as the LB technique, with the following notable alteration. The vertical transfer of GO sheets on to a hydrophilic substrate in MLB technique is effected by a controlled draining of the subphase, instead of the lifting of the substrate, as in conventional LB technique. Moreover, the MLB technique uses a relatively simple and low cost apparatus, is relatively fast, consumes negligible energy and most importantly, it is amenable to scale up. It is particularly useful for the transfer of a single layer of isolated GO monolayer sheets of controlled morphology and surface density on to hydrophilic surfaces, for which the conventional LB technique appears somewhat cumbersome.

The influence of the MLB process parameters on the morphology and surface density of GO sheets transferred on technologically important hydrophilic substrates, namely Si and SiO_2 was also studied. These parameters include, subphase pH, target pressure and the downward ‘meniscus speed’, which has been used in lieu of the upward ‘lifting speed’ of substrate in conventional LB technique. Interestingly, our results show that compared to the conventional LB technique, the MLB technique allows the transfer of morphologically flat and adherent GO sheets over a much larger window of process conditions and displays a significantly higher tolerance towards the deviations from ideal hydrophilicity, in the case of Si substrates. In order to further establish the usefulness of the new technique, the GO sheets transferred under optimized conditions by MLB technique have been chemically reduced by the commonly used method of exposure to hydrazine vapor, followed by thermal treatment in argon ambient, as reported earlier for LB transferred GO sheets [20]. The composition and chemical structure of these RGO sheets have been studied by XPS and Raman spectroscopy and their electrical conductivity and field effect mobility have been measured by fabricating bottom gated field effect transistors (FETs) on isolated RGO sheets as channel. These studies have demonstrated that the MLB technique offers a simple route to obtain RGO sheets and RGO based device structures, which display all the features, which are normally associated with those obtained by the conventional LB technique followed by chemical/thermal reduction.

2. Experimental details

The set-up developed and used for the MLB deposition in this work is shown in figure 1. It consists of a glass/teflon vessel of height, 10–12 cm and diameter, ~ 10 cm, called the reservoir. The reservoir is used to hold the subphase, the temperature of which can be controlled by external heating or cooling. A draining tube is provided at the bottom of the reservoir, as shown in the figure. A plastic tube is connected to the draining tube with a valve to control the draining rate of the subphase. A glass rod is fixed or held across the rim of the vessel to suspend the substrate vertically into the subphase, with the help of a clip.

The synthesis of GO was carried out by Hummers–Offeman’s method [21] followed by the ultrasonication and centrifugation, as described earlier [18, 20]. The GO sheets thus obtained were dispersed in water–methanol (1:5) mixture and suitably diluted to prepare the standard spreading solution, by calibrating its absorbance at 230 nm. Ultrafiltered and de-ionized water (Millipore, $18.2 \text{ M}\Omega \text{ cm}$) was used as the subphase, and its pH was varied in the range of 3.5–6.5 by adding an adequate quantity of dilute $\text{HCl}/\text{NaHCO}_3$. Both Si and SiO_2/Si substrates were used in this work, and were subjected to RCA-1 ($\text{NH}_4\text{OH}:\text{H}_2\text{O}_2:\text{H}_2\text{O} = 1:1:2$) treatment, prior to use. This is required to improve the hydrophilicity of the substrates to an appropriate level, so as to facilitate the



LB transfer of morphologically flat and adherent GO sheets, which are amphiphilic in character due to the presence of carboxylic groups at the edges and OH groups on the basal plane [18].

The transfer process of GO sheets is shown schematically in figure 2. The reservoir shown in figure 2(a) was filled with an adequate quantity of the aqueous subphase of desired pH, so that a substantial part of the substrate is immersed in it, as shown in figure 2(b). The spreading solution was then spread over the subphase with a micro-syringe, as shown in figure 2(c). The transfer of GO sheets from the air–water interface on to the substrate (figure 2(d)) took place by allowing the liquid meniscus to move down during the draining of the subphase. The rate of downward movement of the meniscus was controlled by the rate of draining (flow) of subphase and will henceforth be called the meniscus speed, as a counterpart to the lifting speed in conventional LB technique.

The transferred GO sheets were chemically reduced by hydrazine vapor treatment followed by heat treatment at 400 °C in argon atmosphere as described earlier [20], to obtain RGO sheets. The surface morphology of as transferred GO as well as RGO sheets was studied by a Digital Instrument Veeco-Nanoscope IV Multimode scanning probe microscope in tapping mode. Scanning electron microscopy (SEM) measurements were carried out in an electron beam lithography system (Raith 150-Two) operated at 10 kV to assess the large area morphology and surface density. XPS data have been recorded by using a Thermo-VG Scientific photoelectron spectrometer (Multilab) equipped with a concentric hemispherical analyzer. The XPS measurements were performed in ultrahigh vacuum at a base pressure $<10^{-7}$ Pa, using a monochromatic 100 W Al K $_{\alpha}$ (1486.6 eV) x-ray source. The Si-2p signal from the substrate was used as internal reference. Advantage V3.9 software was used for XPS data analysis. For XPS peak fitting, the background was considered to

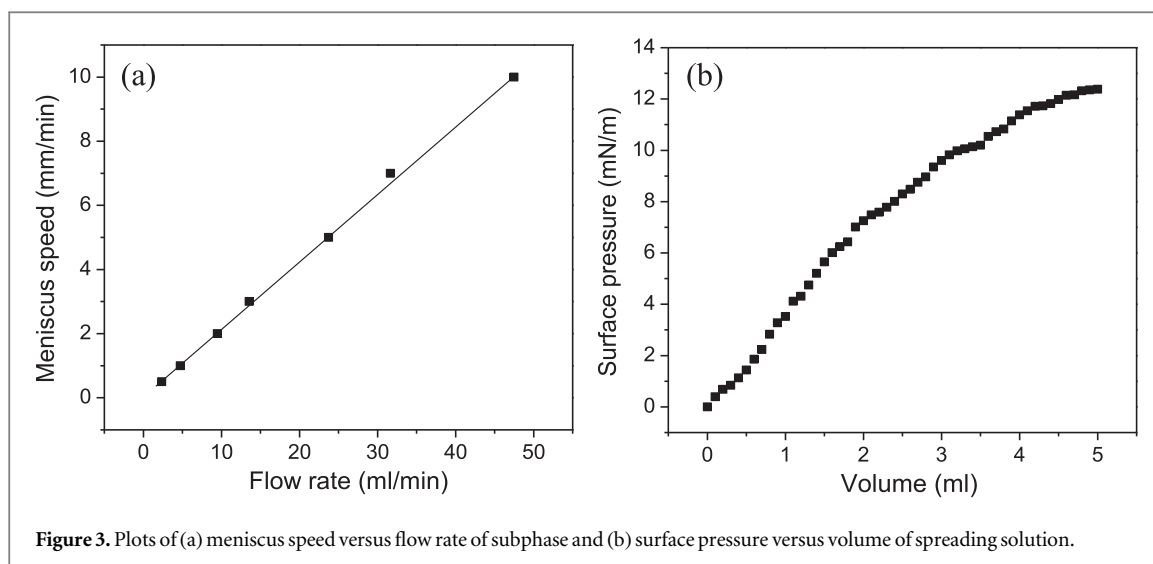


Figure 3. Plots of (a) meniscus speed versus flow rate of subphase and (b) surface pressure versus volume of spreading solution.

be of Shirley type and each component peak was considered to be a mixture of Lorentzian and Gaussian (L/G $\sim 30\%$). Micro-Raman spectroscopy was carried out using a Horiba Jobin Yvon HR800 confocal Raman microprobe equipped with a 514 nm Ar^+ laser. Electrical characterization of GO and RGO monolayers was measured using a two-probe arrangement and bottom-gated FET geometry was employed to measure the field effect mobility. The device structures were fabricated by transferring isolated GO monolayer sheets on to the SiO_2/Si substrate by MLB technique, over which, Cr/Au (5 nm/100 nm) source and drain electrodes were patterned by photolithography and electron beam lithography and deposited by sputtering. A 150 nm thick aluminium back gate contact was deposited by thermal evaporation. The channel length and width were in the range of 10–20 μm and 10–40 μm , respectively. The device characterization was carried out using a Keithley 4200-SCS semiconductor characterization system.

3. Results and discussion

Before carrying out the transfer of GO sheets, a control experiment was performed to calibrate the meniscus speed as a function of the draining (flow) rate of the subphase. The meniscus speed was calibrated against the flow rate of subphase by measuring its position on a vertical scale attached to the wall of the reservoir. The dependence of meniscus speed as a function of flow rate for the set-up used in this work is shown in figure 3(a). A standard spreading solution consisting of the dispersion of GO sheets in water:methanol mixture (1:5), was used. The variation of surface pressure as a function of the volume of spreading solution was studied with a Wilhelmy plate of a KSV 3000 system and is shown in figure 3(b). With increase in the volume of spreading solution up to 2 ml, the surface pressure was found to increase nearly linearly, beyond which, it increased slowly, deviating from linearity, possibly due to the overlapping of sheets or their partial dispersion into the subphase.

The transfer of GO sheets on Si and SiO_2/Si substrates was carried out under different deposition conditions, by varying the subphase pH (3.5–6.5), target (surface) pressure (1–12 mN m^{-1}) and meniscus speed (1–10 mm min^{-1}), in the ranges mentioned in parentheses. A standard GO solution was spread in steps of $\sim 100 \mu\text{l min}^{-1}$ over the subphase to achieve the desired target pressure, and was allowed to equilibrate for ~ 30 min, during which, methanol is expected to evaporate. Typically, 0.5–5 ml spreading solution was required to achieve the target pressure in the above range, for the set up used in this work. The change in surface pressure was also measured during the MLB transfer of GO sheets on the substrate. Typically, for the set up used in this work and a substrate of size about 1 cm \times 1 cm, the surface pressure was found to change well within 5%, during the transfer of GO sheets.

The AFM images of GO sheets transferred on Si and SiO_2/Si substrates at a target pressure of 7 mN m^{-1} , meniscus speed of 1 mm min^{-1} , but at different subphase pH values of 3.5, 5.5 and 6.5 are shown in figure 4. GO sheets transferred on both Si and SiO_2/Si substrates at subphase pH of 3.5 and 5.5 are found to be well defined, uniformly distributed, morphologically flat and adherent, without any wrinkles or curled-up edges. The size of the GO sheets was in the range of 15–30 μm in most of the cases. The height profiles show the sheet thickness to be ~ 1 nm, which is the typical thickness reported for GO sheets [22, 23]. This confirms the monolayer character of GO sheets transferred on to solid substrates by the new technique. The GO sheets transferred on Si at subphase pH of 3.5 exhibit marginally lower surface density compared to those transferred at subphase pH of

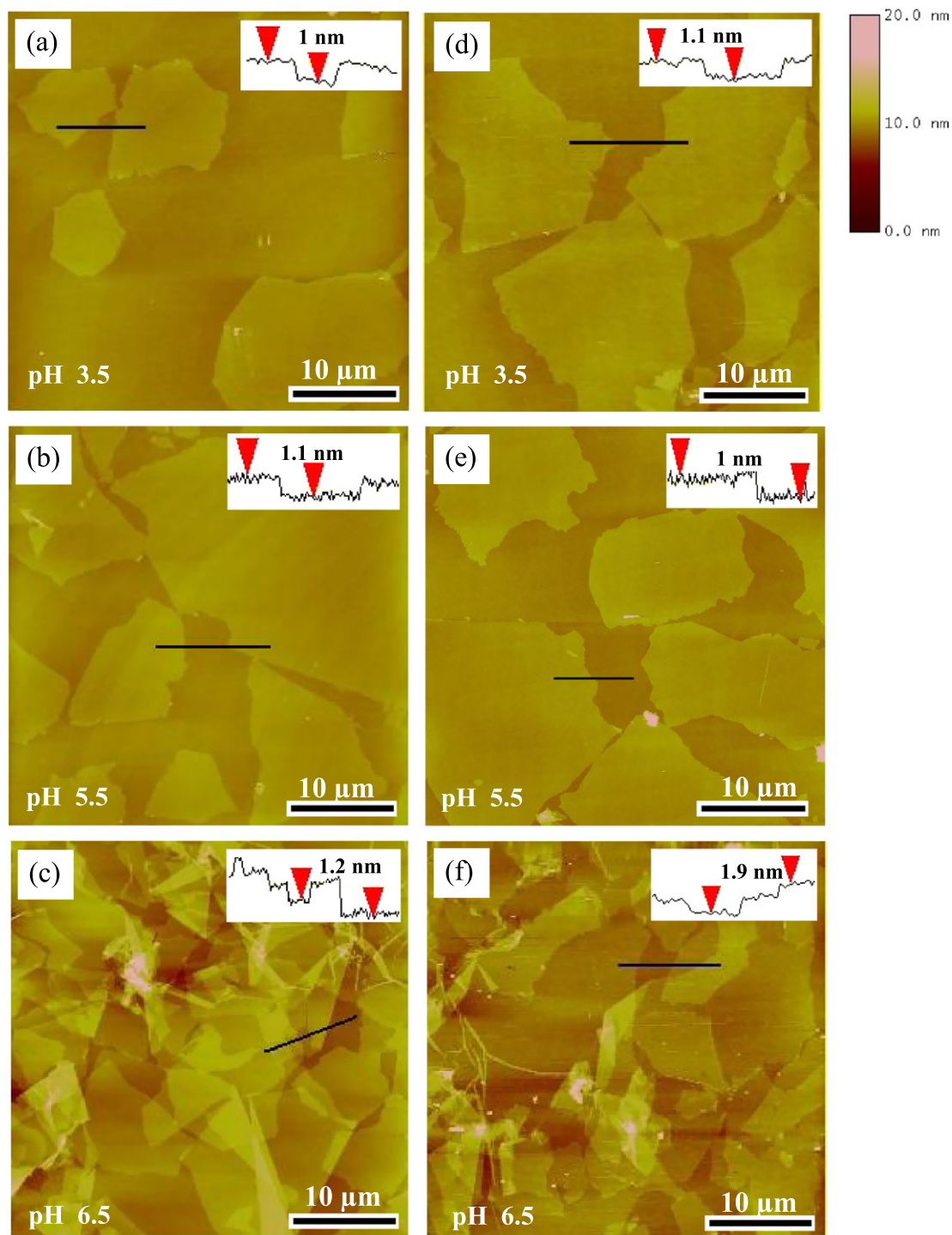
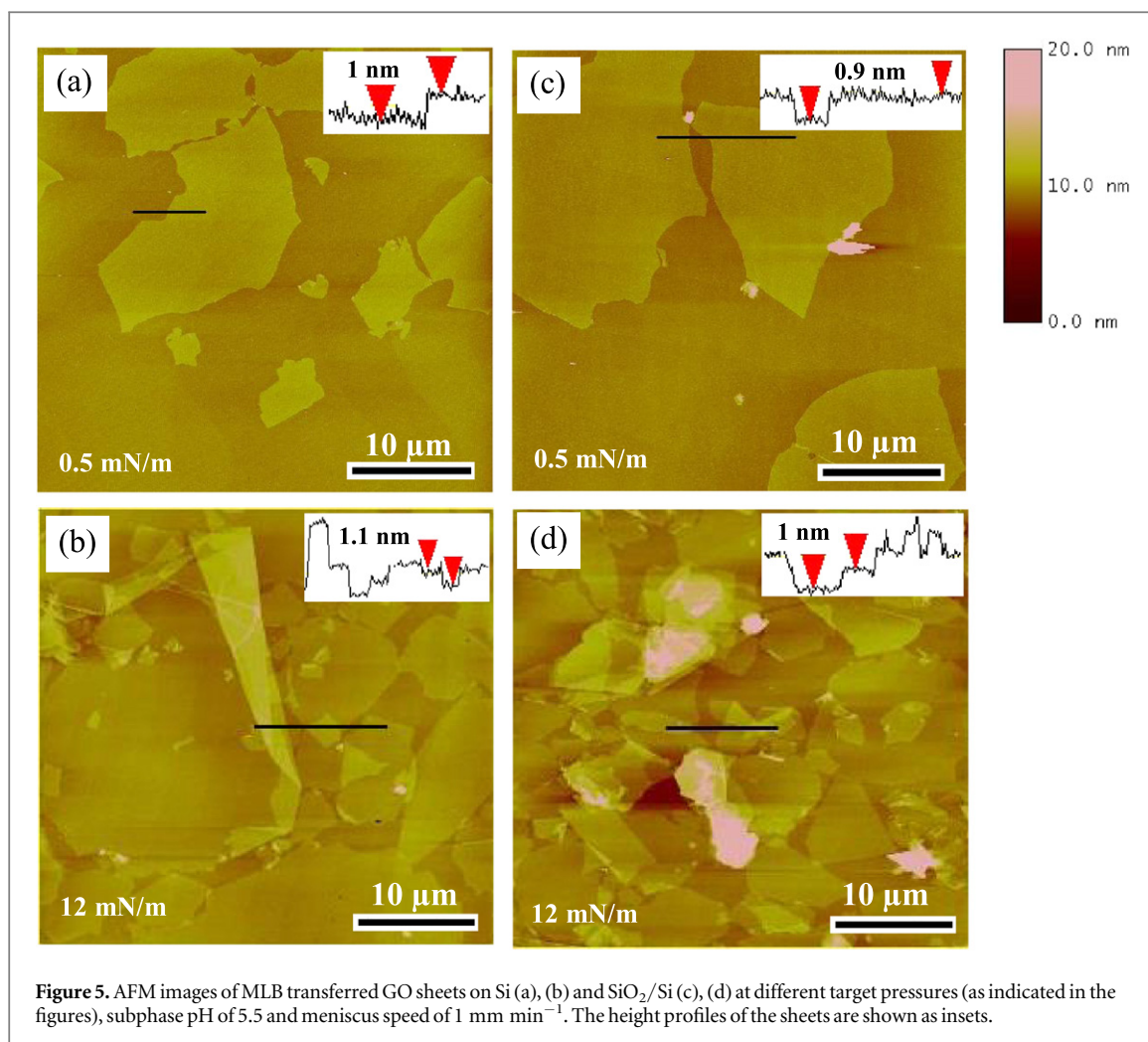


Figure 4. AFM images of MLB transferred GO sheets on Si (a)–(c) and SiO₂/Si (d)–(f) at different subphase pH (as indicated in the figures), target pressure of 7 mN m^{−1} and meniscus speed of 1 mm min^{−1}. The height profiles of the sheets are shown as insets.

5.5. The AFM images of GO sheets transferred on Si and SiO₂/Si substrates at subphase pH of 6.5 are also included in figure 4, which show that the GO sheets transferred on both substrates at higher subphase pH \sim 6.5 display substantial overlap. The increase in surface density of GO sheets with increase in subphase pH on both substrates is similar to that observed in the case of LB transfer of GO sheets and is attributed to increase in their hydrophilicity with subphase pH and the effect of a water lubricating layer, which may facilitate the overlapping of sheets [18, 24].

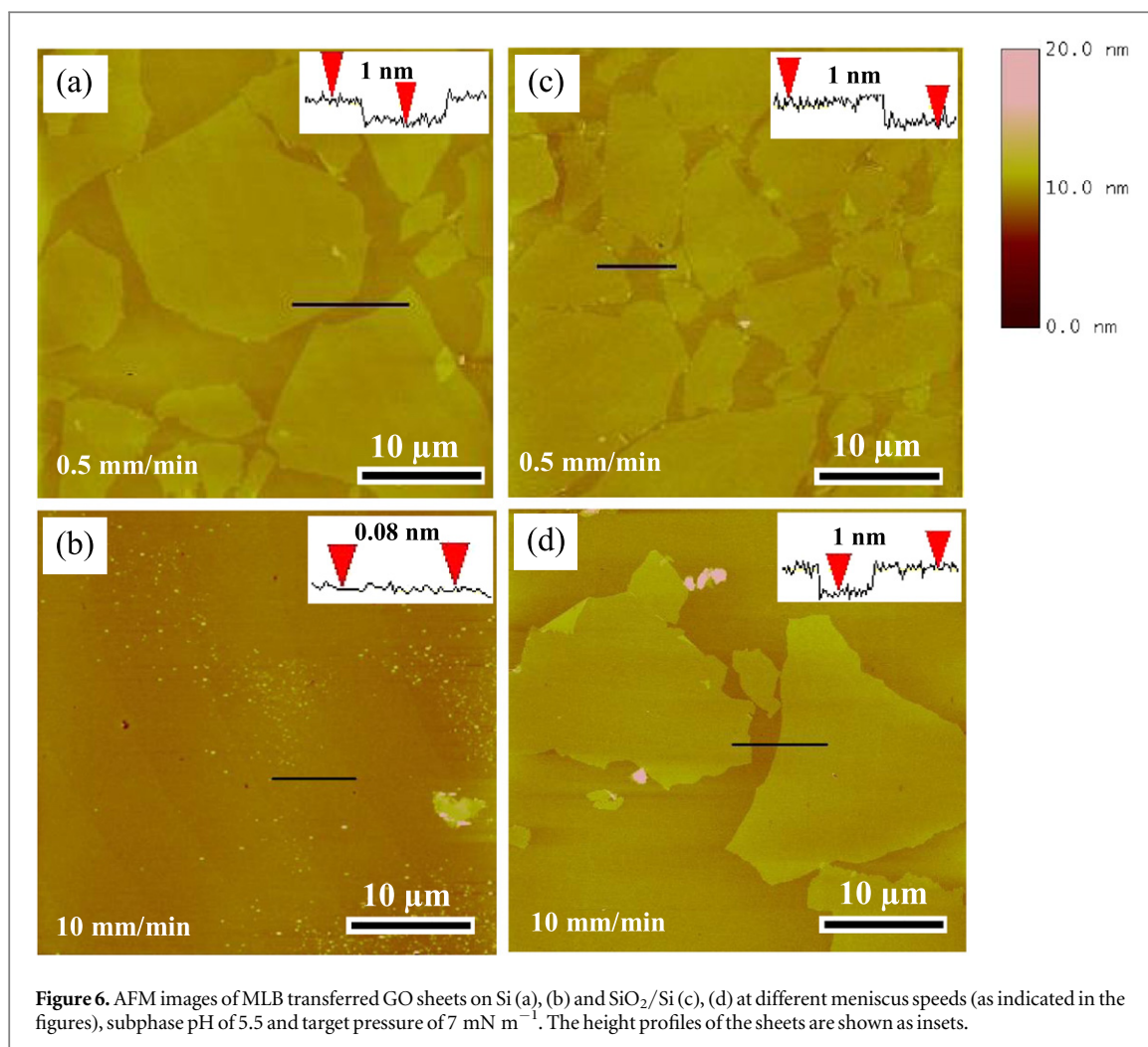
In order to study the effect of target pressure, GO sheets were transferred at a subphase pH of 5.5 and a meniscus speed of 1 mm min^{−1} at different values of target pressure. Typical AFM images of GO sheets transferred on Si at target pressures of 0.5 mN m^{−1} and 12 mN m^{−1} are shown in figures 5(a) and (b), and the images of those transferred on SiO₂/Si are shown in figures 5(c) and (d).



A comparison of figures 5(a)–(d) with figures 4(b) and (e) shows that the density of GO sheets decreases drastically with decrease in target pressure and increases substantially, resulting in overlapping of sheets, at higher values of target pressure. It is, however, interesting to note that well defined and morphologically flat GO sheets are transferred under a very wide range of surface pressure from less than 1 mN m⁻¹–12 mN m⁻¹, unlike with LB technique, which yields crumpled GO sheets at target pressure of 5 mN m⁻¹ or lower on Si substrate, as reported earlier [18]. Such a behavior demonstrates the insensitivity of the MLB process to changes in surface pressure during the transfer of GO sheets, which have been found to be in the range of ~5%.

The effect of meniscus speed on the transfer of GO sheets on Si and SiO₂/Si has been studied at subphase pH of 5.5 and target pressure of 7 mN m⁻¹. Figure 6 shows the AFM images of GO sheets transferred at different meniscus speeds of 0.5 mm min⁻¹, 3 mm min⁻¹ and 10 mm min⁻¹ on both substrates. A comparison of figures 6(a)–(d) with figures 4(b) and (e) shows that the morphology and surface density of the sheets are not significantly affected by the variation of meniscus speed in the range of 0.5–3 mm min⁻¹, as uniformly distributed GO monolayer sheets of comparable surface density are transferred on both substrates. However, at higher meniscus speeds ~10 mm min⁻¹, no GO sheets are transferred on Si substrate, whereas on SiO₂/Si substrate, the surface density of GO sheets is found to decrease considerably.

The above results show that MLB technique facilitates the transfer of well defined, morphologically flat and adherent GO sheets on RCA-1 treated SiO₂/Si surface under a wide range of subphase and transfer conditions, as has also been reported earlier [18] in the case of LB transfer of GO sheets on SiO₂. This is primarily attributed to the strong and uniform hydrophilic character of SiO₂ surface and its strong polar interaction with the hydrophilic groups of the ‘amphiphilic’ GO sheets [18]. It was however shown earlier [18] that the LB transfer of GO sheets on an RCA-1 treated Si surface, which displayed nearly similar contact angle as SiO₂/Si, was highly sensitive to subphase and transfer conditions. A critically optimized window of operation defined by subphase pH of 5.0–6.5, target pressure of 10–15 mN m⁻¹ and lifting speed of 3–5 mm min⁻¹ was thus required to transfer morphologically flat and uniformly distributed GO sheets on Si substrates by LB technique. Outside this window of operation, particularly at low subphase pH, low target pressure as well as low and high lifting speeds,



crumpled GO sheets with large surface roughness, along with GO particles and multilayers (under certain conditions), were found to transfer on even RCA-1 treated Si substrates. For the purpose of comparison, typical AFM images of GO sheets transferred by the conventional LB technique on RCA-1 treated Si under these conditions are reproduced from [18] and presented in figure 7.

In sharp contrast, such a critical dependence of transfer behavior and sheet morphology on subphase and transfer conditions is not seen in the case of GO sheets transferred on RCA-1 treated Si by MLB technique. It is particularly noteworthy that well defined and morphologically flat GO sheets of reasonable surface density are transferred by MLB technique at subphase pH as low as 3.5, target pressure as low as 0.5 mN m⁻¹ and meniscus speed as low as 0.5 mm min⁻¹ on both Si and SiO₂/Si substrates. It is inferred from these observations that the MLB technique offers a substantially larger tolerance for subphase and transfer conditions during the transfer of GO monolayer sheets, even on ‘weakly’ hydrophilic (RCA-1 treated) Si substrates, as well as under conditions, wherein the GO monolayer is expected to have high compressibility (e.g., low surface pressure) and inadequate hydrophilicity (e.g., low subphase pH). The factors that appear to be responsible for the relative insensitivity of MLB technique towards substrate hydrophilicity as well as subphase and transfer conditions are identified as the absence of a lateral compressive force on GO sheets and the vertical motion of the substrate during the MLB transfer process, which are essential components of the ‘conventional’ LB process. Our observations suggest that the membrane-like, 1 nm thick GO sheets are highly vulnerable to morphological deformations during the transfer process, and the mechanism of transfer in this case is significantly different from that of Langmuir monolayers of long chain amphiphilic molecules, which, under high surface pressure and deforming influences during transfer process can respond through orientational adjustments. It is known that during the LB transfer of amphiphilic molecules as ordered molecular films, a close packed structure of the corresponding Langmuir monolayer in solid or liquid condensed region of isotherm and its maintenance during the transfer process through the control of surface pressure are crucial. It appears that these conditions are not strictly required for the transfer of the membrane like GO sheets from the air–water interface, rather a ‘relaxed’ monolayer (without external influences) is better suited to attain morphological stability through the polar interaction with even

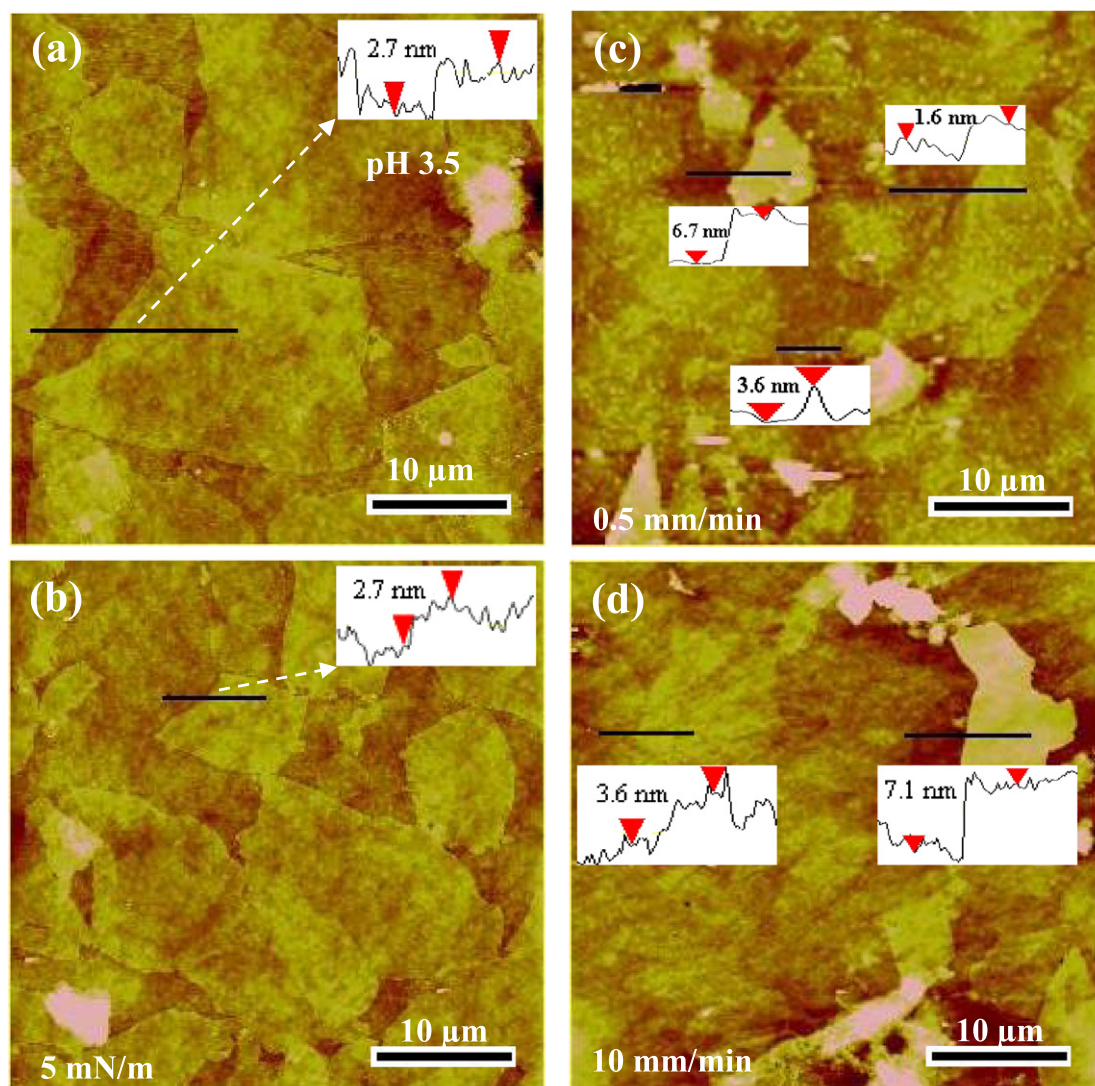


Figure 7. AFM images of LB transferred GO sheets transferred on Si under the following subphase conditions: (a) subphase pH 3.5, target pressure 10 mN m^{-1} and lifting speed 3 mm min^{-1} ; (b) subphase pH 5.5, target pressure 5 mN m^{-1} and lifting speed 3 mm min^{-1} ; (c) and (d) subphase pH 5.5, target pressure 10 mN m^{-1} and lifting speeds 0.5 mm min^{-1} and 10 mm min^{-1} , respectively. The insets show the height profiles of the sheets. (Reproduced from [18] with permission.)

‘weakly’ hydrophilic surfaces. It is thus inferred that a much simpler arrangement than that required for the LB transfer of organized molecular films is sufficient to transfer GO monolayer sheets from the air–water interface.

The above described features of MLB technique make it a simple, scalable and versatile approach to transfer GO monolayer sheets from air–water interface on to a variety of solid substrates under a wide range of subphase and transfer conditions. The technique also makes it conveniently feasible to control the surface density of GO monolayer sheets, as may be required for different kinds of applications, by controlling the target pressure, which simply depends on the volume of spreading solution. This feature is demonstrated over a large scale by typical SEM images of GO sheets transferred at different target pressures at subphase pH ~ 5.5 and meniscus speed of 1 mm min^{-1} , and is shown in figure 8. These SEM images reveal that GO sheets of low surface density are transferred below 1 mN m^{-1} . The surface density of the sheets increases with increase in target pressure and overlapping sheets are transferred above 10 mN m^{-1} .

In order to evaluate the quality of RGO sheets obtained by MLB technique, the reduction behavior of MLB transferred GO sheets and their morphological stability during the widely used chemical reduction with hydrazine hydrate [20, 24] was studied. Figure 9 shows the AFM images of GO sheets transferred by MLB technique under typical optimized conditions on both Si and SiO_2/Si substrates, subsequent to their reduction by exposure to hydrazine vapor followed by heat treatment in argon at 400°C . The RGO sheets on both substrates are found to be morphologically flat and adherent, without any curling up at the edges. The AFM height profiles at the edges of RGO sheets show a thickness of $\sim 1 \text{ nm}$, confirming their morphological stability

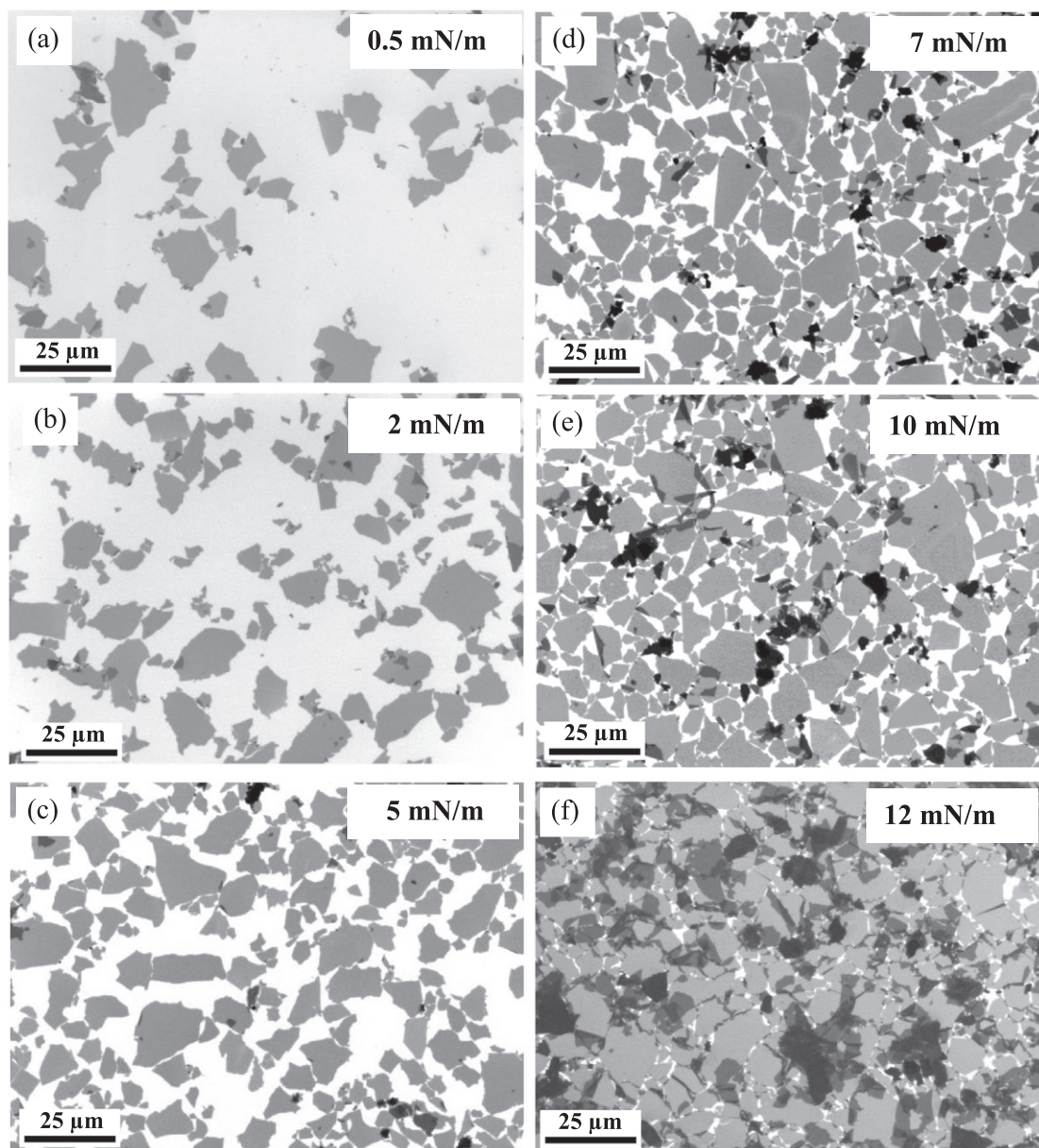


Figure 8. SEM images of MLB transferred GO sheets on SiO_2/Si at different target pressures of (a) 0.5 mN m^{-1} , (b) 2 mN m^{-1} , (c) 5 mN m^{-1} , (d) 7 mN m^{-1} , (e) 10 mN m^{-1} and (f) 12 mN m^{-1} . The subphase pH was ~ 5.5 the meniscus speed was 1 mm min^{-1} .

and monolayer character, after being subjected to the chemical reduction process. These features are similar to those reported earlier [20] for GO sheets transferred by conventional LB technique.

The chemical nature of bonding defects present in as-transferred GO sheets obtained by MLB technique and RGO sheets obtained by subsequent chemical reduction, as described above, has been studied by recording the C-1s core level XPS spectra, shown in figure 10. The GO sheets exhibit a broad peak in the range of 281–291 eV, which has been deconvoluted and the various components of C-1s peak are shown in figure 10(a), following the assignments given in [20]. The major peaks at 284.4 eV and 285.3 eV are respectively assigned to undamaged alternant hydrocarbon structure or graphitic carbon and damaged alternant hydrocarbon structure, while the other component peaks are assigned to contributions from the oxygen functional groups [22, 25]. The weak and broad component at $\sim 290 \text{ eV}$ may be attributed to $\pi-\pi^*$ shake up satellite of the 284.4 eV peak, which has been usually reported to be in the range of 290–292 eV [22, 26, 27]. Figure 10(b) does not show significant changes in peak positions, but the $\text{sp}^2\text{-C}$ content (based on relative integrated intensities) is found to increase from 40% to 70%, while the peaks corresponding to C–O, C=O and COOH groups decrease substantially after the chemical/thermal reduction of GO sheets. The O/C ratio estimated from C-1s peak is also found to be 0.50 for GO sheets, which decreases to a value of 0.16 after reduction. The decrease in the oxygen content is associated with the

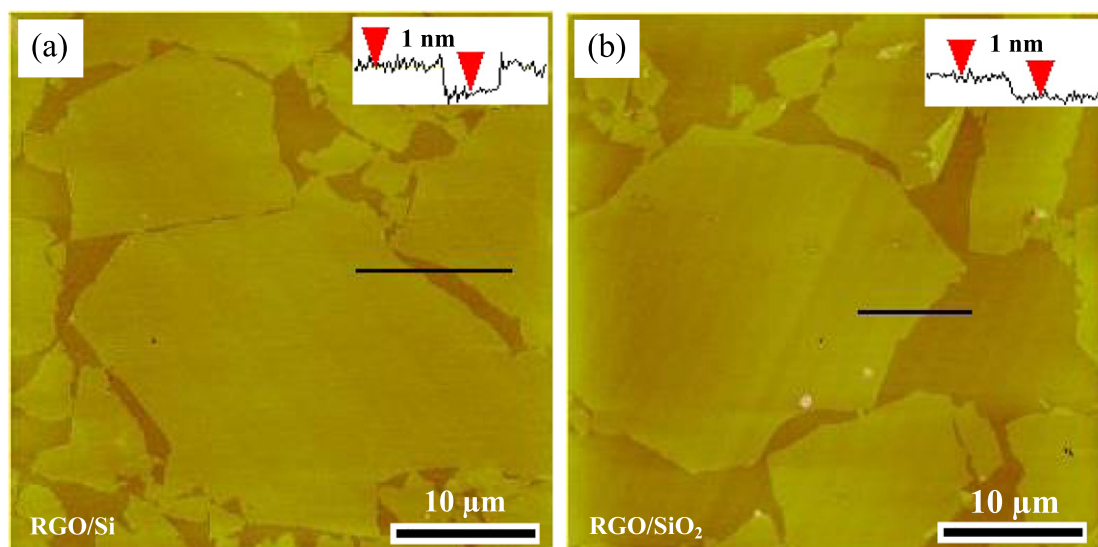


Figure 9. AFM images of RGO sheets on (a) Si and (b) SiO₂/Si substrates. The height profiles of the sheets are shown as insets.

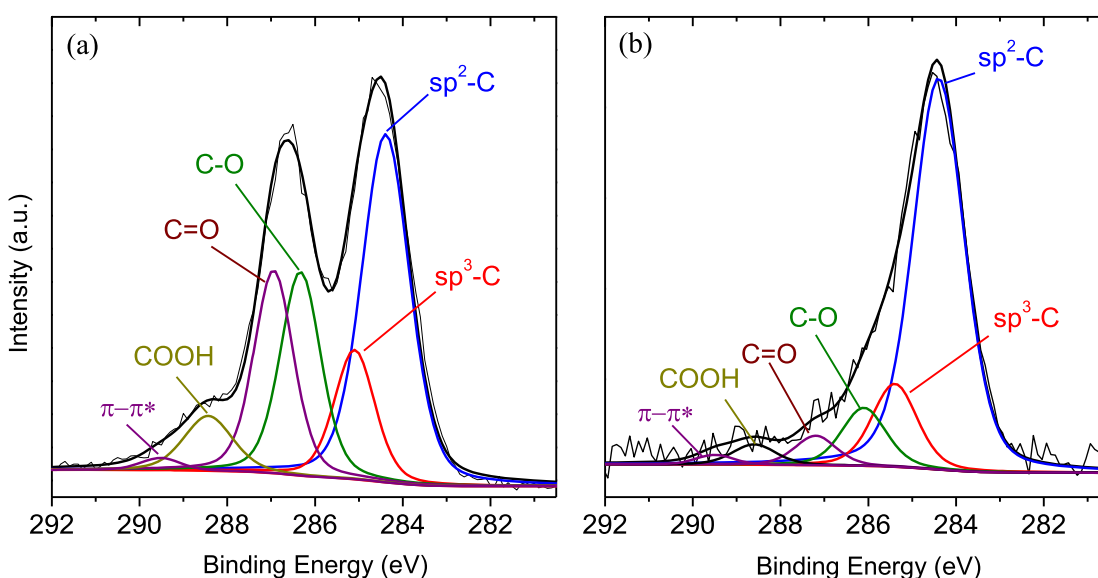


Figure 10. De-convoluted C-1s core level XPS spectra of (a) MLB transferred GO sheets on Si substrate and (b) RGO sheets on Si substrate obtained subsequently by hydrazine vapor treatment followed by heat treatment in vacuum/argon atmosphere at 400 °C.

disappearance/decrease of the contributions of oxygen functional groups in RGO sheets, similar to the behavior reported earlier [20] for LB transferred GO sheets.

Typical Raman spectra of GO and RGO sheets obtained by MLB technique under optimized conditions are shown in figure 11. The spectra show two prominent peaks, which are associated with the well known G- and D-modes of graphite [18, 28]. The G-mode appears at $\sim 1600\text{ cm}^{-1}$ for GO sheets and shifts to $\sim 1590\text{ cm}^{-1}$ for RGO sheets, which is indicative of the reduction of GO films, as its peak values in the range of $1580\text{--}1588\text{ cm}^{-1}$ have been attributed in literature to sp^2 bonded carbon in graphene [1]. The $I(\text{D})/I(\text{G})$ ratio is also found to decrease substantially from (1.6 ± 0.1) for GO sheets to (1.0 ± 0.2) for RGO sheets, which in the case of crystalline graphitic forms, implies increase in sp^2 fraction and size of sp^2 bonded domains [22, 30]. The red-shift of G-mode and the decrease of $I(\text{D})/I(\text{G})$ ratio, together with the increase in $\text{sp}^2\text{-C}$ content (from XPS results), are indicative of the reduction of defects and disorder in GO sheets and show that a substantial recovery of graphitic network takes place during the chemical/thermal reduction process [1, 29]. These features are consistent with the increase in $\text{sp}^2\text{-C}$ content and decrease of the O/C ratio, as inferred from XPS results and are similar to those observed for GO sheets transferred by the conventional LB technique [20].

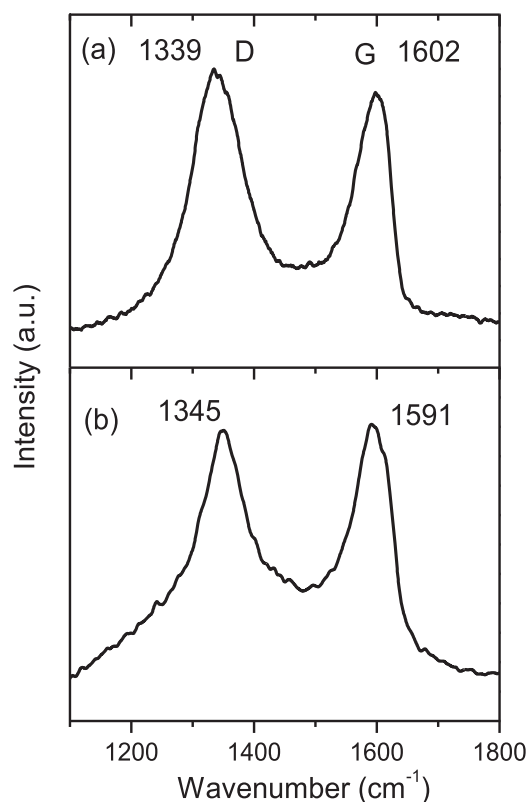


Figure 11. Typical Raman spectra of (a) MLB transferred GO and (b) RGO sheets.

The effect of chemical/thermal reduction on the electrical properties of GO monolayer sheets transferred by MLB technique has been investigated. For this purpose, isolated GO sheets were transferred on SiO_2/Si at a subphase pH of 5.5, target pressure of 0.5 mN m^{-1} and meniscus speed of 1 mm min^{-1} and reduced, in the manner described above. The bottom gated FETs were fabricated with isolated GO and RGO sheets, as described in section 2. Figure 12(a) shows the schematic of the bottom gated FET and the SEM image of a GO monolayer sheet in two-probe contact geometry. Figure 12(b) shows typical $I_{\text{DS}}-V_{\text{DS}}$ curves of the as-transferred GO and chemically reduced GO sheets. The GO monolayer sheet shows nearly insulating behavior with conductivity in the range of $(10^{-6}-10^{-5}) \text{ S cm}^{-1}$. In contrast, the conductivity increases to $(2-10) \text{ S cm}^{-1}$ for the RGO sheet, implying a sheet resistance of $\sim 1000 \text{ k}\Omega \text{ sq}^{-1}$. The conductivity and sheet resistance values obtained in this work are comparable to the values usually reported for RGO monolayer sheets obtained by chemical/thermal reduction [16, 30–32]. Figure 12(c) shows the typical transfer characteristics of an isolated RGO sheet with gate voltage in the range of -40 to $+40 \text{ V}$, recorded at a V_{DS} of 1 V under ambient conditions. The FET exhibits ambipolar behavior with the charge neutrality point on the positive side of the gate voltage. This is indicative of the dominantly p-type nature of RGO sheets which is consistent with the reported [33–35] electrical behavior of RGO under ambient conditions. The field-effect mobility values for holes and electrons has been extracted from the slope $(\Delta I_{\text{DS}}/\Delta V_{\text{G}})$ in the linear regions of the $I_{\text{DS}}-V_{\text{G}}$ curve, using the equation $\mu = (L/WC_{\text{G}}V_{\text{DS}})(\Delta I_{\text{DS}}/\Delta V_{\text{G}})$, where L and W are the channel length and width, respectively, and C_{G} is the gate capacitance. The field effect mobilities have been estimated for 30 devices and their values were found to be in the ranges of $(0.05-2.5) \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and $(0.05-0.5) \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$, for holes and electrons respectively. These field effect mobility values are comparable to those reported for RGO sheets obtained usually by hydrazine treatment followed by heat treatment at temperatures in the range of $400^\circ\text{C}-1000^\circ\text{C}$ [30–32].

4. Conclusions

A MLB technique has been developed as an alternative to ‘conventional’ LB technique for the transfer of GO monolayer sheets from air–water interface on to solid substrates. This technique uses a simple apparatus consisting of a glass or teflon beaker as reservoir, is low cost, fast, energy efficient and scalable and is particularly suitable for obtaining single layers of GO sheets of controlled surface density over large areas. Unlike the LB technique, it uses a vertically held stationary substrate, and the transfer of monolayer from the air–water

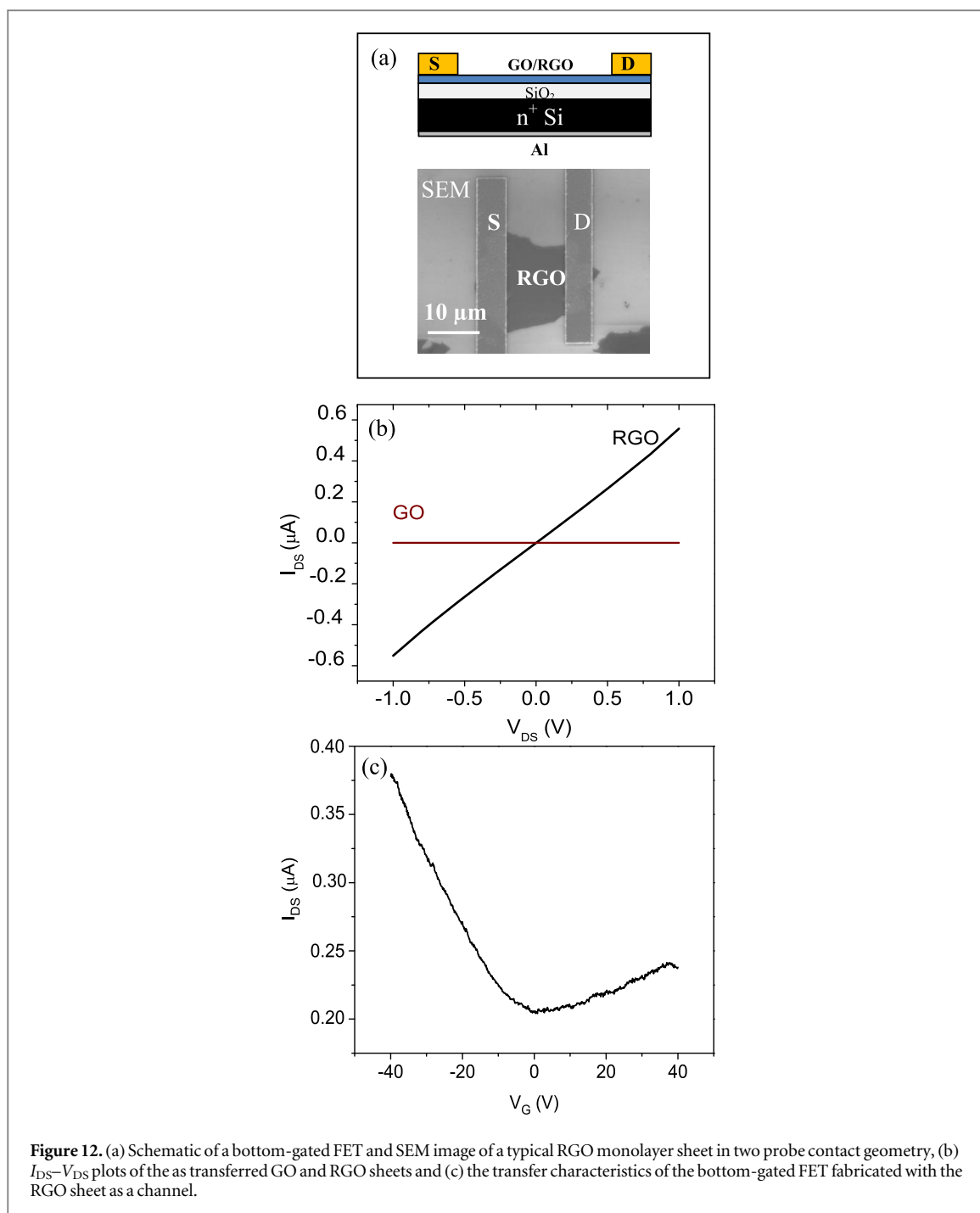


Figure 12. (a) Schematic of a bottom-gated FET and SEM image of a typical RGO monolayer sheet in two probe contact geometry, (b) I_{DS} - V_{DS} plots of the as transferred GO and RGO sheets and (c) the transfer characteristics of the bottom-gated FET fabricated with the RGO sheet as a channel.

interface onto a hydrophilic surface takes place by a downward movement of the subphase meniscus, which is achieved by a controlled draining of subphase from the bottom of the reservoir. The total volume of the spreading solution controls the target pressure and the variation of target pressure due to the transfer of GO sheets does not appear to affect the morphology of the sheets, particularly their flatness. The MLB technique facilitates the transfer of well defined, uniformly distributed, morphologically flat and adherent GO monolayer sheets under a very wide range of subphase and transfer conditions, namely subphase pH (3.5–6.5), target pressure (0.5–8 mN m⁻¹) and meniscus speeds (0.5–3 mm min⁻¹) on RCA-1 treated SiO₂/Si as well as Si substrates. Interestingly, it offers a substantially larger tolerance for subphase and transfer conditions even on ‘weakly’ hydrophilic (RCA-1 treated) Si substrates, as well as under conditions, wherein the GO monolayer has high compressibility (e.g., at low surface pressure) and inadequate hydrophilicity (e.g., at low subphase pH). Most importantly, the crumpling of GO sheets observed after transfer under non-optimal conditions by conventional LB technique on Si, is clearly absent in the case of MLB transferred GO sheets. These features are attributed to the absence of a lateral compressive force during the MLB transfer process, which does not appear to be a requirement for the transfer of membrane-like GO sheets, which are highly vulnerable to morphological

deformations during the transfer process, unlike in the case of closely packed amphiphilic molecules in a Langmuir monolayer. Hence, the much simpler arrangement of MLB technique, which does not necessitate a strict control of surface pressure during transfer and allows the GO sheets to attain morphological stability through polar interaction with a hydrophilic surface, appears to suffice.

The morphological features of MLB transferred GO monolayer sheets and the subsequently obtained RGO sheets by chemical/thermal reduction are comparable to those obtained by conventional LB technique. There is an added advantage of an excellent control over the surface density of the sheets, exercised simply through the control of the volume of spreading solution, which fixes the target pressure. The RGO sheets obtained by the MLB technique followed by chemical reduction with hydrazine and thermal treatment at a moderate temperature of 400 °C display graphitic carbon content of ~70% and the O/C ratio ~0.16 and the red shift of Raman G-mode $\sim 1590\text{ cm}^{-1}$, together with a substantial decrease of $I(D)/I(G)$ ratio to ~ 1 . These features are comparable to those reported for the RGO sheets obtained by LB as well as other routes. The suitability of RGO sheets obtained by the MLB route for device processing has been explored by fabricating back gated FETs on isolated sheets on SiO_2/Si . The electrical conductivity and field effect mobility values of RGO monolayer sheets thus obtained are also comparable to those fabricated by most other approaches, including LB processing. These results show that the newly developed MLB process opens a simple, versatile and scalable route to obtain device quality GO and RGO monolayer sheets on solid substrates.

Acknowledgments

The authors would like to thank FIST (Physics)-IRCC Central SPM facility, IIT Bombay for SPM measurements, Centre for Excellence in Nanoelectronics (CEN), IIT Bombay for electrical measurements, Centre for Research in Nanotechnology and Science (CRNTS), IIT Bombay for Raman spectroscopy measurements and Central Surface Analytical Facility, IIT Bombay for XPS measurements.

References

- [1] Allen M J, Tung V C and Kaner R B 2010 *Chem. Rev.* **110** 132
- [2] Becerril H A, Mao J, Liu Z, Stoltenberg R M, Bao Z and Chen Y 2008 *ACS Nano* **2** 463
- [3] Eda G and Chhowalla M 2010 *Adv. Mater.* **22** 2392
- [4] Loh K P, Bao Q, Eda G and Chhowalla M 2010 *Nat. Chem.* **2** 1015
- [5] Gengler R Y N, Spyrou K and Rudolf P 2010 *J. Phys. D: Appl. Phys.* **43** 374015
- [6] Lerf A, He H, Forster M and Klinowski J 1998 *J. Phys. Chem. B* **102** 4477
- [7] Lv X, Huang Y, Liu Z, Tian J, Wang U, Ma Y, Liang J, Fu S, Wan X and Chen Y 2009 *Small* **5** 1682
- [8] Lee D H, Kim J E, Han T H, Hwang J W, Jeon S, Choi S Y, Hong S H, Lee W J, Ruoff R S and Kim S O 2010 *Adv. Mater.* **22** 1247
- [9] Wang X, Zhi L and Mullen K 2008 *Nano Lett.* **8** 323
- [10] Gilje S, Han S, Wang M, Wang K L and Kaner R B 2007 *Nano Lett.* **7** 3394
- [11] Chen J, Shepherd R L, Razal J M and Minett A I 2013 WO 040636
- [12] Eda G, Fanchini G and Chhowalla M 2008 *Nat. Nanotechnol.* **3** 270
- [13] Lee V, Whittaker L, Jaye C, Baroudi K M, Fischer D A and Banerjee S 2009 *Chem. Mater.* **21** 3905
- [14] Dideykin A, Aleksenskiy A E, Kirilenko D, Brunkov P, Goncharov V, Baidakova M, Sakseev D and Vul Y A 2011 *Diam. Relat. Mater.* **20** 105
- [15] Huang N M, Lim H N, Chia C H, Yarmo M A and Muhamad M R 2011 *Int. J. Nanomed.* **6** 3443
- [16] Cote L J, Kim F and Huang J 2009 *J. Am. Chem. Soc.* **131** 1043
- [17] Zhou X and Liu Z 2010 *Chem. Commun.* **46** 2611
- [18] Botcha V D, Narayanam P K, Singh G, Talwar S S, Srinivasa R S and Major S S 2014 *Colloids Surf. A* **452** 65
- [19] Zheng Q, Ip W H, Lin X Y, Yousefi N, Yeung K K, Li Z and Kim J K 2011 *ACS Nano* **5** 6039
- [20] Sutar D S, Narayanam P K, Singh G, Botcha V D, Talwar S S, Srinivasa R S and Major S S 2012 *Thin Solid Films* **520** 5991
- [21] Hummers W S and Offeman R E 1958 *J. Am. Chem. Soc.* **80** 1339
- [22] Paredes J I, Villar-Rodil S, Martinez-Alonso A and Tascon J M D 2008 *Langmuir* **24** 10560
- [23] Wang H, Wang X, Li X and Dai H 2009 *Nano Res.* **2** 336
- [24] Cote L J, Kim J, Zhang Z, Sun C and Huang J X 2010 *Soft Matter* **6** 6096
- [25] Yang D Q and Sacher E 2006 *Langmuir* **22** 860
- [26] Mattevi C, Eda G, Agnoli S, Miller S, Andre M K, Celik O, Mastrogianni D, Granozzi G, Garfunkel E and Chhowalla M 2009 *Adv. Func. Mater.* **19** 2577
- [27] Fan X, Peng W, Li Y, Li X, Wang S, Zhang G and Zhang F 2008 *Adv. Mater.* **20** 4490
- [28] Stankovich S, Piner R D, Chen X, Wu N, Nguyen S T and Ruoff R S 2006 *J. Mater. Chem.* **16** 155
- [29] Malard L M, Pimenta M A, Dresselhaus G and Dresselhaus M S 2009 *Phys. Rep.* **473** 51
- [30] Ferrari A C and Robertson J 2000 *J. Phys. Rev. B* **61** 14095
- [31] Becerril H A, Mao J, Liu Z, Stoltenberg R M, Bao Z and Chen Y 2008 *ACS Nano* **2** 463
- [32] Su C Y, Xu Y, Zhang W, Zhao J, Tang X, Tsai C H and Li L J 2009 *Chem. Mater.* **21** 5674
- [33] Wu J, Becerril H A, Bao Z, Liu Z, Chen Y and Peumans P 2008 *Appl. Phys. Lett.* **92** 263302
- [34] Lopez V, Sundaram R S, Navarro C G, Olea D, Burghard M, Herrero J G, Zamora F and Kern K 2009 *Adv. Mater.* **21** 468
- [35] Tung V C, Allen M J, Yang Y and Kaner R B 2009 *Nat. Nanotechnol.* **4** 25



A NOVEL APPROACH FOR R-PEAK DETECTION IN THE ELECTROCARDIOGRAM (ECG) SIGNAL

B. Khaleelu Rehman, Adesh Kumar and Paawan Sharma
Department of EI and C, University of Petroleum & Energy Studies, Dehradun, India
E-Mail: krehman@ddn.upes.ac.in

ABSTRACT

The research article proposes the effective method for R-peak detection in the ECG signal. The improper beating of the heart called cardiac arrhythmia which is risk to human. The ECG samples are taken from physionet (physio bank ATM). Analysis of ECG signal and detection of R-Peaks is discussed in this paper. Initially the noise is removed from the signal using FFT technique, windowing technique and thresholding technique to detect R-peaks. In the ECG signal processing one can encounter the difficulties like unequal distance between peaks, irregular peak form, occurrence of low-frequency components due to patient breathing etc., In order to resolve and reduce the effect of these factors processing pipeline should contain particular stages which is discussed in the paper and the R-peak detection algorithm is implemented in MATLAB R 2012b.

Keywords: electrocardiogram, MATLAB R 2012b, cardiac arrhythmia, R-peaks detection.

1. INTRODUCTION

Electrocardiogram (ECG or EKG) signal plays a vital role in the clinical diagnosis of the health conditions of human heart. ECG is an investigative tool which measures and registers the electrical motion of the heart in detail. ECG is produced by a nerve impulse stimulus to a heart. The electric current is diffused around the body surface through which the current at the body surface will build on the potential drop (voltage) which is very small ranging from microvolt to mill volts with an impulse variation, this voltage is small in amplitude of impulse, which needs several hundred times of amplification. It's a periodic signal which lasts for 0.9 seconds and it consists a waveform P,Q,R,S and T as shown in Figure-1 and the normal amplitude values of P,Q,R,S and T are shown in Table-1(a)[7].

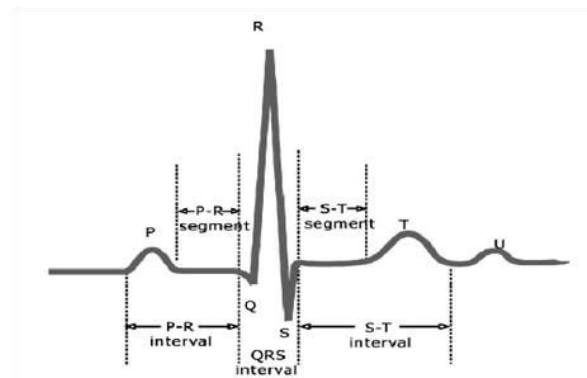


Figure-1. Normal ECG signal.

Table-1(a). Standard ECG data.

Amplitude		Duration	
P wave	0.25 mV	P-R interval	0.12 to 0.20 Sec
R wave	1.60mV	Q-T interval	0.35 to 0.44 Sec
Q wave	25% of R wave	S-T segment	0.05 to 0.15 Sec
T wave	0.1 to 0.5mV	P wave Interval	0.11 sec
		QRS complex	0.09 Sec
		PR segment	0.06 to 0.10 Sec
		ST segment	0.10 to 0.15 sec
		T wave	Varies

The rapid depolarization of the ventricles is characterized by the QRS complex which takes less than 0.09 Seconds [3]. QRS complex is the most important part in the ECG signal; especially the features of R-Wave in the QRS complex play a vital role in the pathological diagnosis. ECG is associated with different noises when it is recorded. It may have variety of noises such as base line

wander the 50 Hz power line interference etc., These type of noises disrupts the original signal and the detection of R-peaks. For this purpose a reliable and efficient R-peak detection is essential.



2. METHODOLOGY

The proposed methodology is depicted in the Figure-2R-peak detection structure

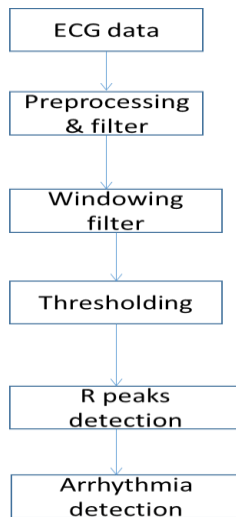


Figure-2. Block diagram of complete process of R-peak detection.

Steps for R-peaks detection in the ECG signal:

Step 1: ECG sampled data is taken from website and FFT technique is applied

A noisy ECG samples are taken from the website physionet.org [4] The graph shown in Figure-3 is an even thus first step is to straighten it. In order to straighten it all the low frequency components should be removed. The X-axis values represent samples and the Y-axis represents the Voltage. Apply the Fast Fourier transform-FFT to the ECG signal Figure-3 using Equation (1)

$$X(K) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi nk/N}$$

Where $K=0,1,2,\dots,N-1$; (1)[2]

Again IFFT is applied to the signal to get back the original time domain signal.]

$$x(n) = \frac{1}{N} \sum_{K=0}^{N-1} X(K) e^{j2\pi nk/N}$$

where $n=0,1,2,\dots,K-1$; (2)[2]

Equation (1) remove low frequencies by using FIR filter and restore ECG with the help of inverse Fast Fourier transform using equation (2) The result of FFT processing is shown in Figure-4.

Step 2: Using a windowing filter.

Second step is to find local maxima that observe only maximum in the window and ignore the others. This method has its background in the processing of signals through which the windowing operation performs the

spectral analysis of non-periodic signals. In this step the window of default size is used as a result in Figure-5.

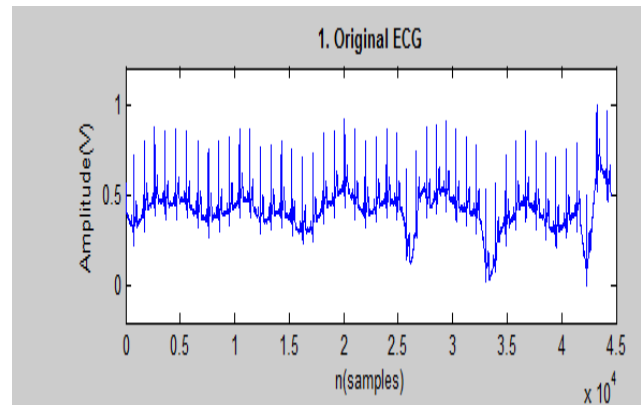


Figure-3. Original ECG signal.

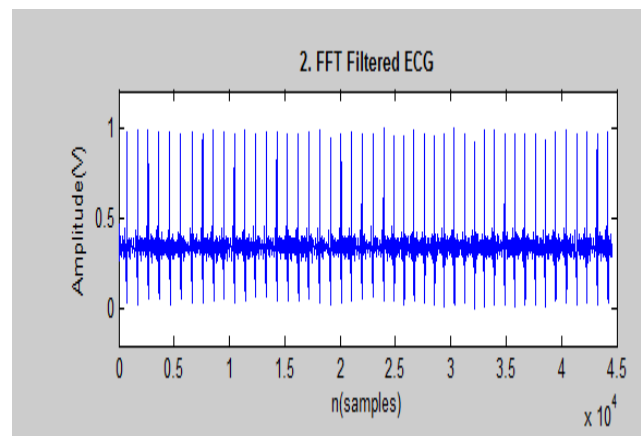


Figure-4. FFT filtered signal.

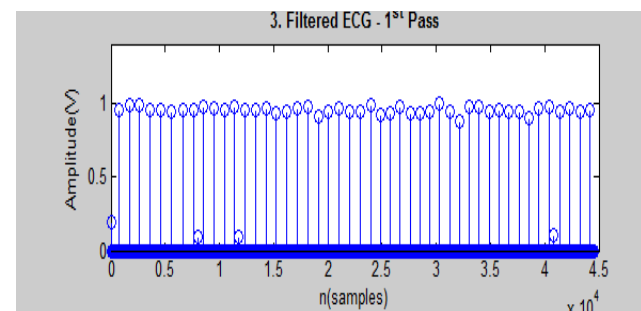


Figure-5. Filtered ECG-1st pass.

Step 3: Thresholding

After the step of windowing filter apply the threshold filter to remove the small peaks and preserve the significant ones. The graph after applying the threshold filter appears as in Figure-6.

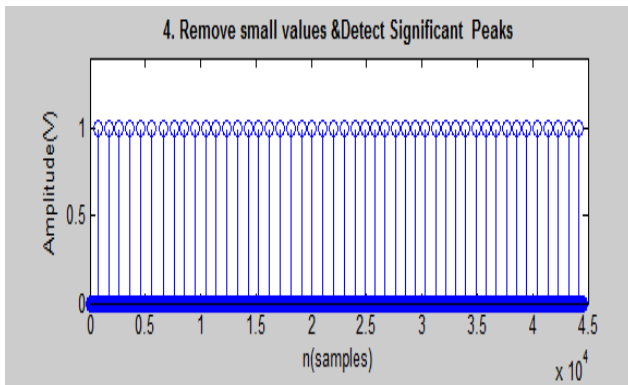


Figure-6. Detected peaks.

Step 4: Adjusted filter

From Figure-6 it is observed that the result is satisfying but it does not assure that it will have all the peaks in general case. To overcome this, the size of the filter window is altered and techniques of filtering are repeated in Figure-7. The quality of graph is more desirable than Figure-6.

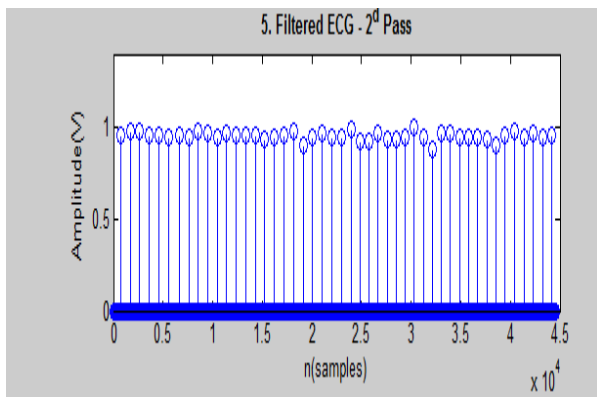


Figure-7. Filtered ECG - Second pass.

Step 5: Final result of peaks is shown in Figure-8 and Figure-9 shows the overlapped R peaks which is used for the heart rate calculation.

Heart rate calculation: The number of times a heart beats per minute is heart rate. For a normal person, heart beats 60 to 100 times per minute so the normal value is 60 to 100 beats per minute.

If the heart rate is slower, then the condition is called Bradycardia. If the heart rate is higher, then it is tachycardia and unevenly spaced cycles specify an arrhythmia [10]. If PR interval is more than 0.2 Sec, blockage of AV node is indicated. The equation to calculate heart rate is given below:

$$\text{Heart Rate} = (1/\text{RR Interval in sec}) * 60 \quad (3)[1]$$

Bradycardia: If heart beats less than 60 BPM [8] then it is slower heart rate. This condition can be observed in athletes and the patients suffering from jaundice, myxedema and in patients with increased intra carinal pressure.

Tachycardia: If heart rate is greater than 100 BPM it is tachycardia. Atrium having ectopic focus that regularly beats at a higher rate [8] causes tachycardia.

The heart rate can be calculated from the Figure-8

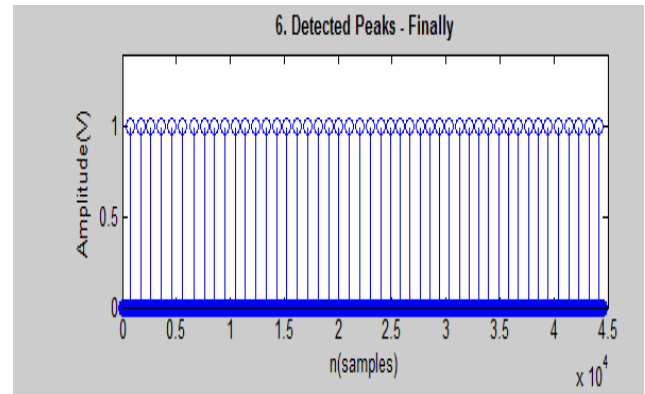


Figure-8. Final detected peaks.

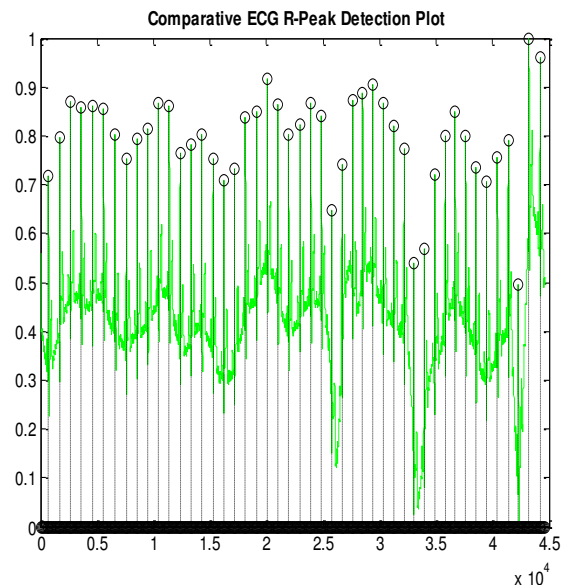


Figure-9. ECG R-peak detection plot.

3. RESULTS AND ANALYSIS

The ECG samples are taken from physionet.org website and the proposed R-peak detection algorithm is executed. The algorithm besides the accuracy of the detection, the processing time of the algorithm is optimized. As discussed in the paper to calculate heart beat and to check the cardiac arrhythmia steps from Figure-3 to Figure-7 are carried out and got the BPM as 68 (i.e. Normal Heart beat) using equation(3) the same steps are repeated for the 10 ECG samples from the database. Each record is slightly over 1 min in length and sampled at 1000 Hz shown in below Table-1(b). From the Table-1 (b) it can be concluded that the patient details of S. No. 1,3,5,7,8 and 10 are normal heart beat whereas S. No. 4 & 9 observed bradycardia and S. No. 2 and 6 observed tachycardia.

**Table-1(b).** Patient details and Heart rate.

S. No.	ECG sample collected from physionet.org database	Heart rate from RR interval	Normal	Bradycardia	Tachycardia
1	Record ptbdb/patient001/s00141re from 0:00.000 to 1:00.000	67	Normal	-----	-----
2	Record ptbdb/patient003/s00171re from 0:00.000 to 1:00.000	112	-----	-----	Tachycardia
3	Record ptbdb/patient007/s00261re (i) , from 0:00.000 to 1:00.000	72	Normal	-----	-----
4	Record ptbdb/patient035/s01451re (ii) from 0:00.000 to 1:00.000	53	-----	Bradycardia	-----
5	Record ptbdb/patient067/s02301re (iii) , from 0:00.000 to 1:00.000	75	Normal	-----	-----
6	Record ptbdb/patient180/s0545_re (avr) , from 0:00.000 to 1:00.000	115	-----	-----	Tachycardia
7	Record ptbdb/patient219/s0441_re (avr) , from 0:00.000 to 1:00.000	78	Normal	-----	-----
8	Record ptbdb/patient286/s0546_re , from 0:00.000 to 1:00.000	84	Normal	-----	-----
9	Record ptbdb/patient289/s0550_re , from 0:00.000 to 1:00.000	56	-----	Bradycardia	-----
10	Record ptbdb/patient294/s0559_re , from 0:00.000 to 1:00.000	68	Normal	-----	-----

4. CONCLUSIONS

R-peak detection algorithm is executed using pre-processing and filtering techniques, Windowing techniques and thresholding. The information obtained from the detection of R-peaks is very helpful for classification analysis and detection of arrhythmia such as tachycardia and Bradycardia. This algorithm gives the best locations of R-peaks from which R-R interval can be easily computed and heart beat rate can be calculated. From the Table-1 (c) the heart rate for the 10 ECG samples are calculated. The main advantage of this algorithm is taking less time for long time ECG signals.

REFERENCES

- [1] AlexakisC.,NyongesalH.O.,SaatchiR.,HarrisN.D.,DaviesC.,Emery C.,IrelandR.H. and HelleS.R.2015. Feature extraction and classification of electrocardiogram (ECG) signals related to hypoglycaemia. Computersin cardiology. 537-540.
- [2] A.V. Oppenheim, RW.Schafer and J.R. Buck. 2014. Discrete-time signal processing, 5th edition. Prentice hall.
- [3] El Mimouni, M. Karim. 2015. Novel Simple Decision Stage of Pan and Tompkins QRS Detector and its FPGA-Based Implementation. ISBN 978-1-4673-2679-7, pp. 331-336.
- [4] https://www.physionet.org/physiobank/database/ptbdb/patient001/S0010_re.
- [5] Izzah T.A., Alhady S.A., Ngah U.K. and Ibrahim W.P. 2013. A journal of real peak recognition of electrocardiogram (ecg) signals using neural network. American Journal of Networks and Communications. pp. 9-16.
- [6] <https://in.mathworks.com/help>
- [7] N. Debbabi, S.El Asmi, H. Arfa. 2014. Real-time Correction of ECG baseline wander Application to the Pan and Tompkins QRS detection algorithm. IEEE.
- [8] Patel P.G., Warriar J.S. and Bagal U.R. 2012. ECG analysis and detection of arrhythmia usingmatlab. International Journal of Innovative Research and Development. 18(3): 59-68.
- [9] Sarkar. 2010. Elements of digital signal processing, Khanna publishers, Delhi, India.
- [10] WillisJ.Topkins. 2012.Biomedical Digital signal processing, prentice hall.

Application of Remote Sensing and GIS technique in rooftop mapping and PV module layout design

Rishabh, Harmeet Singh Kathuria, Ashish Aggarwal and Saurabh Mittal
University of Petroleum and Energy Studies,
smittal@ddn.upes.ac.in

ABSTRACT

The rooftop solar market is at an early stage in India with many potential participants. India will likely see a record growth in the rooftop solar segment in the coming years. The current project aims at mapping the Solar Rooftop and designing the layout for the Photovoltaic (PV) modules on the roofs of a subset of New Delhi region. The project includes estimation of the height of the buildings using a high resolution optical imagery, from shadow of the building, without involving any physical measurement. These heights are then utilized to prepare 3D Models for Solar Rooftop Mapping using Geographical Information System (GIS) and Remote Sensing. Occlusions and shadows, which lead to poor distinction of neighboring buildings, can have a significant impact on 3D modeling. Important data included the angle and alignment of the roof, the sun's path across the sky, shadows cast by a chimney or another rooftop over the course of the day, and the seasonal change in hours of sunlight. 3D building model was designed using various softwares. After solar irradiance calculations and finding out the shadow free area on the roofs, i.e. Photovoltaic (PV) layout of the rooftops can be used for effective and efficient installation of PV modules. This project is an attempt to utilize renewable energy resources by placing photovoltaic modules on rooftops for fulfilling residential electricity purposes.

Keywords: Rooftop mapping, 3-D Modelling, Shadow Analysis, Photovoltaic Layout, ArcGIS, Solar Generation.

1. Introduction

Renewable Energy that replaces conventional fuels and is served as an alternative source of energy has contributed 19 percent to our global energy consumption and 22 percent to our electricity generation in 2014 and 2015, respectively (REN21, 2016). Solar Energy is one of the most important sources of renewable energy available nowadays. Harnessing the solar energy depends upon geographic location, weather condition, altitude and diurnal variation. Annual average solar energy on entire planet may produce 1.3 - 1.4 kW/m² electric energy (Muhammad Luqman, 2015).

India's Solar Power Potential Estimation is around 750 GW, but India's current solar power installed capacity is just around 4 GW (Clean Technica, 2015), or less than 0.55% of the estimated potential. Naturally there exists a massive opportunity to tap this potential. This can be best done by installing solar panels on the roofs of the buildings. The main purpose of our project is to do the rooftop mapping and design the Photovoltaic (PV) Module layout which helps in depicting the pattern and design for the installment of solar modules at the available area on the roof of the buildings.

A rooftop solar mapping and its assessment helps in discovering the solar potential of every single rooftop in a city. Most rooftop analyses use GIS-based methods for estimating the

suitable space for rooftop PV. GIS-based methods use primarily 3-D models to determine solar resource or shadow effects on buildings. The 3-D models are most often generated from orthophotography or light detection and ranging (LiDAR) data, and they are combined with slope, orientation, and building structure data to estimate total solar energy generation potential. Though the availability of Lidar data is less, but because of advancement in the GIS technology the height of the buildings can be computed directly using the mensuration tools in ArcGIS 10.3.1.

2. Study area

Study area in this project is a subset of New Delhi region from longitude 77°12'35.615"E to 77°15'45.956"E and latitude 28°37'44.646"N to 28°34'28.48"N. A high resolution multispectral imagery (Fig. 1) from Worldview 3 sensor with a spatial resolution of 1.28m is used for rooftop mapping.

Climatologically, in New Delhi, summers start in early April and peak in May, with average temperatures near 32 °C (90 °F) and receives an average annual rainfall of 790 mm. Minimum rainfall occurs in months of November & December i.e. 9 mm, while maximum rainfall occurs in months of July with an average of 237 mm. Winter starts in November and peaks in January, with average temperatures around 12–13 °C (54–55 °F) (Wikimedia Foundation, 2015).



Figure 1: Study area

3. Methodology

3.1 Sampling

The data is collected from various sources including

- i) 1.28m multispectral spatial resolution imagery of the study area was taken from www.digitalglobe.com
- ii) Multi-Temporal resolution digital imageries were available from Google Earth.
- iii) Actual building height data was obtained from www.emporis.com

3.2 Statistical procedure

1. A multispectral imagery of Worldview-3 sensor with 1.28m spatial resolution of the study area is acquired for performing the solar rooftop mapping. After verifying appropriate image preprocessing checks, the image is used for determining the height of various buildings using the imagery in ArcGIS 10.3.1 software applying mensuration tools. There are 3 tools available to find out height of any structure.
2. After finding the height of the buildings through software, the observed height of the buildings is compared with the original height to find out the percentage error. Regression analysis is performed for observed and true height of the buildings.
3. Then, 3D model of the buildings is created in SketchUp software. Objective of 3D modelling is to use roof of the building for shadow analysis such that the rooftop area shaded by other buildings, structures and vegetation etc. that affect the solar yield can be differentiated from the rest of the rooftop area.
4. Shadow free area mapping is followed up by designing of PV module layout design on the rooftop in AutoCAD software such that maximum ground ratio can be achieved.
5. Now, since photovoltaic module layout is prepared, PVSyst software is used for techno-commercial optimization of tilt and pitch of the modules
6. Solar potential and solar generation on the rooftop for specified inverter and photovoltaic modules are calculated.

3D model is used to obtain shadow-free rooftop. Shadow analysis is performed in order to find out the rooftop area shaded by other buildings, structures and vegetation that affect the solar yield and finally, once we got the shadow free area, the designing of the rooftop is done in AutoCAD to prepare a layout for Solar PV modules. Designing of a layout is followed up by PVSyst software, in order to find out the solar generation on the rooftops for specified inverter and PV module.

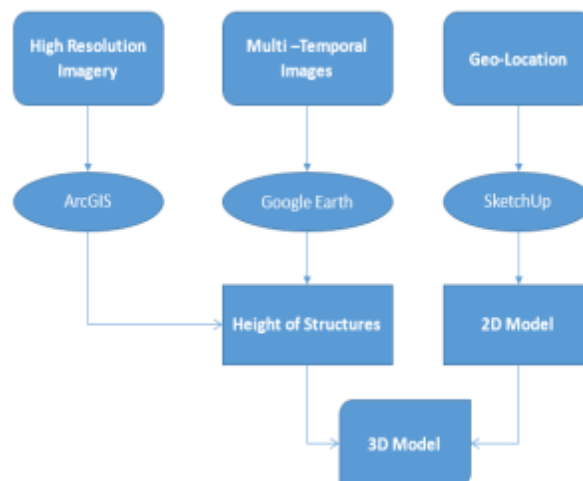


Figure 2: Flow diagram for 3D model

4. Results and discussion

4.1 3-Dimensional model

3D modelling for a sample of thirteen buildings was performed. 3D modelling also included height estimation from the satellite imagery. Now, using SketchUp, footprints of the

buildings are designed first, and then elevation is provided to the respective buildings. After this, one particular building, Hansalya building, is chosen to design the layout of PV Modules and then, solar generation of the rooftop is calculated.

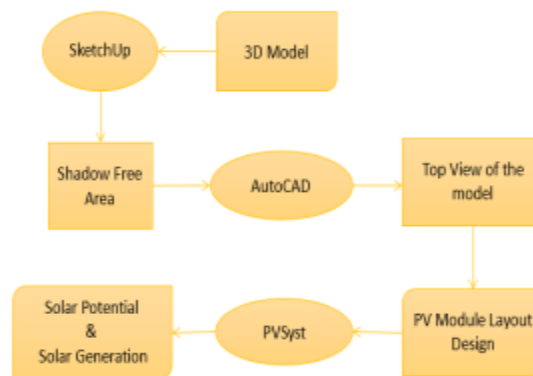


Figure 3: Flow diagram for Solar Generation estimation



Figure 4: Flow diagram for Solar Generation estimation

4.2 Shadow analysis

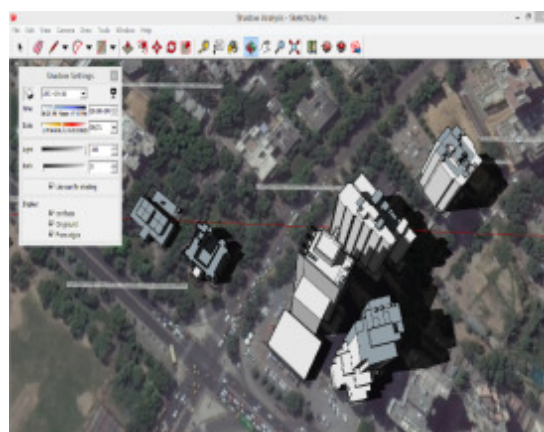


Figure 5: June 21st, 2014, 9:00 am

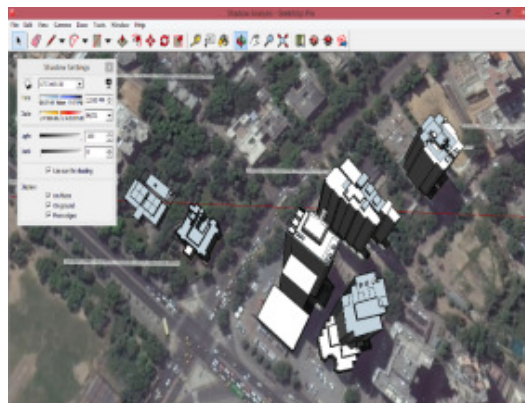


Figure 6: June 21st, 2014, 12:00 pm

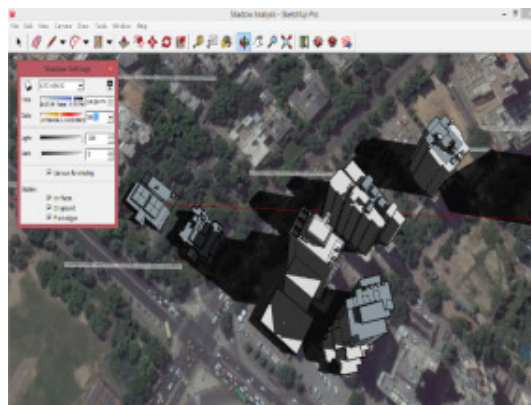


Figure 7: June 21st, 2014, 04:00 pm

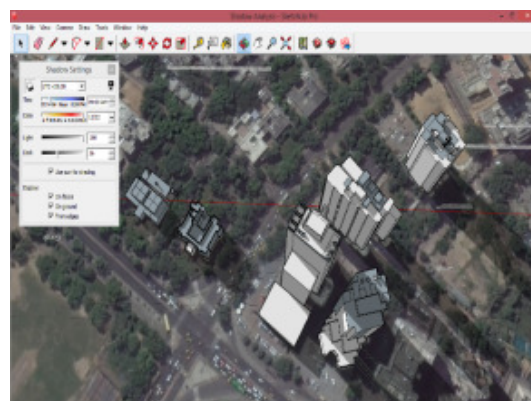


Figure 8: Dec 22nd, 2014, 09:00 am

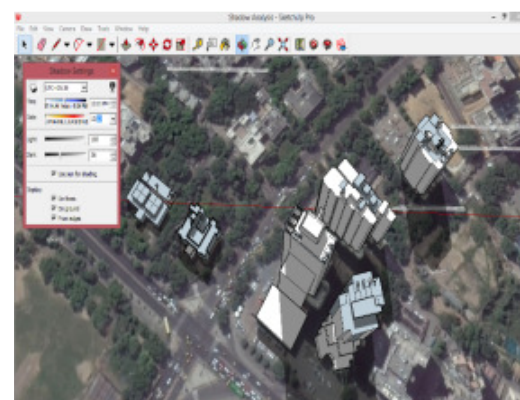


Figure 9: Dec 22nd, 2014, 12:00 pm

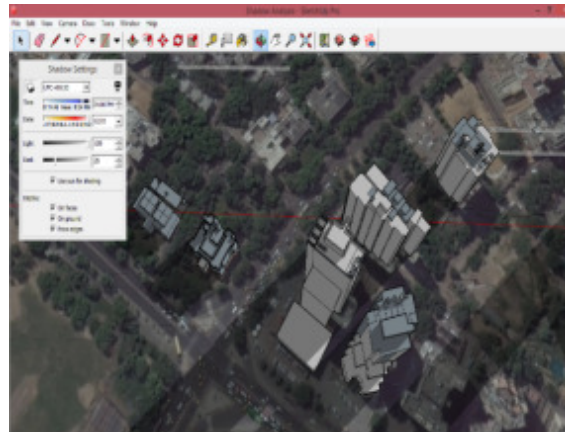


Figure 10: Dec 22nd, 2014, 04:00 pm

4.3 Module installation layout design

Hansalaya Building, New Delhi model is used for designing the layout for the rooftop installation of PV Modules and to find out its monthly and annual solar generation. With a height of 88 meters, Hansalaya is one of the highest buildings in New Delhi. The green colored rectangular boxes represent layout of PV modules and orange color shows the area which receives shadow during a year, thus, this area is not recommended for installation of PV modules.

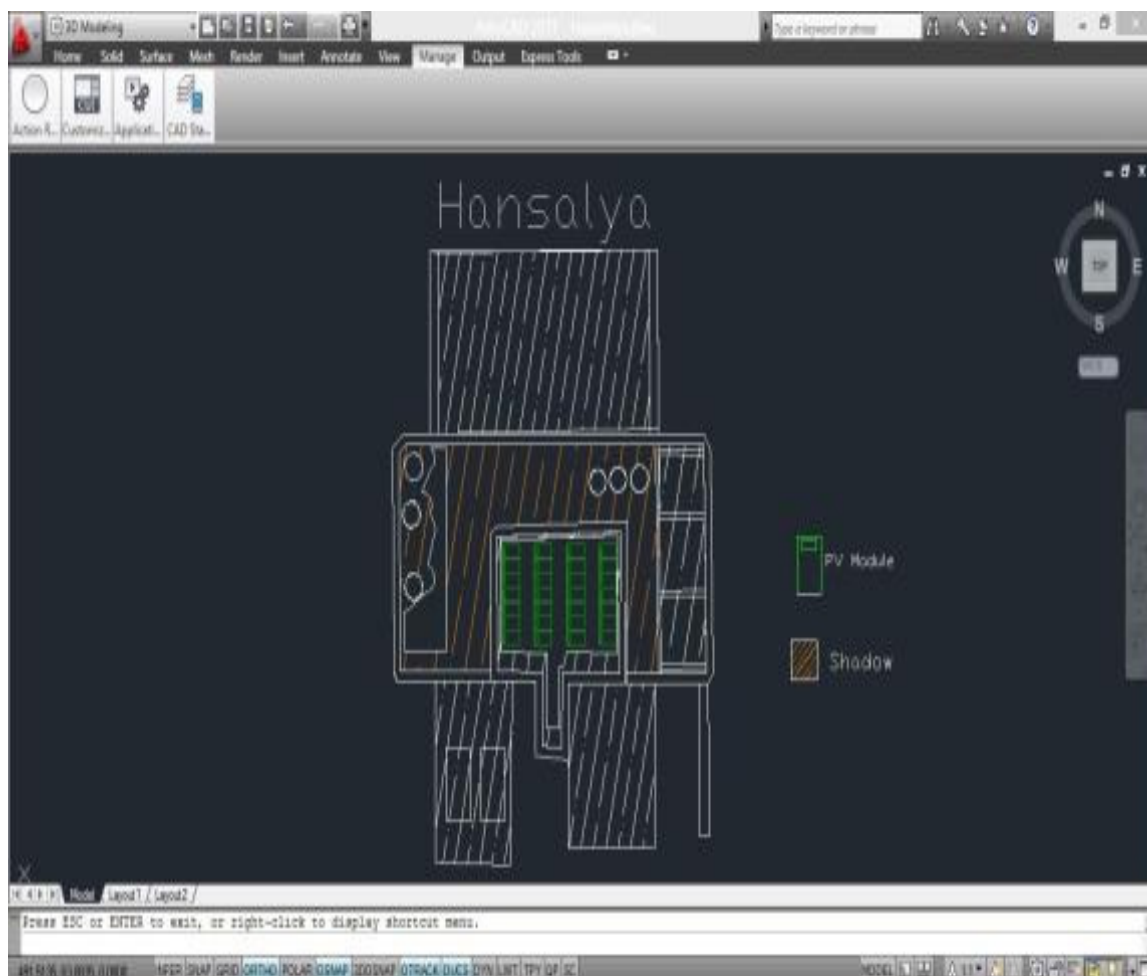


Figure 11: Module Layout Plan

4.4 Solar Generation

PVSYST V6.38		15/07/15 Page 1/4	
Grid-Connected System: Simulation parameters			
Project : Grid-Connected Project at Project/New Delhi			
Geographical Site	Project/New Delhi	Country	India
Situation	Latitude 28.6°N	Longitude	77.2°E
Time defined as	Legal Time Time zone UT+5.5	Altitude	208 m
Meteo data:	New Delhi Synthetic - Meteonorm 7.1 (2001-2010)		
Simulation variant : Hansabays, New Delhi			
Simulation date 18/07/15 14h02			
Simulation parameters			
Collector Plane Orientation	Tilt 19°	Azimuth	32°
4 Sheds	Pitch 3.40 m	Collector width	1.96 m
Inactive band	Top 0 m	Bottom	0 m
Shading limit angle	Gamma 22.42°	Occupation Ratio	57.6 %
Models used	Transposition Perez	Diffuse	Erls, Meteonorm
Horizon	Free Horizon		
Near Shadings	Mutual shadings of sheds		
PV Array Characteristics			
PV module	Si-poly	Model CS6X - 320P	
	Manufacturer	Canadian Solar Inc.	
Number of PV modules	In series 14 modules	In parallel	2 strings
Total number of PV modules	Nb. modules 28	Unit Nom. Power	320 Wp
Array global power	Nominal (STC) 8.96 kWp	At operating cond.	7.90 kWp (50°C)
Array operating characteristics (50°C)	U mp 456 V	I mp	16 A
Total area	Module area 83.7 m²	Cell area	49.1 m²
Inverter	Model PVS300-TL-8000W-2		
	Manufacturer ABB		
Characteristics	Operating Voltage 330-800 V	Unit Nom. Power	8.00 kWac
Inverter pack	Nb. of inverters 1 units	Total Power	8.0 kWac
PV Array loss factors			
Array Soiling Losses		Loss Fraction	2.0 %
Thermal Loss factor	Uc (const) 29.0 W/m²K	Uv (wind) 0.0 W/m²K / m/s	
Wiring Ohmic Loss	Global array res. 444 mOhm	Loss Fraction	1.5 % at STC
LID - Light Induced Degradation		Loss Fraction	1.0 %
Module Quality Loss		Loss Fraction	1.5 %
Module Mismatch Losses		Loss Fraction	1.0 % at MPP
Incidence effect, ASHRAE parameterization	IAM = 1 - bc (1/cos i - 1)	bc Param.	0.05
System loss factors			
AC wire loss inverter to transfo	Inverter voltage 230 Vac in		
	Wires: 2x240.0 mm²	Loss Fraction	0.5 % at STC
External transformer	Iron loss (24H continuous) 9 W	Loss Fraction	0.1 % at STC
	Relative/Inductive losses 61.8 mOhm	Loss Fraction	1.0 % at STC
User's needs :	Unlimited load (grid)		
Auxiliaries loss	Proportional to Power 5.6 W/kW	... from Power thresh.	0.0 kW

Project Evaluation mode

PVSYST V6.38		15/07/15 Page 2/4	
Grid-Connected System: Simulation parameters (continued)			

System loss factors			
AC wire loss inverter to transfo	Inverter voltage 230 Vac in		
	Wires: 2x240.0 mm²	Loss Fraction	0.5 % at STC
External transformer	Iron loss (24H continuous) 9 W	Loss Fraction	0.1 % at STC
	Relative/Inductive losses 61.8 mOhm	Loss Fraction	1.0 % at STC
User's needs :	Unlimited load (grid)		
Auxiliaries loss	Proportional to Power 5.6 W/kW	... from Power thresh.	0.0 kW

Project Evaluation mode

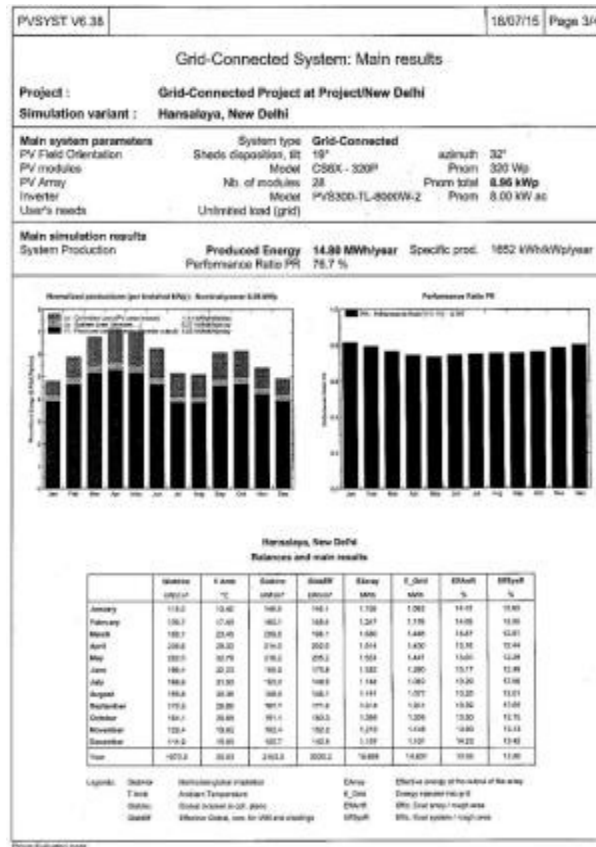


Figure 12: PV Syst based solar power generation statistics for the study area

4.5 Model Validation

Table 1: True heights, calculated heights and % error

Building Name	True Height, y (m)	Calculated Height, x (m)	Difference (m)	Error (%)
The Park	38.4	37.0	-1.41	-3.671
India Gate	42.0	39.5	-2.50	-5.952
The Taj Palace Hotel	45.0	43.0	-2.00	-4.444
Surya Kiran	46.0	43.0	-3.00	-6.522
DLF (Janpath)	46.0	46.0	0.00	0.000
Prakash Deep	46.1	43.0	-3.09	-6.704
Mercantile House	46.1	48.0	1.91	4.144
Vandana	49.9	53.0	3.06	6.127
Jeevan Prakash	51.2	55.2	3.99	7.791
Himalaya House	52.0	56.0	4.00	7.692
Sansad Bhavan	61.5	61.0	-0.46	-0.748
BOB (Sansad Marg)	64.0	64.0	-0.01	-0.016
HT Media House	80.0	75.0	-5.00	-6.250
Vikas Minar	82.0	79.5	-2.50	-3.049
Ambadeep	84.0	83.0	-1.00	-1.190
Hansalaya	88.0	85.0	-3.00	-3.409
Mean Error (%)	-1.080067			
RMSE	2.706239			
Correlation Coefficient, R	0.987320			

Where,

y: true height of the building (in meters) (*Tallest buildings in New Delhi, 2015*)

x: calculated height of the building (in meters)

- Error (E) = (Calculated Height – True Height) / True Height
- Root Mean Squared Error (RMSE)

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - x_i)^2}$$

Where, N: Total number of buildings

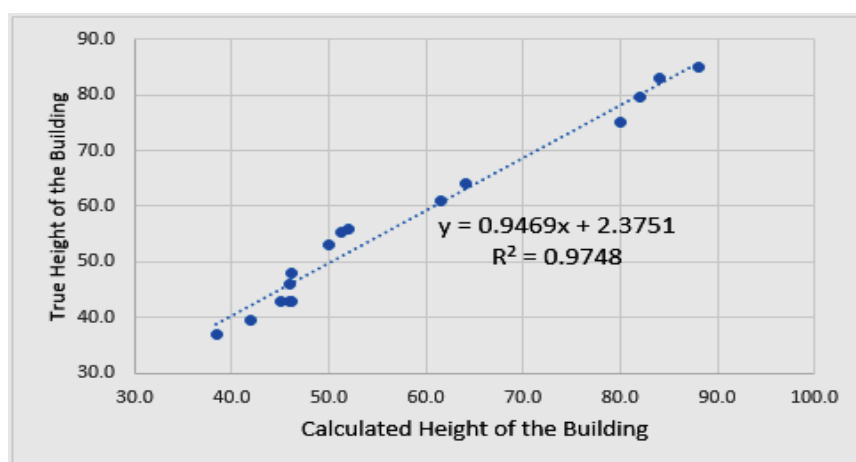


Figure 13: Data Correlation

1. Mean error of 1.080067% is observed for the above set of calculated heights.
2. Root Mean Square error (RMSE) in the above set of calculated heights is 2.706239.

3. 97.48% of the variability in actual heights can be predicted by the variability in calculated heights of the buildings in this model. ($R^2 = 0.9747$)
4. Calculated height for a building may vary depending upon analyst's accuracy. Thus, error will vary from analyst to analyst.

5. Conclusion

In this study, ArcGIS 10.3.1 software is used to calculate the height of a building by using a high-resolution imagery of 1.28m spatial resolution, of Worldview-3 sensor. Correlation coefficient (R) proves that 97.48% (i.e. $R^2 = 0.9748$) of the variability in true height of the buildings can be predicted by the variability in calculated height of the buildings by this method. But it becomes very difficult to calculate heights which are obstructed due to occlusions. Using photovoltaic layout, solar generation on the rooftop has also been calculated. This study proposed a method for significant reduction of time and cost involved in field visits for the installation of photovoltaic modules on a roof in any Rooftop Solar Power Project.

6. References

1. Climate of Delhi, Wikimedia Foundation.
2. Delhi, Wikipedia (2015, July 5), Retrieved from <https://en.wikipedia.org/wiki/Delhi>
3. Luqman, M., Ahmad, S.R., Khan, S., Ahmad, U., Raza, A. and Akmal, F., (2015), Estimation of Solar Energy Potential from Rooftop of Punjab Government Servants Cooperative Housing Society Lahore Using GIS, Smart Grid and Renewable Energy, 6, pp 128-139.
4. Press Trust of India (2014, July 15), Power demand in Delhi touches all-time high. Delhi, India.
5. Profile of Delhi: National Capital Territory – Delhi. (N.D).
6. REN21 (2016), Renewables 2016 Global Status Report: key findings, Renewable Energy Policy Network for the 21st century.
7. Smiti Mittal, (2015, June 22), India Achieves 4 Gigawatts Installed Solar Power Capacity, Clean Technica.
8. Tallest buildings in New Delhi, Published on July 8th, 2015.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/295907307>

Big Data Security – Challenges and Recommendations

Research · March 2016

CITATIONS

2

READS

5,621

3 authors, including:



[Vaibhav Hans](#)

University of Petroleum & Energy Studies

3 PUBLICATIONS 23 CITATIONS

[SEE PROFILE](#)



[Neelu J. Ahuja](#)

University of Petroleum & Energy Studies

19 PUBLICATIONS 52 CITATIONS

[SEE PROFILE](#)

Big Data Security – Challenges and Recommendations

Renu Bhandari¹, Vaibhav Hans^{2*} and Neelu Jyothi Ahuja³

^{1,2*}University of Petroleum and Energy Studies, India

³Centre of Information Technology, University of Petroleum and Energy Studies, India

www.ijcseonline.org

Received: Dec/11/2015

Revised: Dec/23/2015

Accepted: Jan/12/2016

Published: Jan/30/ 2016

Abstract— This paper focuses on key insights of big data architecture which somehow lead to top 5 big data security risks and the use of top 5 best practices that should be considered while designing big data solution which can thereby surmount with these risks. Big data architecture, being distributive in nature can undergo partition, replication and distribution among thousands of data and processing nodes for distributed computation thus supporting multiple features associated with big data analytics like real time, streaming and continuous data computation along with massive parallel and powerful programming framework. These series of characteristics are put into effect via a key setup that somehow leads to certain crucial security implications. The challenges induced by this can be handled via big data technologies and solutions that exist inside big data architecture compound characterized for specific big data problems. Big data solutions should provide effective ways to be more proactive against fraud, management and consolidation of data, proper security against data intrusion, malicious attacks and many other fraudulent activities. In particular, this paper discusses the issues and key features that should be taken into consideration while undergoing development of secured big data solutions and technologies that will handle the risks and privacy concerns (e.g. Data security, insecure computation and data storage, invasive marketing etc.) associated with big data analysis in an effective way to increase the performance impact, considering that these risks are somehow a result of characteristics of big data architecture.

Keywords— Big Data; Hadoop; MapReduce; Secure Computation

I. INTRODUCTION

The term big data is coined to describe voluminous amount of unstructured and semi-structured data with three characteristics: ‘volume, variety and velocity’ considering that the instance, the volume (amount of data), variety (complexity of multiple data types) and velocity (data in motion) of the data produced increases, data is defined as Big Data. Big Data architecture is distributive in nature scaling upto thousands of data and processing nodes. Among these thousands of nodes, the data gets partitioned, replicated and distributed for powerful computation, and because of performance reasons, data is also segmented into classes. Features like auto-tiering, real-time processing and streaming of data have been major trends in big data analysis. A growing number of companies are using the technology to store and analyze petabytes of data including web logs, click stream data and social media content to gain better insights about their customers and their business. Hence, classification of information is becoming more critical.

A number of software firms have been working on applications and solutions related to big data basically designed to bring the power of analytics to the masses. One of the chief risk factor surrounding big data computation and management is the unawareness of the potential future

downsides being associated with failure to manage it, clearly making the risk factors transparent for all big data sets – unstructured, structured and all grey areas in between – becoming the top business priorities.

In this paper, we highlight top five security and privacy challenges that are specific to big data. We went through a number of journals, studied published researches, and big data related books to initialize a list of high-priority security and privacy problems and finally arrived at the top five challenges to big data computation along with basic recommendation of how to deal with these problems. The list of challenges is as follows:

1. Insecure Computation
2. Input Validation and Filtering
3. Granular Access Control
4. Insecure Data Storage
5. Privacy concerns in Data Mining and Analytics.

II. KEY INSIGHTS OF BIG DATA ARCHITECTURE

Big Data Architecture is premised on a set of skills for the development of scalable, reliable, and completely automated pipelines of huge amount of data. Big data is still till date quite a young field which therefore leads to ‘no standard big data architecture available’ that has been used for a long time. Properties like latency, volume, velocity, variety, veracity, capability for ad-hoc queries, scalability,

robustness and fault tolerance have become key features that are mandatory for choosing any big data architecture. Although some intrinsic properties like auto-tiering, easy shift of code through disks for analysis have been into practice for a long time, these also undergo some security issues. These issues have been discussed thoroughly in the coming sections.

A. Basic Big Data Framework

Big Data Architecture is distributed in nature and can scale up to thousands of data and processing nodes. In big data architecture the data is partitioned, replicated and distributed among those thousands of nodes. Because of performance reasons, data is partitioned into two classes – hot data and cold data giving a nice feature to big data architecture called auto-tiering [6]. Easy move of code through disks rather than the data has also been another big paradigm shift from traditional to modern architecture. Modern architecture supports real time computation (real – time analytics), collection of data from variety of input sources [fig. 2] and transfer of that data to the big data solutions on a regular basis. Ad-hoc queries along with massive parallel and powerful programming layout provides flexibility. Many other frameworks like mapreduce where a program gets divided into multiple maps that get executed on respective data nodes and when finally they are reduced to a single result set, Storm Topology (Spouts & Bolts) ,

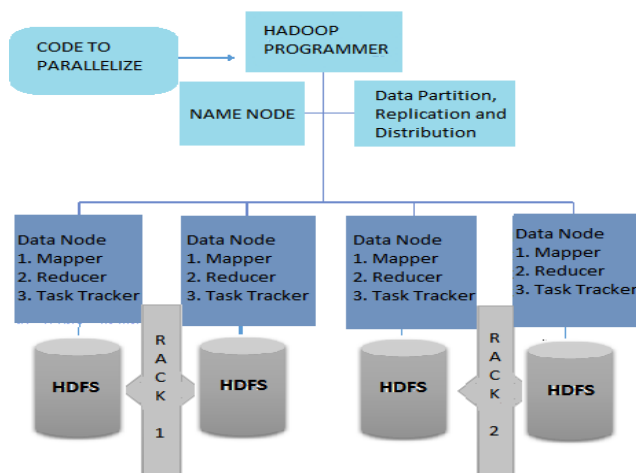


Fig. 1 DISTRIBUTED ARCHITECTURE

where Spouts are data sources and Bolts are data processing nodes following network topology for real time computation are being used. However there is no single silver bullet as Hadoop is already unsuitable for many big data problems like real time analytics, graph computation, low latency queries etc. and Storm topology.

B. Auto-tiering

Storage arrays have become good at managing huge amount of data that can dynamically move between different disk types and RAID levels in an array. Automated Storage Tiering, a storage software management feature, manages the need of maintaining space, performance and cost requirements for data processing. General policies are set-up by storage administrators that deal with partition of data into two major classes followed by assigning of memory addresses to data according to their respective classes.

For performance reasons, data is partitioned into two classes, hot data and cold data. Data that is used frequently, say for analysis or prediction analysis is classified as hot data. On the other hand, data that is used temporarily or less frequently falls under the category of cold data. Cold data is assigned to slower, less-expensive SATA storage, and hot data is moved to high-performing SAS or SSDs. However, data is automatically classified and migrated to the optimum tier of storage disk drives as its activity level rises or falls away. Auto tiering has security implications, there comes a certain amount of time when you do not know where the data resides.

C. Real-Time, Streaming and Continuous Computation.

Performing real time computation is the next big trend in big data. With applications like google analytics, real time monitoring of websites, webpages can be done. Huge amount of terabytes of data is collected from various sources, filtered, analyzed via multiple data mining, data classification and prediction algorithms and hence reports are maintained of all these analysis. These reports thereby help in decision making for better performance of organizations. Stream Processing Language is a real – time data processing language used to process data streams coming from multiple sources [6]. SPL comes with three basic types of operators – Utility, Relational and Arithmetic which take data through Input Source Operator and give output through Output Source Operators. These multiple operators present in between the source filter, aggregate, join multiple data streams accordance to the need of the user. Arrangements of the operators can be done manually by the users as per the requirements. This adds-up as a more efficient way of processing streaming data.

Another key feature of big data is that it supports ad-hoc queries. As an example from traditional database is that, end-users are able to create SQL queries at their end and are able to submit them to respective web applications directly. This will definitely give more flexibility, more power. But there are greater security concerns of having ad-hoc queries.

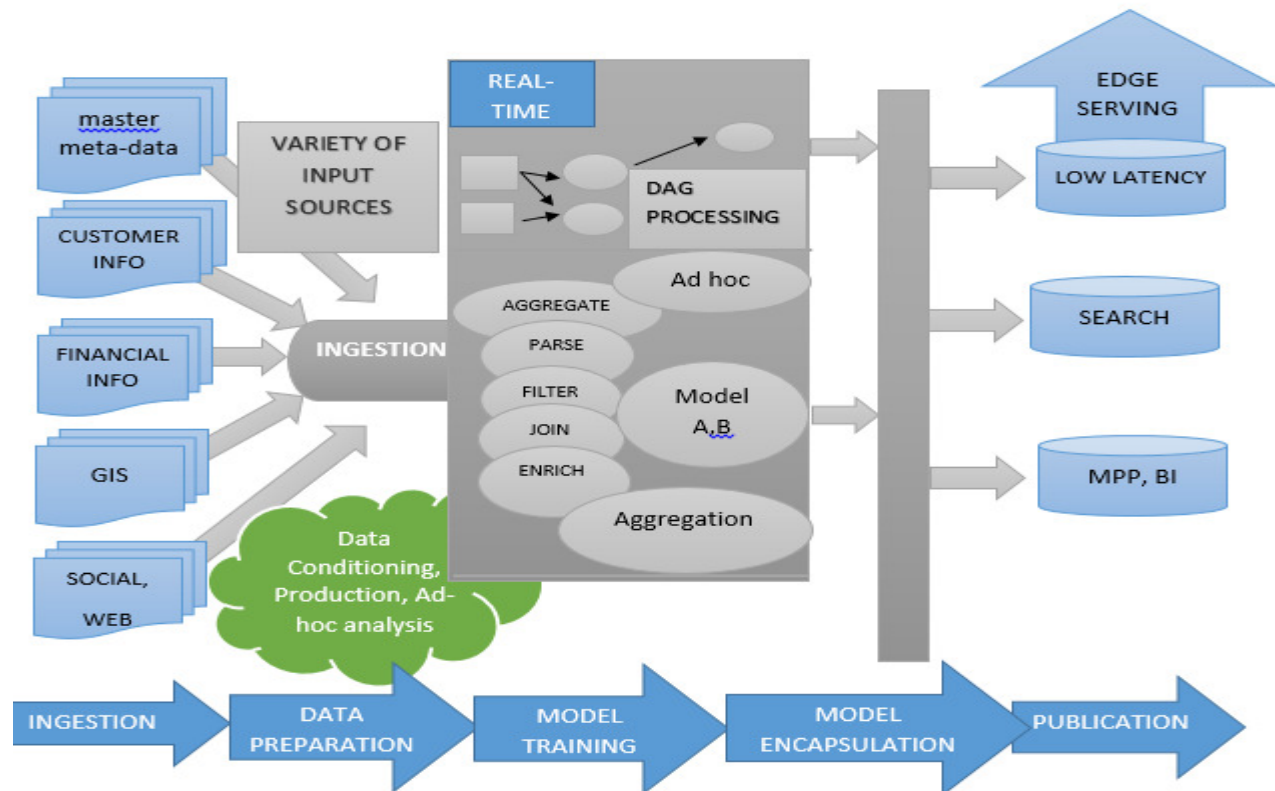


Fig. 2 REAL-TIME STREAMING, AND CONTINUOUS COMPUTATION

D. Parallel and Powerful Programming Framework

Big Data has massive parallel and powerful programming framework. Suppose if you have 16 TB of data, and the data is divided into 128 MB chunks, then your program will be divided into 82000 Maps or functions that will run concurrently on data processing nodes. Big Data Programming framework is very powerful. User can develop program in Java rather than SQL/PLSQL databases [6].

There are multiple frameworks available for Big data computation. MapReduce framework is used by Hadoop where a program gets divided into multiple map that get executed at multiple data nodes and then finally the results are merged together into single result set [3].

Topology based computation is yet another major framework associated with big data computation. This paradigm is used by a real time analysis big data solution called Storm. Storm utilizes a network topology of Spouts and Bolts. Here Spouts act as data sources, whereas Bolts act as data processing nodes. A lot of frameworks and topologies exist for data processing and management in the field of big data computation.

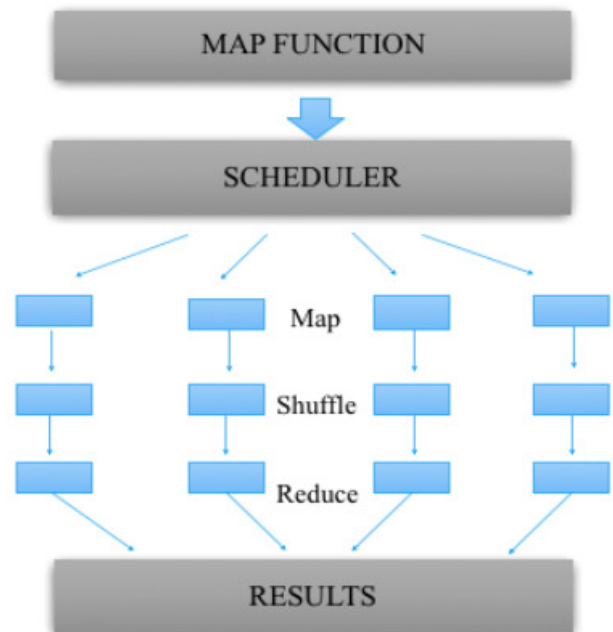


Fig. 3 MAPREDUCE FRAMEWORK

III. TOP 5 SECURITY RISKS

The security mechanisms in big data technology is generally weak. Hence implementing the above 3 features in big data architecture has been inevitable and of a big concern. Finding robust security mechanisms for the purpose of using features like auto – tiering, parallelism etc. has been a challenging problem. Issues like invasion of privacy, complexity of disk drive storage, invasive marketing etc. have led to challenges in implementing Big Data Analytics tools for Big data solutions and applications.

A. Insecure Computation

Untrusted computational programs are used by attackers in order to extract and turnout sensitive information from data sources. Insecure computation apart from causing information leak can also corrupt your data, leading to incorrect results in prediction or analysis. It can also result into Denial of Services (DoS) on your big data solution disabling the property of using massively parallel programming language.

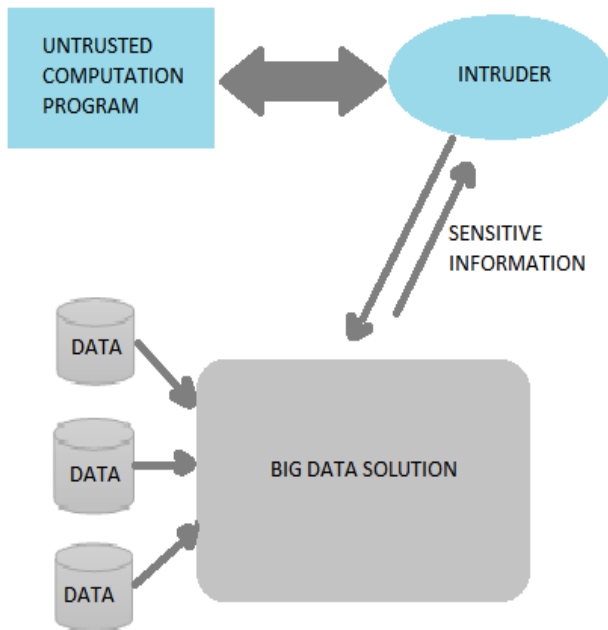


Fig. 4 INSECURE COMPUTATION

B. Input Validation and Filtering

Big Data needs to collect input from variety of sources, therefore it is quite important and mandatory to validate the input [4]. This involves making a decision of what kind of data is untrusted and what are the untrusted data sources. It also needs to filter rogue or malicious data from the good one. These two challenges are not new, i.e. these two challenges are also present in traditional databases, however in big data the huge amount of gigabytes and terabytes of

continuous data flow makes it really very difficult to perform input validation or data filtering on the incoming batch of data [5]. Signature based data filtering also has limitations like it cannot filter rogue or malicious data having some behavioral aspect. When a large amount of malicious data is inserted into the dataset, its influence on the result produced is *massive*. Signature based data filtering is incapable of tracking down such attacks, thus individual custom algorithms need to be designed to deal with such cases.

C. Granular Access Controls

Big data was traditionally designed for performance and scalability with almost no security in mind. Traditional databases have very comprehensive table, row and cell level access control, and these have been really gone missing in big data solutions. Ad-hoc queries pose another additional challenge to big data solutions where user can retrieve sensitive information out of the data using ad-hoc queries. Even though being provided by a big data solution, access control is disabled by default. EX: NO-SQL databases depends upon access control provided by third party softwares, but it is actually disabled by default and you need to enable it explicitly.

D. INSECURE DATA STORAGE

As data is stored at thousands of nodes – ‘authentication, authorization and encryption of data at those nodes becomes a challenging work’. Auto-tiering moves cold data to lesser secure medium – ‘What if cold data is sensitive?’ .If any solution provides encryption of real time data, it may not be useful, as encryption of real time data may have performance impacts. Secure communication amongst various nodes, middlewares, and end users is disabled by default, hence it needs to be enabled explicitly [4].

E. Privacy Concerns in Data Mining and Analytics

Monetization of Big Data involves Data Mining and Analytics and sharing of those analytical results involves multiple challenges like invasion of privacy, invasive marketing and unintentional disclosure of information. Quite a few examples of these include - AOL release of anonymized search logs where users could be identified easily, which is really concerning.

IV. TOP 5 BEST PRACTICES.

Traditional security mechanisms that focus on securing small scale data that is basically static (not streaming), are practically inadequate. Big data security issues are magnified with the velocity, volume and variety of big data. Big data processing requires ultra-fast response times for computation, and adding up multiple feature to the framework adds up the security implications. However,

these security issues can be handled by taking proper adequate steps. Five of the best recommendations for big data security have been discussed below.

A. *Secure Your Computation Code*

In order to prevent malicious data from entering your big data solution, implement access control, code signing and dynamic analysis of the computational code. Proper strategies need to be made having the capability of controlling the impact of untrusted code on the data, once it has been able to get into the big data solution. There are generally two ways of preventing attacks: securing the data when insecure mapper is present, and securing the mapper (related to MapReduce framework of Big Data Hadoop).

B. *Implement Comprehensive Input Validation and Filtering.*

For better processing and security practices, implementation of comprehensive input validation and filtering on almost all internal and external sources is mandatory. On the other hand, proper evaluation of key input validation and filtering features of respective big data solution is required to see if it can scale up the data requirement for respective big data solutions [5]. This may be implemented by building algorithms that will validate the input for large sets of data.

C. *Implement Granular Access Control.*

Reviewing and configuring the role and privilege matrix of different the kinds of users of big data which can be the admin, knowledge workers, end users, developers etc. is the core part for the implementation of granular access control. Ad-hoc queries are required for data computation, and the queries processed should be verified for what access has to be given to the respective ad-hoc queries [5]. By default the access control is disabled, it needs to get enabled explicitly for proper access to the data and its sources. This section of recommendation in simple words means preventing the data to be accessed by the people that should not have access to the data.

D. *SECURE YOUR DATA STORAGE AND COMPUTATION.*

Protecting data storage and computation sections of big data analytics becomes a prime area to focus on as much part of sensitive data leakage portions are encountered in this phase. For this, the sensitive data should be segregated. Enabling Data Encryption for sensitive data and audit administrative access on Data Nodes marks to be a major step in this scenario [5]. And finally the verification of proper configuration of API security of all components of big marks to be the final step for secure data storage and computation.

E. *Review and Implement Privacy Preserving Data Mining and Analytics.*

For proper preservation of sensitive information, verification of analytical algorithms designed for data mining, pattern classification and recognition is necessary. This will reduce the rate of disclosure of sensitive information. Before any further actions, big data implementation should be pen tested. It is important to establish guidelines and recommendations for the prevention of inadvertent privacy disclosures.

CONCLUSION

Big data is trending. No new application can be imagined without it producing new forms of data, operating on data-driven algorithms, and consuming specified amount of data. With data storing and computing environments becoming more cheaper, cloud environments becoming more capable of storing and sharing system and analytics applications, software applications becoming more networked, data security, access control, compression – encryption and compliance have introduced several challenges that practically need to be handled and addressed in a very systematic manner.

In this paper, we have walked through the top five big data security challenges and have laid some recommendations for making big data processing and computation more reliable and in turn making its infrastructure more secure. Some common elements in this list of the top five security issues that are specific to big data arise from the multiple infrastructure tiers – (both computing and storage) used for big data processing, the new computation infrastructures like NoSQL databases used for fast throughput that are necessary for big volumes of data are not thoroughly secured from major security threats, the non – scalability of real – time monitoring techniques, the heterogeneous layout of devices producing data, confusion with diverse legal restrictions that somehow lead to ad-hoc approaches for security and privacy. There is a big ecosystem existing for specific big data problems. Topics in this paper serve to clarify specific aspects of the vulnerable areas in the entire big data processing infrastructure that need to be analyzed for certain threats. Major recommendations for dealing with the top five security risks have also been suggested in this paper.

Our hope is that this paper will collaboratively increase the focus of the research and development community towards the top five challenges, which will ultimately lead to greater security and privacy in respective big data platforms.

ACKNOWLEDGMENTS

We would like to acknowledge our fellow colleagues and researchers for their work and support.

REFERENCES

- [1] Big data. In *Wikipedia, The Free Encyclopedia*. Retrieved 08:36, November 10, **2015**
- [2] Apache Hadoop. In *Wikipedia, The Free Encyclopedia*. Retrieved 10:28, November 20, **2015**
- [3] MapReduce. In *Wikipedia, The Free Encyclopedia*. Retrieved 08:43, January 15, **2016**
- [4] IBM Security Intelligence with Big Data, In IBM. Retrieved 09:38, November 22, **2015**
- [5] Big Data Research, Security in big data research papers, Retrieved 08:10, December 10, **2015**
- [6] Anuja Pandit, Amruta Deshpande and Prajakta Karmarkar, Log Mining Based on Hadoop's Map and Reduce Technique, Int. Journal of Computer Sciences and Engineering, Volume -05, Issue -04, Page No (1-4), April **2013**



Dr. Neelu Jyothi Ahuja

Neelu Jyothi Ahuja has received her PhD on development of a rule based expert system for seismic data interpretation, from University of Petroleum and Energy Studies, Dehradun. She has 17 years of experience in teaching, research and project proposal development and has published papers in journals and conferences at international and national level. She is

currently executing 03 research projects and supervising doctoral work of 04 scholars. With keen interest in intra-disciplinary research, her research interests include Expert Systems, Knowledge-based tutoring, Artificial Intelligence, Object oriented development and Programming Languages. She is head computing research institute, a virtual R & D center and an Associate Professor in computer science and engineering department, at College of Engineering Studies at University of Petroleum and Energy Studies, Dehradun.

AUTHORS PROFILE



Ms. Renu Bhandari

Renu Bhandari, an undergraduate scholar from University of Petroleum and Energy Studies is pursuing her degree in Computer Science with specialization in Business Analytics and Optimization. She has worked on many analytics based projects that involve innovation and R&D. Her research interest is in the area of Data Mining

and Analytics and she is currently working as an intern in Indian Statistical Institute, Bangalore. Her work and experience strongly deal with Machine Learning, Big Data, Data Mining, and Python.



Mr. Vaibhav Hans

Vaibhav Hans is a student of University of Petroleum & Energy Studies. He's pursuing his degree in Computer Science Engineering with specialisation in Business Analytics. He's currently working in upcoming field of Analytics and Internet of Things. His areas of expertise is object oriented programming, Machine Learning

Data Analytics & Big Data.

Design and Verification of 16-Bit Vedic Multiplier Using 3:2 Compressors and 4-Bit Novel Adder

K. Venkata Siva Reddy, P. Vishnu Kumar, K. Maheswari
and B. Khaleelu Rehman

Abstract This paper describes the implementation of a 16-bit Vedic multiplier enhanced in terms of propagation delay when compared with array multiplier and Vedic multiplier using ripple carry adder. In our design, we have utilized ripple carry adder, 4-bit novel adder, 3:2 Compressor. The propagation delay comparison was extracted from the synthesis report and static timing report as well as circuits have to be designed and simulated using DSCH3.1 and the results are to be compared with different technologies.

Keywords Ripple carry adder · 4-bit novel adder · 3-2 Compressor propagation delay

1 Introduction

Multipliers are an integral part of most processing units hence the performance of processors greatly depends upon the functioning of their multiplication units.

Several new architectures have been proposed for improving the functioning of multiplier units to meet the constraints of reducing the delay, power, layout, and hence small area or even combination of them which makes suitable for various very high speed, low power, and compact VLSI implementations. Though, an efficient multiplier design is yet to come.

K. Venkata Siva Reddy (✉)

Ravindra College of Engineering for Women, Andhra Pradesh, India

e-mail: kvenkatsivareddy@gmail.com

P. Vishnu Kumar · K. Maheswari

G Pullaiah College of Engineering and Technology, Andhra Pradesh, India

e-mail: vishnu4b8@gmail.com

K. Maheswari

e-mail: Kotlamahi435@gmail.com

B. Khaleelu Rehman

University of Petroleum and Energy Studies, Dehradun, India

e-mail: krehman@ddn.upes.ac.in; afridi.1156@gmail.com

In order to address the disadvantages associated with multiplier architectures, Vedic mathematic approach was proposed. Multipliers were designed using Urdhwa Tiryagbhyam [1]. In this paper, an even more efficient approach to improve multiplier units compared to the Vedic multipliers is being proposed.

2 Existing Techniques

2.1 Array Multiplier

Array multiplier is an efficient layout of a combinational multiplier [2]. To multiply two unsigned 8-bit integers together, we can produce eight partial products by performing eight 1-bit multiplications one for each bit is multiplicand to produce the product, add all eight partial products. Consider an array multiplier for two binary numbers X and Y , of p bits each as shown in Fig. 1. There will be p^2 partial

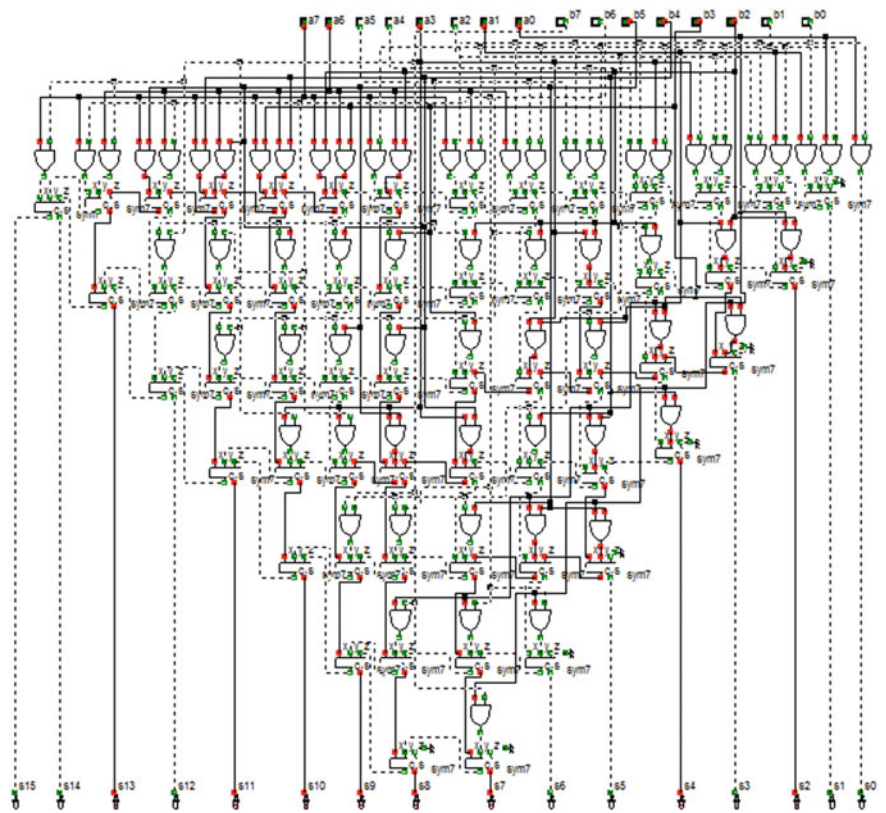


Fig. 1 A 8-bit array multiplier

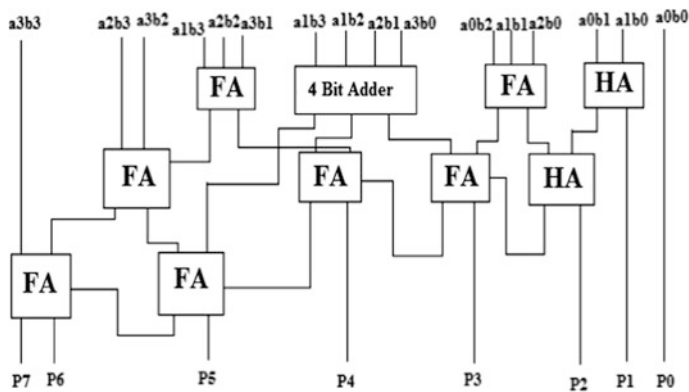


Fig. 2 4-bit novel adder-based multiplier

products that are produced in parallel by a set of p^2 and gates. A $p \times p$ multiplier requires $p(p - 1)$ full adders and Sp^2 gates.

The limitations in the array multipliers are high power consumption as well as number of devices utilization. Delay is the time taken by the signals to propagate through the gates and in array multiplier, the worst case delay would be $(2p + 1) t_d$ due to the gates that form the multiplication array. Thus array multipliers are less economical with more hardware complexity.

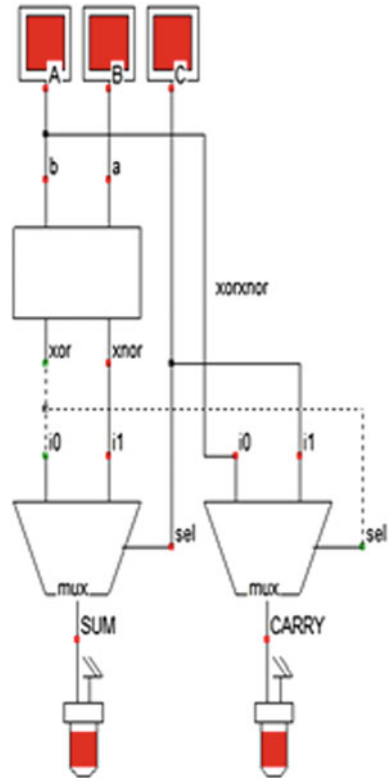
2.2 Vedic Multipliers: Using Urdhwa Tiryagbhyam Sutra and 4-Bit Novel Adder

Vedic multiplier designed using Urdhwa Tiryagbhyam sutra and 4-bit novel adder [3] is shown in Fig. 2. The novel 4-bit adder performs the addition of 4 bits at a time and produces three output bits. These three output bits comprise of one sum bit and two carry bits. The 4-bit adder adds the four input bits at a time and the speed of the multiplication increases.

By using this 4×4 multiplier we can design the architecture for 8×8 multiplier also. Though the multiplier reduces the design complexity and power drastically the delay can still be reduced further [4].

3 Proposed Multiplier

In this proposed multiplier design, we are introducing compressors in the existing Vedic multiplier using novel adder. A compressor is a device that reduces the combination of input bits at the output. Shown in Fig. 3, is a 3:2 adder compressor that functions similar to a full adder.

Fig. 3 3:2 compressor

It takes three inputs A, B, and C to generate two outputs, the sum, and the carry bits. Equations for sum and carry bits are governed by 1 and 2

$$\text{Sum} = (A \oplus B) * \overline{C} + \overline{(A \oplus B)} * C \quad (1)$$

$$\text{Carry} = (A \oplus B) * C + \overline{(A \oplus B)} * A \quad (2)$$

This compressor is built using xor-xnor and multiplexer modules. We are replacing the full adders in the Vedic multiplier design using novel 4-bit adder by 3:2 compressor. Even though a 3:2 compressor works the same as full adder, the difference lies in propagation delay. A full adder needs two half adders which are in turn built using 'xor' and 'and' gates. The delay produced by a full adder is 0.027 ns whereas a 3:2 compressor adder produces a delay of only 0.019 ns (Fig. 4).

Hence we are improving the delay and power consumption very efficiently in the above multiplier architecture though the area constraint can still be improved further using other techniques (Fig. 5).

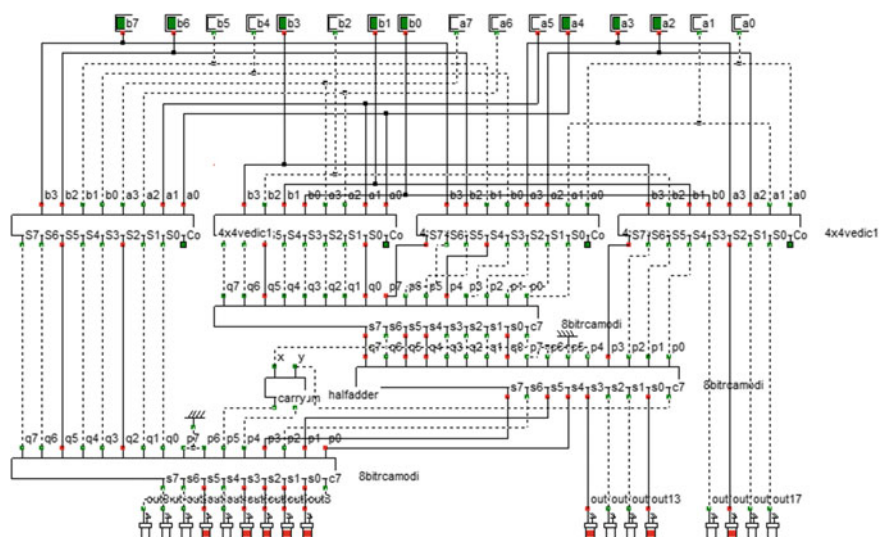


Fig. 4 8-bit Multiplier using 3:2 compressor adder and 4-bit novel adder

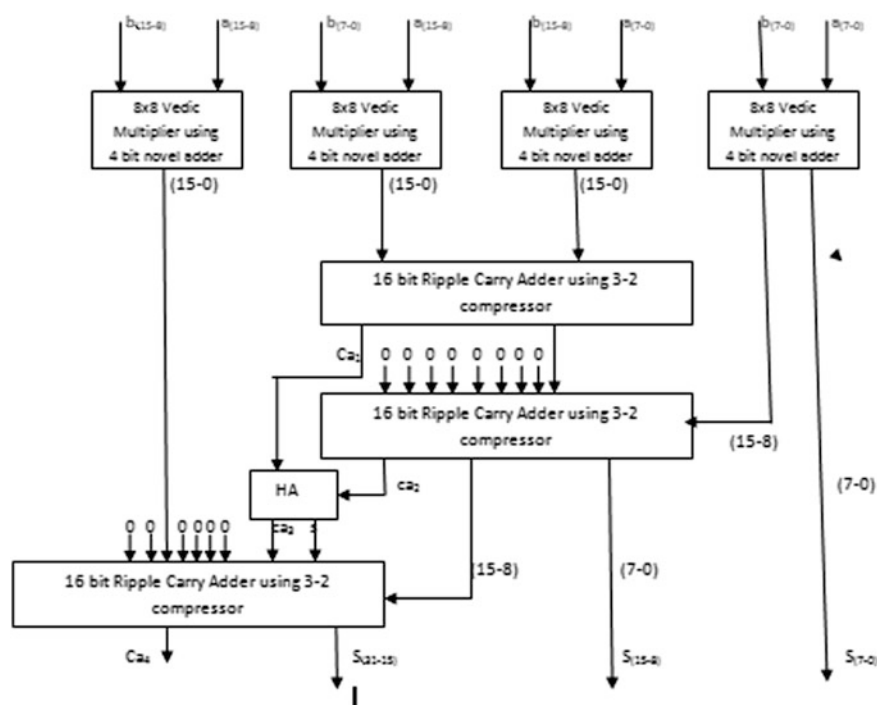


Fig. 5 16-bit Multiplier using 3:2 compressor adder and 4-bit novel adder

4 Results

4.1 Simulation Results

Using the Xilinx 9.2i version the simulation results were found as shown

- (i) For 8-bit proposed multiplier (Fig. 6).
- (ii) For 16-bit proposed multiplier (Fig. 7).

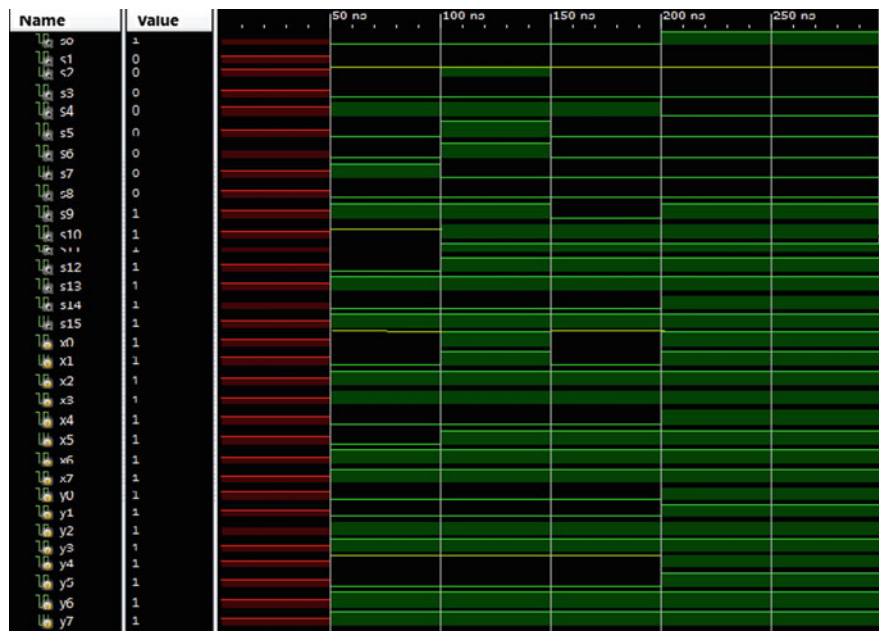


Fig. 6 Simulation output

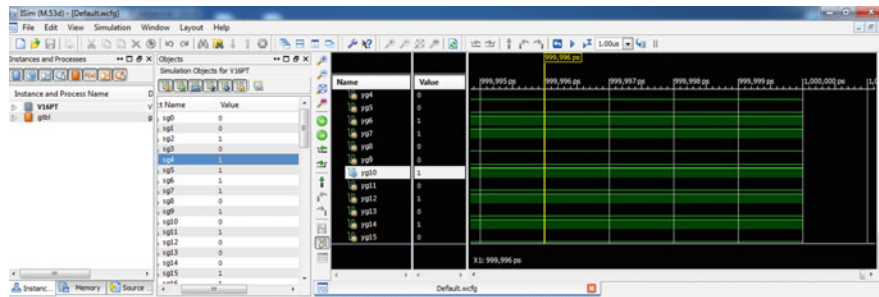


Fig. 7 Simulation output

4.2 Delay Tables and Graphs

- (i) For 4-bit proposed multiplier (Fig. 8 and Table 1).
- (ii) For 8-bit proposed multiplier (Fig. 9 and Table 2).

Fig. 8 Delay table, graph for 4 bit

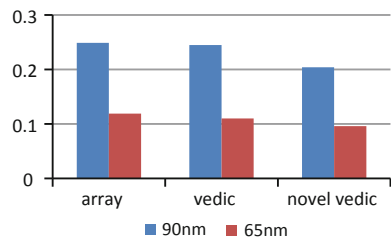


Table 1 Comparison of Delay for various 4-bit multipliers in multiple technologies

Multiplier	Technology in nm	
	65	90
Array multiplier	0.119 ns	0.249 ns
Vedic multiplier	0.110 ns	0.245 ns
Proposed Vedic multiplier	0.096 ns	0.204 ns

Fig. 9 Delay table, graph for 8 bit

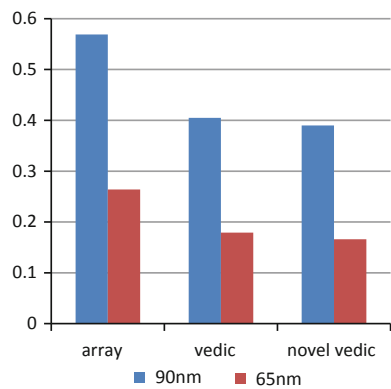


Table 2 Comparison of Delay for various 8-bit multipliers in multiple technologies

Multiplier	Technology in nm	
	65	90
Array multiplier	0.264 ns	0.569 ns
Vedic multiplier	0.179 ns	0.405 ns
Proposed Vedic multiplier	0.166 ns	0.370 ns

4.3 RTL Schematics

- (i) For 8-bit proposed multiplier (Fig. 10).
- (ii) For 16-bit proposed multiplier (Fig. 11).

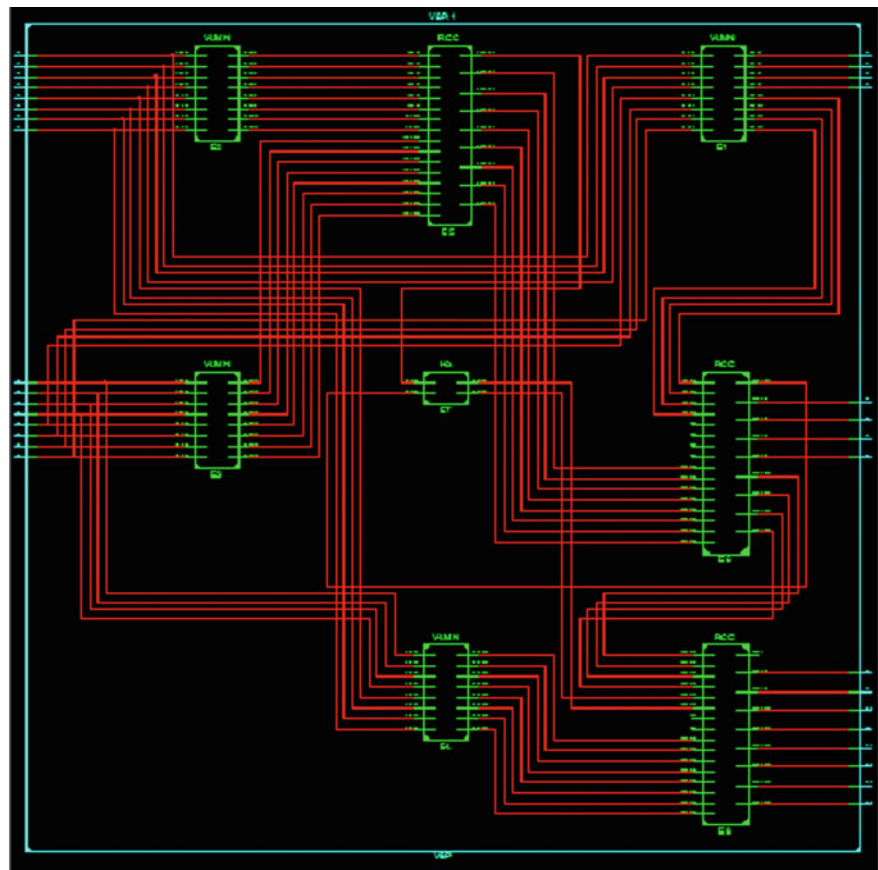


Fig. 10 Schematics for 8 bit

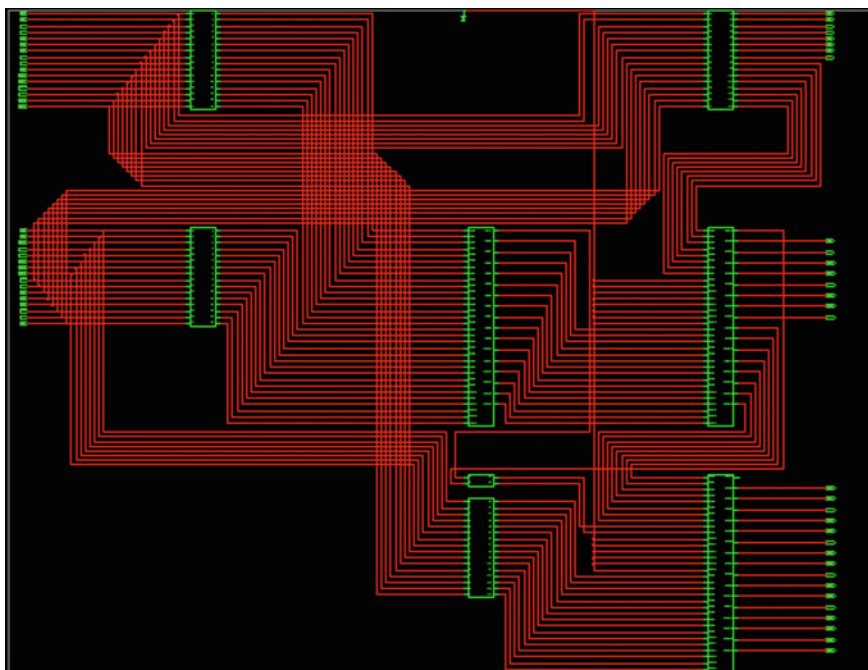


Fig. 11 Schematics for 16 bit

5 Conclusion

The proposed Vedic multiplier using 4-bit novel adder and 3:2 compressors has produced an improved performance compared to its predecessors by reducing the delay. These multipliers can improve the performances of applications in which they are used.

6 Future Scope

The proposed Vedic multiplier using novel 4-bit adder and 3:2 compressors can also be implemented for 32, 64 bits, and for higher bits. The hardware area can be decreased considerably, and it can be further improved.

References

1. MD. Belal Rashid, Balaji B.S. and Prof. M.B. Anandaraju, "VLSI Design and Implementation of Binary Number Multiplier based on Urdhva Tiryagbhyam Sutra with reduced Delay and Area", *International Journal of Engineering Research and Technology*, ISSN 0974-3154 Volume 6, Number 2 (2013), pp. 269–278.
2. M.Padmaja, A.Saida, "Design and Implementation of High Speed and Low Power Multiplier using Urdhwa Tiryagbhyam Sutra", *IJERT*, ISSN: 2278-0181 Vol.3 Issue 3, March-2014.
3. Rajasekhar. N, Shanmuganatham. T, "A Novel 4 Bit Adder Based Urdhwa Tiryagbhyam Multiplier", *IJCSMC*, Vol. 2, Issue. 10, October 2013, pp. 219–225.
4. Pavan Kumar, Sai Prasad Goud, A.Radhika, "FPGA Implementation of High Speed 8 bit Vedic multiplier using barrel shifter", 978-1-4673-6150-7/13/\$31.00 ©2013 IEEE.
5. Sushma R. Huddar and Sudhir Rao Rupanagudi, Kalpana M., Surabhi Mohan, "Novel High Speed Vedic Mathematics Multiplier using Compressors", 978-1-4673-5090-7/13/\$31.00 ©2013 IEEE.
6. Hsiao, Shen-Fu, Ming-Roun Jiang, and Jia-Sien Yeh, "Design of high speed low-power 3-2 counter and 4-2 compressor for fast multipliers," *IEEE Electronics Letters*, vol. 34, no.4, pp. 341–343, Feb. 1998.
7. Prabir Saha, Arindam Banerjee, Partha Bhattacharyya, Anup dandapat, "High speed ASIC design of complex multiplier using vedic mathematics", *proceeding of the 2011 IEEE students technology Symposium* 14–16 January, 2011, IIT Kharagpur, pp. 237–241.

Design Improvement and Assessment of Efficiency of Three Phase Induction Motor Operating Under the Rated Voltage

Rajeev Gupta, Devender Kumar Saini,
Raj Kumar Saini and Piush Verma

Abstract Three-phase induction motors which are designed for rated voltages but operating under the rated voltages particular in the rural areas increases stator and copper losses which leads to reduction in efficiency, temperature rise, and shorter life time of the machine. Far from the utility centers the voltage profile in most of the distribution networks is poor due to reactive power demand. This paper describe a consideration in design of a three-phase induction motor which can give desirable performance under the rated voltages even when operating without any protection system. The objective of this paper is to improve the design of the induction motors to get its rated performances which are designed for the rated voltage, but generally operating under the rated voltages.

Keywords Three-phase induction motor • Improved design • Efficiency • Under rated voltages • Ampere conductor

Rajeev Gupta
Faculty of Applied Sciences, University of Petroleum
and Energy Study, Dehradun, Uttarakhand, India
e-mail: rajeevgupta@ddn.upes.ac.in

D.K. Saini
Faculty of Electrical Engineering, University of Petroleum
and Energy Study, Dehradun, Uttarakhand, India
e-mail: dksaini@ddn.upes.ac.in

R.K. Saini (✉)
Department of Electrical Engineering, University of Petroleum
and Energy Study, Dehradun, Uttarakhand, India
e-mail: rajsaini.acet@gmail.com

Piush Verma
Rayat-Bahra Group of Institutes, Patiala Campus (Pb.), Mohali, India
e-mail: pverma19@yahoo.co.uk

1 Introduction

Apart from the industrial sector, induction motors find extensive use in the agricultural sector to drive irrigation pumps in rural areas even operating under the rated voltage. In Ref. [1], the author describes some methods to improve the efficiency of three-phase induction motors operating at low voltages in industry and for the general purposes. In this paper, twenty five elements have been taken in order to calculate the design of the motor and twelve variable elements have been taken which affects the efficiency of the machine. These twelve variables are valued independently to study the effects on the efficiency and characteristic of the motor. The author of Ref. [2], carried out a detailed study for the modification of a three-phase induction motor under the wide range of load and supply voltage and proofed that the efficiency of a new design is about 2.5 % more than that of the standard motor and also find that the input kva is about 3 % less than that of standard motor, but also investigated that the cost of proposed design is about 15 % more than that of the standard design and said that this extra cost is pay back within a reasonable period of time. In Ref [3], the author describes that the unbalanced voltages reduces the efficiency and power factor of the machine. By increasing the positive sequence of the voltage, a reduction in power factor also has been absorbed. In Ref. [4], author proofed that the worst scenario will appear when under voltages and over voltages appear with unbalanced voltages. Jawad Faiz a senior member of IEEE, in Ref [5], pointed out that the available definitions of unbalanced voltages are uncompleted. To prove the claim, a three phase 25-hp squirrel cage induction motor is analyzed under the unbalanced voltage and verified the theoretical results with experimental results. In Ref. [6], the IEEE112-B, IEC34-2, and JEC37 international standard for induction motor efficiency are evaluated and found that the rated efficiency of the machine depends upon the standard followed. On the bases of the results IEEE112-B can be considered the best suitable standard for the stray load loss measurement, hence for the motor efficiency measurement IEC34-2 and JEC 37 overestimate the motor efficiency because they only define instead of measuring the stray load losses. The author aided that stray load loss calculations are very important to calculate the correct value of efficiency. In Ref. [7], the author pointed out that the energy efficiency of continuous motor (S1) is a subject of more interest compared to intermittent duty motors (S3). The efficiency of such types of motors can be increased by proper selection of stack length and rotor material of the machine. In Ref. [8], the author concluded with the help of analytical and experimental study that the derating factor depends upon voltage unbalanced factor and magnitude of positive sequence voltage and also suggested that positive sequence voltage must be consider together with unbalanced voltage factor. In Ref. [9], the main focus of this paper is to calculate the efficiency of the induction motor under the variation of load with the variation of unbalanced, over and under voltages. In Ref. [10, 11], the authors of these research papers describe the various effects and definitions of unbalanced voltages on the operation performance of induction motors. The major effects of unbalanced voltages will lead to change in characteristic of an induction

Table 1 Input Specifications

Particulars	Inputs
Rating of the machine (Pi)	2.2 Kw
Number of phases (m)	3
Supply voltage (Es)	415 V
Power factor (pf)	0.825
Supply frequency (f)	50 Hz
Full load efficiency (n)	0.8 %
Synchronous speed (Ns)	1500 R.P.M
Electric loading (ac)	21,000

motor, due to this reason the real and reactive power consumption of an induction motor will be changed which will affect the voltage stability margin. Due to positive sequence voltage which is higher than that of the rated voltage, may be another cause of unbalance of voltages. The voltage unbalance cases may be due to single phase under voltage unbalance, two phases under voltage unbalanced, three phases under-voltage unbalanced, single phase over-voltage unbalanced, two phases over voltage unbalanced, three phases over voltage unbalanced, unequal single phase and angle displacement and unequal two phase angle displacement. To calculate the unbalanced voltages, different definitions are discussed by different authors. The definition of voltage unbalanced generally used by the power community is the ratio of negative sequence voltage to the positive sequence voltage. However in electrical machine community IEEE and NEMA use the following definitions of voltage unbalanced.

$$\% \text{ unbalanced} = \text{Maximum deviation from average voltages} / \text{Average voltage} * 100$$

In order to investigate the performance of the three-phase squirrel cage induction motor in terms of rated voltage and under the rated voltage, the input parameters of the three phase induction motor are shown in Table 1.

1.1 Analysis of Efficiency Under the Rated Supply Voltage

In this research, article the performance of three-phase induction motor operating under the rated voltage mostly due to the presence of unbalanced voltage present in the three phases have been investigated. This is studied that in the rural areas the phase-to-phase voltage in the three phases are less than the normal voltage. The variable losses and efficiency of the motor are analyzed at rated and under the rated voltages as discussed in Table 2. Constant losses have been taken as 2.5 % of the output of the machine. Still up to date research on the induction motors, the effects of the unbalanced voltages have been pointed out, the remedies of the system is still pending. The rectification of the voltage from the utility side is not much possible,

Table 2 Calculation of efficiency under the rated voltage

S. no.	% Unbalanced	Ave(v)	% below the rated voltage	Stator copper losses (w)	Rotor copper losses (w)	% Efficiency
1	0.55	363	12.30	226.96	145.22	80.69
2	1.97	370.67	10.68	22.27	142.25	80.91
3	1.70	371.67	11.65	226.34	144.82	80.72
4	2.70	370	10.84	222.18	142.50	80.89
5	1.16	370.67	11.95	222.27	142.25	80.91
6	1.33	374	9.87	220.29	141.00	81.00
7	0.18	365.67	11.96	225.51	144.30	80.76
8	0.09	365.67	11.88	225.30	144.17	80.77
9	1.06	376	9.38	219.12	140.26	81.06
10	1.15	375.67	9.48	219.31	140.38	81.05
11	2.51	370.67	10.68	222.27	142.25	80.91
12	1.2	367.33	11.48	224.28	143.52	80.81
13	0.36	365.67	11.88	225.30	144.17	80.77
14	1.10	362	12.77	227.59	145.61	80.66

but to save the energy crises the modification in the design of the motor is possible to achieve the desired performance of the machines operating under the rated voltage. The aim of this research paper is to modify the design of the machine by adjusting the ampere conductors and others parameters. In Table 2 the efficiency of the machine is worked out under the rated voltage (by taking the average voltage of the three unbalanced phases) without modifying the ampere conductors and other related parameters, while in Table 3 parameters are calculated by applying the rated voltage 415 V and under the rated voltage 368 V. 368 V is the approximate average voltage of all the average of 14 observations.

The voltage in all the three phases have been calculated regularly, discussed in Table 2. From this unbalanced voltage average voltage of all the three phases, percentage below the rated voltage, stator and copper losses are calculated to calculate the efficiency.

1.2 Results and Discussions

The calculations for the average voltage of the three phases which generally lies under the rated voltage due to the presence of unbalanced voltages in the three phases have been discussed in Table 2. Variation of supply voltage gives rise to additional losses in the machines. It means that millions of induction motors in the rural areas are running under the rated voltages which directly affect the efficiency of the machines. The proposed efficiency of the machine can be maintained up to

Table 3 Calculations of stator and rotor parameters at rated voltage and under rated voltage

Parameters	Computer added parameters (415 V)	Unmodified parameters (368 V)	Modified parameters (368 V)	% change in parameters
Output coefficient	97.066	97.066	92.444	-4.99
Product of $D^2 L$	0.00137 m	0.00137 m	0.00144 m	4.86
Diameter of core	0.10527 m	0.10527 m	0.10699 m	1.60
Length of core	0.1239 m	0.1239 m	0.1259 m	1.58
Net iron length	0.1023 m	0.1023 m	0.1039 m	1.53
Flux per pole	0.004507	0.004507	0.004656	3.20
Stator turns/phase	434	434	373	-16.35
Stator slot pitch	0.01377 m	0.01377 m	0.01399 m	1.57
Stator conductors	2604	2604	2238	-16.35
Conductors/slot	2604	2604	2238	-16.35
Slot pitch	13.785 mm	13.785 mm	14.0113 mm	1.61
Teeth width	4.985 mm	4.985 mm	5.211 mm	4.33
Flux density	1.473 web	1.473 web	1.432 web	-2.86
Length of turns	0.6779 m	0.6779 m	0.6851 m	1.19
Area of stator core	0.00187 m ²	0.00187 m ²	0.00194 m ²	3.60
Depth of stator core	0.01836 m	0.01836 m	0.01866 m	1.60
Stator resistance	9.231 Ω	8.1860 ohms	7.1099 Ω	-15.13
Outer diameter of laminations	159.59 mm	183.997 mm	186.325 mm	1.24
Total copper loss in stator	198.524 w	233.879 w	194.45 w	-20.27
Air gap length	0.428 mm	0.428 mm	0.432 mm	0.925
Diameter of rotor core	104.413 mm	104.41 mm	106.131 mm	1.61
Rotor pole pitch	12.609 mm	12.609 mm	12.817 mm	1.62
Current in each rotor bar	206.589 A	232.974 A	200.229 A	-16.35
Area of each bar	34.431 mm ²	38.82 mm ²	33.371 mm ²	-16.35
Width of rotor slots	4.130 mm	4.529 mm	4.0337 mm	-12.29
Copper loss in bars	117.73 W	132.766 W	115.439 W	-15.0
End ring current	427.652 A	482.271 A	414.486 A	-16.35
Area of each end ring	71.275 mm ²	80.37 mm ²	69.081 mm ²	-16.35
Depth of end ring	7.919 mm	8.931 mm	7.675 mm	-16.35
Total copper loss	127.239 w	143.269 w	124.88 w	-14.72
Rotor frequency	2.708 Hz	3.0286 Hz	2.66 Hz	-13.83
Full load efficiency	82.05 %	80.83 %	82.24 %	1.71

Table 4 Analysis of copper losses and efficiency

S. no	Average voltage	Modified ampere conductors	Losses (S + R) without modified parameters	Losses (S + R) with modified parameters	% Efficiency without modified parameters	% Modified efficiency
6	374	20000	361.29	319.25	81.00	82.24
11	370.67	19080	364.52	319.57	80.01	82.24
12	367.33	19070	367.8	319.05	80.81	82.25
13	365.67	19065	369.47	318.77	80.77	82.26
1	363	19061	372.18	319.38	80.69	82.24
14	362	19059	373.2	319.39	80.66	82.24

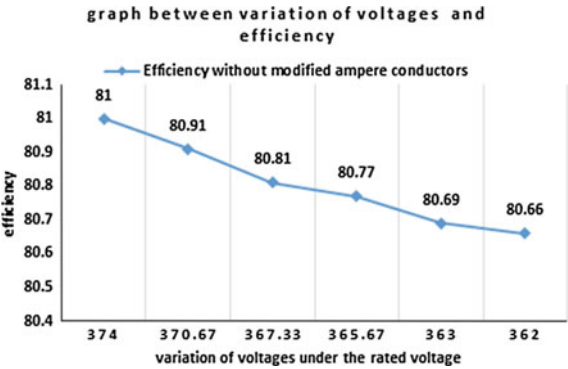
desirable limit by selecting the proper selection of the electric loading, and others parameters as discussed in Table 3, generally when the machine is operating under the rated voltage.

In Table 3. The change in parameters has been investigated without modified and with modified ampere conductors under the rated voltage 368 V. To study the performance of the motor and other characteristics under the rated voltage approximately 56 parameters are observed by putting the input data into the computer with the help of Mat Lab-Gui as a research tool, out of which 32 major parameters are modified to increase the efficiency.

In Table 4 six values of average voltage are exerted from Table 2 to observe the variation of efficiency with and without modifying the value of ampere conductors and other parameters. This is observed that the efficiency of the motor can be improved up to its rated efficiency with the modification of the ampere conductors.

Figures 1 and 2 show that the efficiency of the machine can be achieved up to its rated efficiency by modifying the ampere conductors and controlling all others parameters discussed in Table 3.

Fig. 1 Variation of voltages Vs Unmodified efficiency



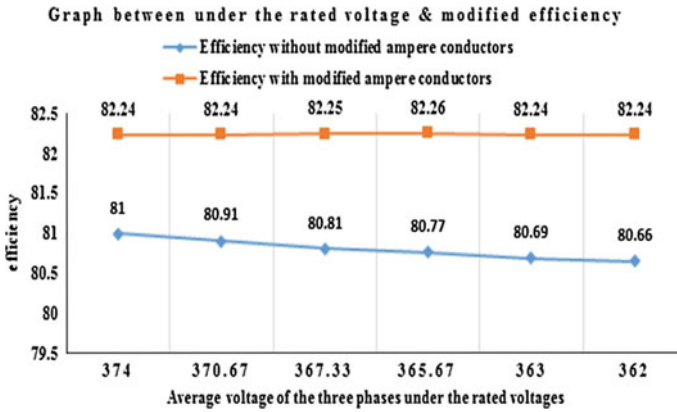


Fig. 2 Variation of voltages Vs Modified efficiency

2 Conclusion

The efficiency of three phase induction motor decreases while operating under the rated voltages. This reduction leads to an increase of total power losses. The present study shows that if a motor operating under the rated voltage is properly derated, even then the efficiency of the motors can be improved up to its rated efficiency. This can be achieved by reducing the value of electric loading thereby controlling the copper losses and 32 others parameters as discussed in Table 3. This type of study is very important for those rural areas where the network of power distribution is poor. To get the maximum efficiency, further scope of research related to this article is to optimize the ampere conductors and other related parameters of three phase induction motor operating under the variation of voltages by adding more experimental results.

References

1. Hasuike, K: Efficiency improvement study of low voltage three phase squirrel cage induction motor for general purpose. IEEE Trans on power apparatus system, VOL.PAS-102, No.12, (1993).
2. Sridhar, L: Design of energy efficient motor for irrigation pumps operating under realistic conduction IEE Roc.-Electr. Power Appl., Vol. 141, No. 6, (1994) 269–274.
3. Lee CY: Effect of unbalanced voltages on the operation of performance of a three phase induction motor. IEEE Trans Energy Convers (1999) 202–208.
4. Gnacinski, p: Winding temperature and loss of life of an induction machine under voltage unbalanced combined with over or under voltages. IEEE Trans Energy (2008) 363–371.
5. Jawad Faiz: Influence of Unbalanced Voltage on the Steady-State performance of a Three-Phase Squirrel-Cage Induction Motor. IEEE Trans on Energy conversion, vol. 19, no. 4 (2004) 657–662.

6. Aldo Boglietti: International Standards for the Induction Motor Efficiency Evaluation: A Critical Analysis of the Stray-Load Loss Determination. IEEE Trans on industry applications, VOL. 40, no. 5, (2004) 1294–1301.
7. Yaw-Juen Wang: Analysis of Effects of Three-Phase Voltage Unbalance on Induction Motors with Emphasis on the Angle of the Complex Voltage Unbalance Factor. IEEE transactions on Energy Conversion, vol. 16, no. 3, September 2001, pp 270–275.
8. Repo, A: Energy efficiency of hoisting motors. Electrical Machines (ICEM), IEEE International conference on 2–5 Sept (2014) 144–149.
9. Enrique C. Quispe: Influence of the Positive Sequence Voltage on the Derating of Three-Phase Induction Motors under Voltage Unbalance Electric Machines & Drives Conference (IEMDC), 2013 IEEE International on 12–15 May(2015) 100–105.
10. Giridhar Kini, P: Effect of Voltage and Load Variations on Efficiencies of a Motor-Pump System. IEEE Transactions on Energy Conversion, vol. 25, no. 2 (2010) 287–292.
11. Jawad Faiz: Precise derating of three phase induction motors with unbalanced voltages. Elsevier, Energy conversion and management (2007) 2579–2586.

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/298711125>

Effect of topography at the proximity of AC electric traction line on the earth fault distance locating algorithm

Article · February 2016

CITATIONS

2

READS

83

3 authors:



C. Ravi Kumaran Nair

1 PUBLICATION 2 CITATIONS

[SEE PROFILE](#)



Devender Kumar Saini

University of Petroleum & Energy Studies

23 PUBLICATIONS 53 CITATIONS

[SEE PROFILE](#)



Madhavan Jayaraju

College of Engineering, Munnar

25 PUBLICATIONS 72 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Electrical Traction system [View project](#)



Performance Enhancement of Power System Stabilizer [View project](#)

Effect of Topography at the Proximity of AC Electric Traction Line on the Earth Fault Distance Locating Algorithm.

Ravi kumaran Nair.C.

Divisional Electrical Engineer, Electric Traction Distribution Branch, Indian Railways, Trivandrum Division, Trivandrum, Kerala, India, nair.ravikumar@rediffmail.com

Dr. Devender Kumar Saini

Asst. Prof. Dept of Electrical & Electronics Engineering, College of Engineering Studies, University of Petroleum & Energy Studies, Dehradun, India, E-mail: dksaini@ddn.upes.ac.in

Prof.(Dr.) Jayaraju.M

(Former Director, Agency for Non-conventional Energy and Rural Technology, Ministry of Power, Govt of Kerala), India, E-mail: jayarajum@gmail.com

Abstract - Electrification of railways became necessary in suburban areas in 19th century itself due to worsening of air pollutions by the fossil fuel driven locomotives. Modern electric traction lines established for railways all over the world are mostly of 25 kV 50 Hz or 60 Hz systems. Railway tracks in suburban area shall generally be laid on plane surface. But in developing countries, the momentums to electrify their railway track continue at non-suburban areas also, where tracks are laid through earth cuttings, tunnels, over line structures etc. Such areas become the reasons for frequent earth faults on traction lines. Earth fault distances located by the relay system in those areas are found potentially deviated in irregular fashion from the actual. Existing models for high voltage AC traction are made with different impedances for different configuration of railway tracks; but with constant impedance angle only with the presumptions that the shunt admittance existing between earth and traction lines are not considerably high, and are uniformly distributed along its entire length. The protection schemes and fault distance locating algorithms are designed accordingly. Experimental studies conducted on a real electric traction network system of Trivandrum Division of Indian Railways and some of the experimental study results, which are beneficial to the railway traction system, are summarized in this paper.

Keywords: AC electric traction, Constant impedance angle, Earth cuttings, Over Head Equipment (OHE), Shunt admittance, Tunnels.

Introduction

Traction Electrification System provides electrical power to the trains by means of the traction power supply system, traction power distribution system, and traction power return system. Traction Power Supply System includes Traction Sub Stations (TSS) located along the railway track at theoretically equal distance. This distance may vary slightly with some other factors, like availability of grid sub station nearby the

proposed TSS, availability of suitable land to build up the TSS etc.. Traction Power Distribution System, also called as Over Head Equipments (OHE), is fed at single end consists of the catenary system, along track feeder system, intermediate auxiliary transformer stations (ATSS) and end-of-feeding Sectioning Post (SP). The length of the OHE fed with electric supply through a circuit breaker shall normally be 35 to 40 km. The working voltage shall be below 28 kV. Traction Power Return System comprised of the running rails, impedance bonds, cross bonds, and the ground (earth) itself. Even though the physical appearance of high voltage AC power line of railway traction has no similarity with the short length AC power transmission line, its modelling is done at par with the short length AC power transmission line. Literatures [1][3][5][7][10][12][13] on power lines characteristics indicate that the model for short length power transmission lines (of length 80 km or lesser, and with voltage 66 kV or below) comprise an inductive reactance in series with line resistance per unit length as shown in Fig.1. Effect of shunt admittance existing between power line and earth is conveniently neglected due to reason that the height of lowest live conductor itself is strung at considerably high altitude, which practically eliminates the effect of shunt admittance. Also, the presence of earth or earthed structures at the proximity shall not be there for a longer stretch.

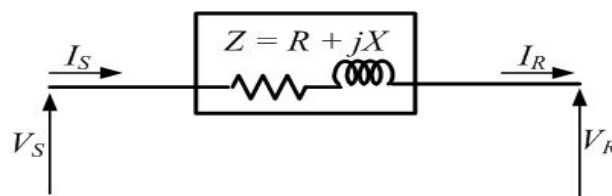


Fig.1 Generalized model of short AC power transmission line

But, the generalized model (Fig.1) can't be fit as it is on AC power high voltage traction lines for developing the algorithm for protection scheme, even when the working voltage and the

length of traction line is far below the standard stipulated for short transmission line. Electric traction lines have some unique features, like, its altitude is as low as 4.58 M from the ground (rail level), and slope profiling (earth cuttings made through hillocks to lay railway tracks, with vertical height maximum 18 M) may be available within 3 to 4 M radial distances for a longer stretch. The presences of tunnels further reduce the clearance between traction conductors and earth to 0.25 M, that too, for very lengthy stretches in hilly terrain. Fig. 2 is the generalized model of an AC traction system.

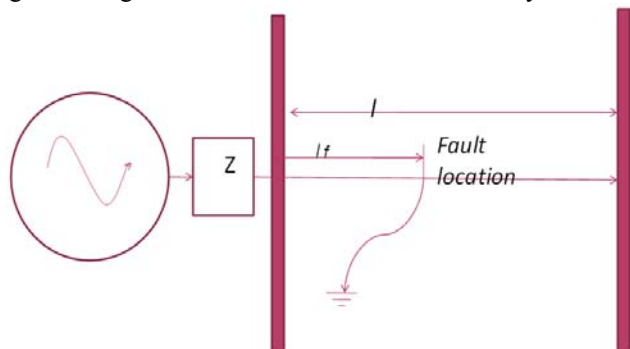


Fig. 2 The generalized AC power electric traction network system, under earth fault condition

In Fig. 2, Z is the impedance of the traction power source, l is the length of the traction system, and l_f is the distance of fault from the TSS. Here, the traction line parameters are assumed to be uniformly distributed [9], where the per unit length resistance, inductance, and capacitance are R_0 , L_0 , C_0 respectively. Where R_0 is 0.1 to 0.3 Ω /km, L_0 is 1.4 to 2.4 mH/km and, C_0 is 10 to 14 nF/km [8]

The prevailing algorithms [9][11][14] for measuring the fault distance by the Distance Protection Relay (DPR) in electric traction were made based on the loop impedance or reactance under different fault type condition of OHE. The assumption is made that the impedance angle of OHE is constant, and the impedance of the line is uniformly distributed [11]. The impedances of traction line for different layouts of railway tracks [4] are given in Table.1

Table.1. Impedances of traction line for different layouts of railway tracks

Sl. No.	Layout of track & OHE	Impedance / kilo meter
1	Single line track OHE with return conductor	$0.70 \angle 70^\circ \Omega$
2	Single line track OHE without return conductor	$0.41 \angle 70^\circ \Omega$
3	Double line track OHE with return conductor	$0.43 \angle 70^\circ \Omega$
4	Double line track OHE without return conductor	$0.24 \angle 70^\circ \Omega$

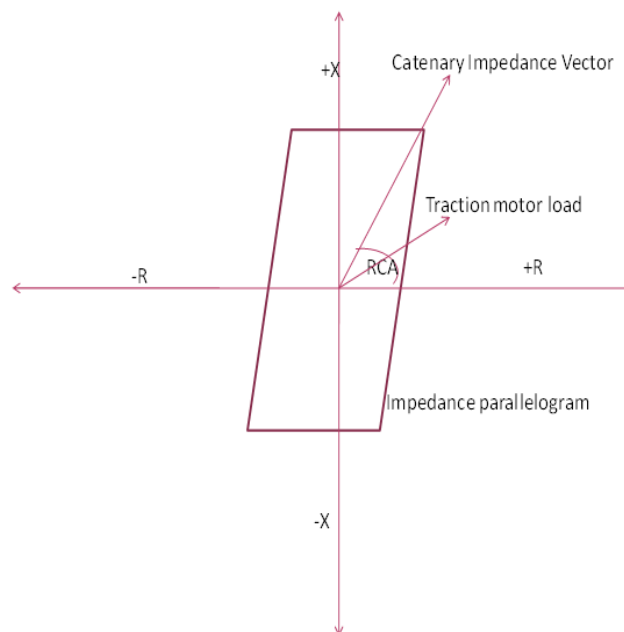


Fig. 3. Impedance parallelogram of Distance Protection Relay

Fig.3. shows how the resistive reach settings are arrived. The magnitude of the impedance of traction line is calculated using general expression, $Z = V/I$, where V is the system voltage, and I is the fault current occurred with zero impedance load.

The Relay Characteristic Angle (RCA) plays the crucial role in differentiating a fault current, and the load current contributed by the electric locomotives.

The angle of load current of the locomotive shall be around 40° , whereas the angle of earth fault current is taken as 70° for all fault distance calculation [4]. The RCA is so determined that the relay does not act at the higher traction load current contributed by the electric locomotives, and it acts when earth fault current occurs even at lesser magnitude. The angle of the earth faults current, which is determined by the angle of loop impedance of OHE, if varies; the magnitude of minimum current required to actuate the DPR shall also vary. Distance between dead earth fault and the source of supply is being determined by the magnitude of the fault current, which is directly related to the loop impedance of OHE.

Table.2 is the records [15] kept at Traction Power Control Room of Indian Railways at Trivandrum Division on naturally occurred incidents of earth fault on OHE. It can be noted from Table-2 that the DPR indicates fault distances differently for different terrain for the similar distance of fault location from the TSS. The variation from the actual was observed to be more or less linear with progressing offset value (span error) where tracks are laid on plane areas. The least variations, but in highly irregular manner, was observed at the sections where railway tracks are laid at terrain where steep slope profiling and tunnels are present at the proximity of OHE.



Fig.4. OHE at plane ground

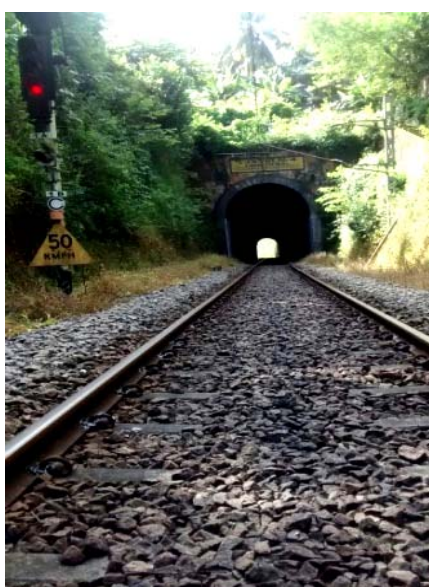


Fig. 5. OHE at Tunnel



Fig.6. OHE at steep slope profile

Table.2. Actual Earth fault distance and distance indicated by DPR in kilo meter, for two different terrains

Railway track laid on plane ground		Railway track laid on area where steep earth profiling are available	
Actual Fault distance, in km	Fault distance indicated by DPR	Actual Fault distance, in km	Fault distance indicated by DPR
5	7	7	7
7	9	8	8
9	11	9	10
17	19	20	21
29	32	32	32

Prolonged power supply interruption on electric traction system even at a smaller stretch of railway section will disrupt the train services badly in a wide rail network. Hence, it is accounted to be disastrous. Some persisting earth fault like, flashover of insulators happened with the electrocution of snake like reptiles, could not be detected from the ground. In the event of such faults on OHE, the technical staffs looking after the maintenance of OHE have to move immediately to the fault location(s) and to clear them off for the re-energization of OHE to resume normal train traffic. For which, the exact location where the earth fault persist is essentially to be informed to the maintenance staffs. In practice, the prevailing algorithms are found inadequate in precisely indicating the fault distance.

Experimental studies

Experimental studies were conducted at two distinct sections of railway traction lines by creating earth faults on OHE with practically zero impedance load (short circuit test) using discharge rod at ten different feasible locations. Six locations were at the section between Eraniel and Thovala railway stations in Tamilnadu state of India where tracks are laid on level ground (Fig.4), and four at the section between Eraniel and Balaramapuram railway stations where railway tracks are laid through steep earth cuttings & tunnels (Fig. 5 & 6) at the valley of Western Ghats in India. Geographical maps of experiment zone are shown as Fig. 8 & 9. Even when the authority is vested with the first author to conduct short circuit tests on OHE for assessing the healthiness of the protection systems, the opportunities to conduct experiments in controlled ambience duly satisfying many test conditions like, (i) to ensure no rolling stocks or trains shall be present in the entire length of experiment zone (ie, from the supply point to the farthest end of OHE), (ii) to conduct experiment only on non-rainy day, (iii) to get other power lines switched off, which run in parallel and across at the vicinity of OHE etc. was very rare. All the ten controlled experiments were conducted in between 2nd October 2014 and 12th October 2015. The sequence of experimentation was, (i) bottom clamp of discharge rod was rigidly connected to the rail, which is earthed through many earth pit electrodes, (ii) the OHE supply was switched off, (iii) spring loaded top end hook of the

discharge rod was rigidly clipped on OHE, and (iv) the OHE supply was switched on for observing the supply trip and fault distance indicated by the DPR. The earth fault distances indicated by the DPR on plane ground are shown in Table.3 and on terrain with earth cuttings & tunnels are shown in Table. 3 A

Table.3. Fault distance indicated by DPR in experimental studies on plane ground

Distance of fault indicated by DPR for the OHE at plane ground (with some minor earth cuttings of short length on en-route), fig. 4		
Experiment date	Fault created distance, in KM	Fault distance shown by DPR in KM
02/10/2014	35	40
29/10/2014	20	23
06/02/2015	14	16
26/03/2015	20	23
02/07/2015	19	22
12/10/2015	9	10

Table.3A. Fault distance indicated by DPR in experimental studies at cuttings & tunnels

Distance of fault indicated by DPR for the OHE at terrain where deep & long earth cuttings & tunnels are available on en-route, fig.5 & 6		
Experiment date	Fault created distance, in KM	Fault distance shown by DPR, in KM
14/06/2015	34	33
27/06/2015	16	18
30/06/2015	37	37
25/08/2015	19	19

Table.4. Nature of topography of the section between Eraniel and Thovala railway stations

Distance from TSS	Nature of terrain through railway track is laid
0 to 05 KM	Plane ground
05 to 10 KM	Slope profiling of length 320 mtrs and peak altitude 4 mtrs at 8 th km from TSS
10 to 15 KM	Plane ground
15 to 20 KM	Plane ground
20 to 25 KM	Plane ground
25 to 30 KM	Plane ground
30 to 35 KM	Plane ground

Table.5. Nature of topography of the section between Eraniel and Balaramapuram railway stations,

Distance from TSS	Nature of terrain through railway track is laid
0 to 05 KM	Plane ground
05 to 10 KM	Plane ground and minor slope profile, one tunnel of 80 mtrs length at 9 th km
10 to 15 KM	Plane ground and ten gradual slope profile of altitude max 3 mtrs, and of short lengths
15 to 20 KM	Fifteen minor slope profiles with altitude less than 3 mtrs, short length, Five steep slope profile of altitudes 6, 8, 8, 9 & 13 mtrs, of lengths 250, 350, 450, 900 and 450 mtrs respectively, and two tunnels of length 130 mtrs and 40 mtrs are available in this stretch.
20 to 25 KM	Eight minor slope profiles with altitude less than 3 mtrs , short length, Four steep slope profile of altitudes 22,11,14 & 6 mtrs, of lengths 600, 300,750 and 550 mtrs respectively, and one tunnel of length 201 mtrs are available in this stretch.
25 to 30 KM	Six minor slope profiles with altitude less than 3 mtrs, short length, six steep slope profile of altitudes 11,7,6,10,8 & 5 mtrs of lengths 500, 200, 350, 300, 360 & 350 mtrs respectively are available in this stretch.
30 to 37 KM	five minor slope profiles with altitude less than 3 mtrs, short length, six steep slope profile of altitudes 8, 6, 5, 6, 5 & 7 mtrs of lengths 250, 60, 210, 400, 350 & 400 mtrs respectively are available in this stretch.

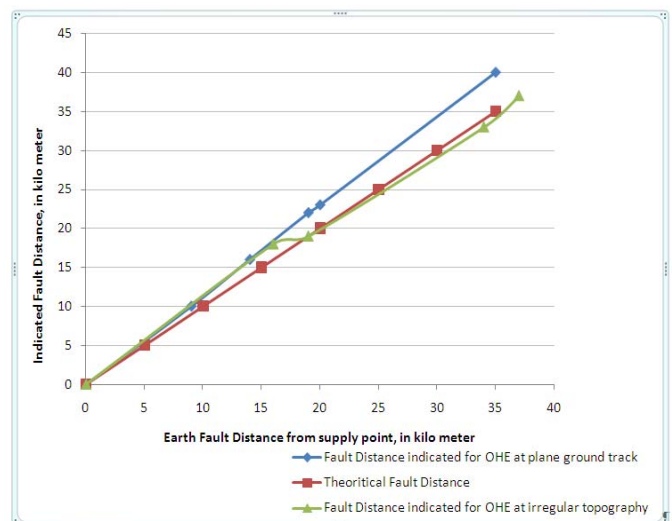


Fig.7. Plots on fault distances indicated by DPR for plane area, irregular terrain, and ideal situation

Result and Analysis

Earth fault distances indicated by DPR for the two distinct terrains and the theoretical curve are plotted as Fig. 7.

The graph plotted on the earth fault distance indicated by the DPR for the OHE at plane area (Fig.4) is more or less linear, and show a span error over and above the actual. The span error is attributed with the incorrect calculation of the shunt admittance present in between the entire length of OHE and the ground. The shunt admittance so calculated is found to be higher than the actual. The theoretical value of shunt admittance is derived from the average height of OHE from the ground, which is lesser than the actual height of OHE at that experimental area.

The graph plotted on the earth fault distance indicated by the DPR for the OHE at terrain where steep earth cuttings and tunnels are present (Fig.5 & 6) is almost linear with positive span error from the actual at its initial 15 KM long stretch due to the reason that topography at those stretch is generally plane ground with some minor gradual slope profiling of short lengths and a short length tunnel. A sudden shift towards the theoretical value is observed at 15 KM to 19 KM, due to the reason that many steep earth cuttings with high altitude & length, and two lengthy tunnels are available in this stretch. Besides, the height of the OHE is very close to its minimum permitted value due the constraints experienced in the tunnels and over head bridges to maintain more clearances. In the remaining stretches upto 37 KM, combinations of plane grounds, few minor slope profiles, large number of steep slope profiles, and one tunnel of length 201 M are available enroute (Table-5).

From the Tables 4 & 5, and from Fig. 7, it can be understood that the significant variations in the fault distances indicated by the DPR from the actual are attributed by the dynamic nature of the shunt admittance existing between the earth/earthed structure and the OHE. But, the prevailing algorithms [9][11][14] use the optimized value of shunt admittance for the fault distance calculation for all the terrains. Chances of significant variations in it due to the variations in topographical features, and variations in the clearance of OHE from ground are yet to be addressed rigorously.

From the above analysis, it is apparent that the slope profiles of earth cuttings, clearance of tunnels & over line structures in the proximity of OHE, variations in height of OHE etc. influence its shunt admittance significantly, and cause the variations in earth fault distances indicated by the DPR.

Conclusion

Each erratic data originated by the Distance Protection Relay on the distance of persisting earth fault on traction power systems cause serious delay in detecting them, which in turn adversely affect the electric traction systems' reliability to a large extent. Monetary loss due to prolonged disruption of train traffic shall also be too high. Hence, locating the exact position of earth fault on railway electric traction lines is a must, for which, foolproof algorithms are to be developed duly formulating the variations of parameters of traction lines contributed by the various topographical features. The relation of shunt admittance of OHE with the surrounding earth & earthed structures for various physical conditions like,

size, clearances from OHE, angle of slope, alignment of track etc. are to be formulated to modify the railway traction system more realistic.

Geographical maps of the experimental zone

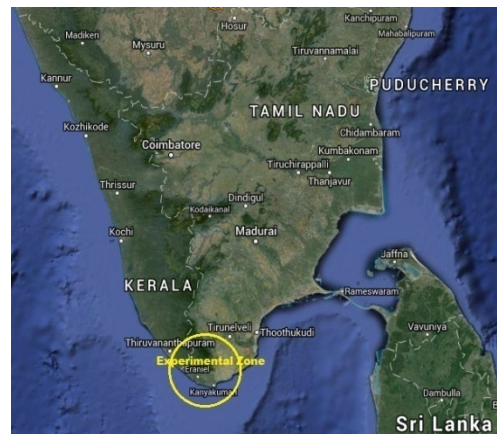


Fig. 8. Map of South India

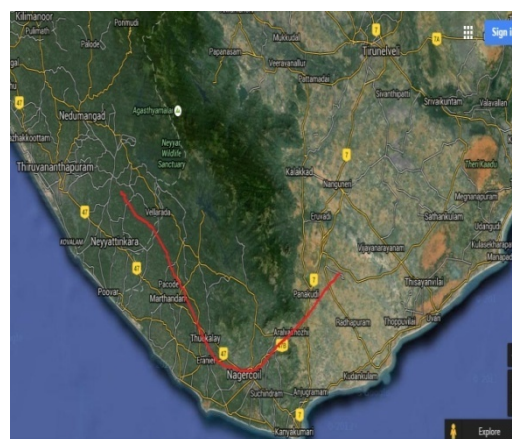


Fig. 9. Railway traction line indicated in red

References

1. M.V.Deshpande, "Electrical Power System Design", Tata McGraw Hill, New Delhi. 1984
2. Gonen. T, "Electric Power Transmission System Engineering, Analysis and Design", John Wiley & Sons, New York. 1988
3. "Modern Power Station Practice, EHV Transmission", Pergamon Press: Oxford, UK, Chap. 9. 1991 "AC Traction Manual for Indian Railways", vol.II, Part-II., pp 101, 1993
4. "Flexible AC Transmission Systems (FACTS)", IEE Power series, no. 30, UK, 1999
5. A.R.Bergan and V.Vittal, "Power System Analysis", Pearson Education Asia, New Delhi, 2001,
6. Glover. J.D., and Sarma. M.S, "Power System Analysis and Design", Thomson Asia Pvt. Ltd. Bangalore, 2003
7. Mariscotti.A. , Pozzobon.P, "Synthesis of line impedance expressions for Railway traction system",

IEEE Trnas. Vehicular Technol. Vol 52, pp 420 – 430, 2003

8. Cheng.W., Xu.G., and Mu.L, “A novel fault location algorithm for traction network based on distributed parameter line model”, Proceedings. CSU – EPSA, Vol 17, pp 66-63, 2005
9. William.H.Kerstring, “Distribution System Modeling and Analysis”, CRC Press, USA, 2006
10. Lin.X., Weng.H., and Wang.B.A., ,“A generalized method to improve the location accuracy of the single- ended sampled data and lumped parameter model based fault locators”, Electrical Power Energy System, vol 31, pp 201-205, 2009
11. Kothari.D.P., and Nagrath. I.J,“Power System Engineering”, Tata McGraw-Hill Publishing Company limited, New Delhi, India, Chap. 5, 2011
12. Singh.S.N,“Electric Power Generation, Transmission and Distribution”, PHI Learning Pvt. Ltd., New Delhi, India, Chap. 11, 2011
13. Yimin Zhou., Guoqing Xu., Yanfeng Chen, “Fault Location in Power Electrical Traction Line System”, Energies Journal, pp.5004, 5005, Nov.2012
14. “Log of Events on Earth faults”, Traction Power Control room, Indian Railways, Trivandrum Division, 2013-14 and 2014-2015

5-2016

Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations

Rajesh Singh

Banaras Hindu University, rsinghstat@gmail.com

Hemant K. Verma

Banaras Hindu University, coolhemant010189@gmail.com

Prayas Sharma

University of Petroleum and Energy Studies, prayassharma02@gmail.com

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

Recommended Citation

Singh, Rajesh; Verma, Hemant K.; and Sharma, Prayas (2016) "Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations," *Journal of Modern Applied Statistical Methods*: Vol. 15 : Iss. 1 , Article 19.

DOI: 10.22237/jmasm/1462076280

Available at: <http://digitalcommons.wayne.edu/jmasm/vol15/iss1/19>

This Regular Article is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized editor of DigitalCommons@WayneState.

Estimation of Population Mean Using Exponential Type Imputation Technique for Missing Observations

Rajesh Singh
Banaras Hindu University
Varanasi, India

Hemant K. Verma
Banaras Hindu University
Varanasi, India

Prayas Sharma
University of Petroleum and
Energy Studies
Dehradun, India

Some imputation techniques are suggested for estimating the population mean when the data values are missing completely at random under a simple random sample without replacement scheme. Two classes of point estimators are proposed. The bias and mean squared error expressions of the proposed point estimators are derived up to first order of approximation. It has been shown that the proposed point estimators are more efficient than some existing point estimators due to Lee, Rancourt, and Sarndal (1994) and Singh and Horn (2000). Theoretical findings are supported by an empirical study based on five populations to show the superiority of the constructed estimators and methods of imputation over others.

Keywords: Missing data, imputation, bias, mean squared error, simple random sampling without replacement

Introduction

Missing data is a common and serious problem in survey sampling. Missing data naturally occurs in sample surveys when a few sampling units refuse to respond or are unable to participate in the survey. There are two types of non-responses which occur in surveys: unit non-response and item non-response. Unit non-response occurs when an eligible sample unit fails to participate in a survey because of failure to establish a contact or explicit refusal to cooperate. Item non-response occurs instead when a responding unit does not provide useful answers to particular items

Dr. Singh is an Assistant Professor in the Department of Statistics. Email him at: rsinghstat@gmail.com. Dr. Sharma, the corresponding author, is a Professor in the Department of Decision Sciences, College of Management and Economics Studies. Email at prayassharma02@gmail.com.

of the questionnaire. Such situations create missing data problem. The imputation is a well-defined methodology by virtue of which such problems can be unraveled.

In the literature several imputation techniques are available and discussed. Rubin (1976) addressed three concepts: observed at random (OAR), missing at random (MAR), and parametric distribution (PD). Rubin defined MAR as the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data. Heitjan and Basu (1996) distinguished the meaning of MAR and missing completely at random (MCAR) in a very nice way. The imputation technique is also applicable when information on auxiliary variable is available. Lee et al. (1994; 1995) used the information on an auxiliary variable for the purpose of imputation, Singh and Horn (2000) suggested a compromised method of imputation, Ahmed, Al-Titi, Al-Rawi, and Abu-Dayyeh (2006) suggested several new imputation based estimators that use the information on an auxiliary variable and compared their performances with the mean method of imputation, and Rao and Sitter (1995) used the imputation techniques for variance estimation under two phase sampling. Kadilar and Cingi (2008) and Diana and Perri (2010) also suggested some imputation techniques in case of missing data. In the present study we implicitly assume MCAR.

Let

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N y_i$$

be the population mean of study variable Y . A simple random sample without replacement (SRSWOR), s , of size n is drawn from $\Omega = \{1, 2, \dots, N\}$ to estimate the population mean \bar{Y} . Let r be the number of responding units out of sampled n , then the number of non-responding units is $(n - r)$. Let the set of responding units be denoted by R and that of non-responding units be denoted by R^c . For every unit $i \in R$, the value y_i is observed. However for the units $i \in R^c$, the y_i values are missing and imputed values are to be derived. We assume that imputation is carried out with the aid of a quantitative auxiliary variable x such that, the value of x for unit i is x_i , known and positive for every $i \in s$. In other words, the data $x_s = \{x_i : i \in s\}$ are known.

Some Available Methods of Imputation and Estimators

There are some classical methods of imputation which are commonly used and given as follows:

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Mean Method of Imputation

In this method of imputation, the study variable y after imputation takes the form as

$$y_{.i} = \begin{cases} y_i, & i \in R \\ \bar{y}_r, & i \in R^c \end{cases} \quad (1)$$

and the point estimator of the population mean \bar{Y} is given by

$$\bar{y}_s = \frac{1}{n} \sum_{i \in s} y_{.i} . \quad (2)$$

Thus, under this method of imputation, the point estimator of the population mean \bar{Y} is

$$\bar{y}_m = \frac{1}{r} \sum_{i \in R} y_i = \bar{y}_r . \quad (3)$$

Lemma 1. The expression of Bias and Variance of the point estimator \bar{y}_m is given as

$$\text{Bias}(\bar{y}_m) = 0 \quad (4)$$

$$V(\bar{y}_m) = \left(\frac{1}{r} - \frac{1}{N} \right) S_y^2 \quad (5)$$

where $S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$.

Ratio Method of Imputation

Following the notations of Lee et al. (1994), in the case of single value imputation, if the i^{th} unit requires imputation, the value \hat{bx}_i is imputed. Thus, the study variable y after imputation takes the form as

$$y_{\cdot i} = \begin{cases} y_i, & i \in R \\ \hat{b}x_i, & i \in R^c \end{cases} \quad (6)$$

where

$$\hat{b} = \frac{\sum_{i \in R} y_i}{\sum_{i \in R} x_i}.$$

Under this method of imputation, the point estimator of the population mean \bar{Y} is given by

$$\bar{y}_{\text{RAT}} = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \quad (7)$$

where $\bar{x}_n = \frac{1}{n} \sum_{i \in S} x_i$, $\bar{x}_r = \frac{1}{r} \sum_{i \in R} x_i$, and $\bar{y}_r = \frac{1}{r} \sum_{i \in R} y_i$.

Lemma 2. The expression of Bias and Mean Square Error (MSE) of the point estimator \bar{y}_{RAT} is given as

$$\text{Bias}(\bar{y}_{\text{RAT}}) = \bar{Y} \left(\frac{1}{r} - \frac{1}{n} \right) (C_x^2 - \rho C_y C_x), \quad (8)$$

$$\text{MSE}(\bar{y}_{\text{RAT}}) = \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) (S_y^2 + R_1^2 S_x^2 - 2R_1 S_{xy}), \quad (9)$$

where S_y^2 is defined as above and

$$S_x^2 = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})^2, S_{xy} = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X})$$

$$R_1 = \frac{\bar{Y}}{\bar{X}}, C_y = \frac{S_y}{\bar{Y}}, C_x = \frac{S_x}{\bar{X}}, \rho = \frac{S_{xy}}{S_x S_y}$$

Compromised Method of Imputation

Singh and Horn (2000) proposed compromised imputation procedure. After imputation the study variable takes form as

$$y_{.i} = \begin{cases} \frac{\alpha n y_i}{r + (1-\alpha)\hat{b}x_i}, & i \in R \\ (1-\alpha)\hat{b}x_i, & i \in R^c \end{cases} \quad (10)$$

where α is a suitably chosen constant such that the variance of the resultant estimator is minimum. Here, we are also using information from imputed values for the responding units in addition to non-responding units.

Thus, under compromised method of imputation, the point estimator of the population mean \bar{Y} is

$$\bar{y}_{\text{COMP}} = \alpha \bar{y}_r + (1-\alpha) \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}. \quad (11)$$

Lemma 3. The expression of Bias and MSE of the point estimator \bar{y}_{COMP} is given as

$$\text{Bias}(\bar{y}_{\text{COMP}}) = \bar{Y} (1-\alpha) \left(\frac{1}{r} - \frac{1}{n} \right) (C_x^2 - \rho C_y C_x) \quad (12)$$

$$\begin{aligned} \text{MSE}(\bar{y}_{\text{COMP}}) = & \left(\frac{1}{n} - \frac{1}{N} \right) S_y^2 + \left(\frac{1}{r} - \frac{1}{n} \right) (S_y^2 + R_1^2 S_x^2 - 2R_1 S_{xy}) \\ & - \left(\frac{1}{r} - \frac{1}{n} \right) \bar{Y}^2 \alpha^2 C_x^2 \end{aligned} \quad (13)$$

where $\alpha_{\text{opt}} = 1 - \rho \frac{C_y}{C_x}$. Thus

$$\text{MSE}(\bar{y}_{\text{COMP}})_{\min} = \left[\left(\frac{1}{n} - \frac{1}{N} \right) - \left(\frac{1}{r} - \frac{1}{n} \right) \rho^2 \right] S_y^2 \quad (14)$$

Along similar lines, Ahmed et al. (2006) proposed several new imputation techniques by introducing some unknown parameters and hence proposed the corresponding estimators for estimating the finite population means \bar{Y} .

Proposed Imputation Methods and Corresponding Estimators

The following two imputation methods are suggested. After imputation for the first proposed imputation of technique, the study variable takes the form as

$$y_{.i} = \begin{cases} y_i, & i \in R \\ \frac{1}{n-r} \bar{y}_r \left[n \exp \left\{ \alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{x}_r^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_r^{\frac{1}{h}}} \right\} - r \right], & i \in R^c \end{cases} \quad (15)$$

where a , h , and α are suitably chosen constants. We optimize α in such a way that the MSE of the resultant estimator is minimum. Thus we have the following theorem:

Theorem 1. Under the proposed method of imputation considered in (15), the point estimator of the population mean \bar{Y} is given as

$$T_p = \bar{y}_r \exp \left\{ \alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{x}_r^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_r^{\frac{1}{h}}} \right\}. \quad (16)$$

Proof:

$$T_p = \frac{1}{n} \sum_{i \in s} y_{.i} = \frac{1}{n} \left[\sum_{i \in R} y_{.i} + \sum_{i \in R^c} y_{.i} \right] \quad (17)$$

where R and R^c are the sets of responding and non-responding units in the sample s of size n .

Now putting the values from (15) into (17), the point estimator of population is obtained as mean \bar{Y} as defined in (16), which completes the proof.

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Table 1. Members of the class of estimators T_P

Estimators		Constants	
$\alpha = 1$	$\alpha = 1$	a	h
$T_{P1} = \bar{y}_r \exp \left[\frac{\bar{X} - \bar{x}_r}{\bar{X}} \right]$	$T_{P5} = \bar{y}_r \exp \left[\frac{\bar{x}_r - \bar{X}}{\bar{X}} \right]$	1	1
$T_{P2} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}} - \sqrt{\bar{x}_r}}{\sqrt{\bar{X}}} \right]$	$T_{P6} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{x}_r} - \sqrt{\bar{X}}}{\sqrt{\bar{X}}} \right]$	1	2
$T_{P3} = \bar{y}_r \exp \left[\frac{\bar{X} - \bar{x}_r}{\bar{X} + \bar{x}_r} \right]$	$T_{P7} = \bar{y}_r \exp \left[\frac{\bar{x}_r - \bar{X}}{\bar{X} + \bar{x}_r} \right]$	2	1
$T_{P4} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}} - \sqrt{\bar{x}_r}}{\sqrt{\bar{X}} + \sqrt{\bar{x}_r}} \right]$	$T_{P8} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{x}_r} - \sqrt{\bar{X}}}{\sqrt{\bar{X}} + \sqrt{\bar{x}_r}} \right]$	2	2

Because the point estimator proposed in (16) after imputing the missing values, belongs to a class of estimators. Some members of the proposed class of point estimator defined in (16) are shown in Table 1 for different choice of a , h , and α .

The study variable after imputation for the second proposed imputation of technique becomes

$$y_{.i} = \begin{cases} y_i, & i \in R \\ \frac{1}{n-r} \bar{y}_r \left[n \exp \left\{ \alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{x}_n^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1)\bar{x}_n^{\frac{1}{h}}} \right\} - r \right], & i \in R^c \end{cases} \quad (18)$$

where a , h , and α are suitably chosen constants. We optimize α in such a way that the MSE of the resultant estimator is minimum. Thus we have the following theorem:

Theorem 2. Under the proposed method of imputation considered in (18), the point estimator of the population mean \bar{Y} is given as

$$\bullet T_g = \bar{y}_r \exp \left\{ \alpha \frac{\bar{X}^{\frac{1}{h}} - \bar{X}_n^{\frac{1}{h}}}{\bar{X}^{\frac{1}{h}} + (a-1) \bar{X}_n^{\frac{1}{h}}} \right\} \quad (19)$$

Proof:

$$T_g = \frac{1}{n} \sum_{i \in S} y_{\cdot i} = \frac{1}{n} \left[\sum_{i \in R} y_{\cdot i} + \sum_{i \in R^c} y_{\cdot i} \right] \quad (20)$$

where R and R^c are the sets of responding and non-responding units in the sample, *s*, of size *n*.

Putting the values from (18) into (20), we get the form of the point estimator of population mean \bar{Y} as defined in (19), which completes the proof.

Some members of the proposed class of point estimator defined in (19) are shown in Table 2 for different choices of *a*, *h*, and α .

Properties of the Estimators T_P and T_g

To obtain the bias and MSE expressions of the estimators to the first degree of approximation, we define

$$e_0 = \frac{\bar{y}_r - \bar{Y}}{\bar{Y}}, e_1 = \frac{\bar{x}_r - \bar{X}}{\bar{X}}, e_2 = \frac{\bar{x}_n - \bar{X}}{\bar{X}}$$

such that $E(e_i) = 0$; $i = 0, 1, 2$, and

$$\begin{aligned} E(e_0^2) &= \left(\frac{1}{r} - \frac{1}{N} \right) C_y^2, E(e_1^2) = \left(\frac{1}{r} - \frac{1}{N} \right) C_x^2, E(e_0 e_1) = \left(\frac{1}{r} - \frac{1}{N} \right) \rho C_y C_x, \\ E(e_2^2) &= \left(\frac{1}{n} - \frac{1}{N} \right) C_x^2, E(e_1 e_2) = \left(\frac{1}{n} - \frac{1}{N} \right) C_x^2, E(e_0 e_2) = \left(\frac{1}{n} - \frac{1}{N} \right) \rho C_y C_x \end{aligned}$$

Using above terminology, the bias and MSE of the proposed estimators are given below.

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Table 2. Members of the class of estimators T_g

Estimators		Constants	
$\alpha = 1$	$\alpha = 1$	a	h
$T_{g1} = \bar{y}_r \exp \left[\frac{\bar{X} - \bar{X}_n}{\bar{X}} \right]$	$T_{g5} = \bar{y}_r \exp \left[\frac{\bar{X}_n - \bar{X}}{\bar{X}} \right]$	1	1
$T_{g2} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}} - \sqrt{\bar{X}_n}}{\sqrt{\bar{X}}} \right]$	$T_{g6} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}_n} - \sqrt{\bar{X}}}{\sqrt{\bar{X}}} \right]$	1	2
$T_{g3} = \bar{y}_r \exp \left[\frac{\bar{X} - \bar{X}_n}{\bar{X} + \bar{X}_n} \right]$	$T_{g7} = \bar{y}_r \exp \left[\frac{\bar{X}_n - \bar{X}}{\bar{X} + \bar{X}_n} \right]$	2	1
$T_{g4} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}} - \sqrt{\bar{X}_n}}{\sqrt{\bar{X}} + \sqrt{\bar{X}_n}} \right]$	$T_{g8} = \bar{y}_r \exp \left[\frac{\sqrt{\bar{X}_n} - \sqrt{\bar{X}}}{\sqrt{\bar{X}} + \sqrt{\bar{X}_n}} \right]$	2	2

Theorem 3. The Bias of the estimator T_P is given by

$$\text{Bias}(T_P) = \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y} \frac{\alpha}{ah} \left[\frac{1}{ah} \left(\frac{\alpha}{2} + a - 1 \right) C_x^2 - \rho C_y C_x \right] \quad (21)$$

and the MSE of the estimator T_P is given by

$$\text{MSE}(T_P) = \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 \left[C_y^2 + \frac{\alpha^2}{a^2 h^2} C_x^2 - 2 \frac{\alpha}{ah} \rho C_y C_x \right], \quad (22)$$

where the optimum value of α is given by

$$\alpha_{\text{opt}} = ah\rho \frac{C_y}{C_x}$$

Proof: Expressing the estimator T_P in terms of the e 's, we have

$$T_P = \bar{Y} (1 + e_0) \exp \left[\alpha \left\{ \frac{1 - (1 + e_1)^{\frac{1}{h}}}{1 + (a-1)(1 + e_1)^{\frac{1}{h}}} \right\} \right]$$

$$\begin{aligned}
&\approx \bar{Y}(1+e_0) \exp \left[\frac{-\alpha e_1}{ah} \left(1 + \frac{e_1}{h} - \frac{e_1}{ah} \right)^{-1} \right] \\
&\approx \bar{Y}(1+e_0) \exp \left(\frac{-\alpha e_1}{ah} \right) \exp \left(\frac{\alpha(a-1)}{a^2 h^2} e_1^2 \right) \\
&\approx \bar{Y}(1+e_0) \left[1 - \frac{\alpha e_1}{ah} + \frac{\alpha(a-1)}{a^2 h^2} e_1^2 + \frac{1}{2} \frac{\alpha}{a^2 h^2} e_1^2 \right] \\
(T_P - \bar{Y}) &\approx \left[e_0 - \frac{\alpha}{ah} e_1 - \frac{\alpha(a-1)}{a^2 h^2} e_1^2 + \frac{1}{2} \frac{\alpha}{a^2 h^2} e_1^2 \right]
\end{aligned}$$

Taking expectation on both sides, we get the bias expression of estimator T_P as

$$\text{Bias}(T_P) = \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y} \frac{\alpha}{ah} \left[\frac{1}{ah} \left(\frac{\alpha}{2} + a - 1 \right) C_x^2 - \rho C_y C_x \right]$$

To find the MSE of the estimator T_P , we have

$$\begin{aligned}
\text{MSE}(T_P) &= E(T_P - \bar{Y})^2 \\
&\approx \bar{Y}^2 E \left(e_0 - \frac{\alpha}{ah} e_1 \right)^2 \\
&\approx \bar{Y}^2 \left(E(e_0^2) + \frac{\alpha^2}{a^2 h^2} E(e_1^2) - \frac{\alpha}{ah} E(e_0 e_1) \right) \\
\text{MSE}(T_P) &= \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 \left[C_y^2 + \frac{\alpha^2}{a^2 h^2} C_x^2 - 2 \frac{\alpha}{ah} \rho C_y C_x \right]
\end{aligned}$$

Partially differentiating above equation with respect to α and equating to zero, we have

$$\frac{\partial(\text{MSE}(T_P))}{\partial \alpha} = \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 \left[2 \frac{\alpha}{a^2 h^2} C_x^2 - 2 \frac{1}{ah} \rho C_y C_x \right] = 0$$

Simplifying the above equation, we get the optimum value of α as

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

$$\alpha_{\text{opt}} = ah\rho \frac{C_y}{C_x}$$

Theorem 4. The Bias of the estimator T_g is given by

$$\text{Bias}(T_g) = \left(\frac{1}{n} - \frac{1}{N} \right) \bar{Y} \frac{\alpha}{ah} \left[\frac{1}{ah} \left(\frac{\alpha}{2} + a - 1 \right) C_x^2 - \rho C_y C_x \right] \quad (23)$$

and the MSE of the estimator T_g is given by

$$\text{MSE}(T_g) = \bar{Y}^2 \left[\left(\frac{1}{r} - \frac{1}{N} \right) C_y^2 + \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{\alpha^2}{a^2 h^2} C_x^2 - 2 \frac{\alpha}{ah} \rho C_y C_x \right) \right], \quad (24)$$

where the optimum value of α is given by

$$\alpha_{\text{opt}} = ah\rho \frac{C_y}{C_x}$$

Proof: The above theorem can be proved in a similar way to the proof of Theorem 3.

Efficiency Comparison

Estimator T_P is more efficient than estimator \bar{y}_m if

$$V(\bar{y}_m) - \text{MSE}[T_P(\min)] > 0.$$

But

$$\begin{aligned} \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 C_y^2 - \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 (1 - \rho^2) C_y^2 &> 0 \\ \left(\frac{1}{r} - \frac{1}{N} \right) \bar{Y}^2 \rho^2 C_y^2 &> 0 \end{aligned}$$

since $\frac{1}{r} > \frac{1}{N}$. Therefore, T_P is more efficient than \bar{y}_m . Similarly,

$$\text{MSE}(\bar{y}_{\text{RAT}}) - \text{MSE}[T_P(\min)] > 0 \text{ if } \left(\frac{1}{r} - \frac{1}{n}\right)(C_x^2 - 2\rho C_y C_x) + \left(\frac{1}{r} - \frac{1}{N}\right)\rho^2 C_y^2 > 0$$

$$\text{MSE}(\bar{y}_{\text{COMP}}) - \text{MSE}[T_P(\min)] > 0 \text{ if } \left(\frac{1}{r} - \frac{1}{n}\right) < \left(\frac{1}{r} - \frac{1}{N}\right)\rho^2$$

$$\text{MSE}[T_g(\min)] - \text{MSE}[T_P(\min)] > 0 \text{ if } \frac{1}{r} > \frac{1}{n}$$

Thus, from the above results, we can say that the estimator T_P is more efficient than other estimators.

Empirical Study

Five populations, A, B, C, D, and E, are considered. Population A is the artificial population of size $N = 200$ from Shukla, Thakur, Pathak, and Rajput (2009), population B is from Ahmed et al. (2006), population C is from Dass (1988), population D is from Murthy (1967, p. 228), and population E is from Singh, Singh, and Kumar (1976, p. 126) with parameters as given in Table 3.

Let $n = 40$, $r = 35$ for population A, $n = 200$, $r = 180$ for population B, $n = 80$, $r = 72$ for population C, $n = 23$, $r = 20$ for population D, and $n = 6$, $r = 5$ for population E respectively. Then the bias and MSE of the proposed point estimators are given in Table 4 and Table 5 for populations A, B, C, D and E respectively.

Table 3. Parameters for study populations

Population	Parameters							
	N	\bar{Y}	\bar{X}	S_y^2	S_x^2	ρ	C_y	C_x
A	200	42.485	18.515	199.0598	48.5375	0.865200	0.37630	0.33210
B	8306	253.750	343.316	338006.0000	862017.0000	0.522231	2.70436	2.29116
C	278	39.070	25.110	3199.2400	1660.0200	0.720000	1.44770	1.62260
D	80	5182.640	285.130	3370161.0000	73129.9400	0.920000	0.35420	0.94840
E	17	33.290	40.060	287.8600	458.3500	0.720000	0.50970	0.54990

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Table 4. Biases of estimators

Estimators	Populations				
	A	B	C	D	E
\bar{y}_m	0.0000	0.0000	0.0000	0.0000	0.0000
\bar{y}_{RAT}	0.0051	0.5749	0.0511	19.9568	0.1117
\bar{y}_{COMP}	0.0039	0.2543	0.0328	6.8572	0.0745
$T_P(\text{min})$	-0.0413	-0.9872	-0.8184	-10.3199	-0.3164
$T_g(\text{min})$	-0.0351	-0.8863	-0.1890	-8.5252	-0.2417

Table 5. MSEs of estimators

Estimators	Populations				
	A	B	C	D	E
\bar{y}_m	4.6921	1837.1169	32.9258	126381.0375	40.6391
\bar{y}_{RAT}	4.2110	1867.2341	31.3361	175668.0261	36.9018
\bar{y}_{COMP}	4.1599	1785.9043	30.6224	107777.7488	35.6648
$T_P(\text{min})$	1.1798	1336.0843	15.8571	19412.1274	19.5718
$T_g(\text{min})$	2.8938	1387.2968	18.9807	38015.4161	24.5460

Tables 4 and 5 exhibits the bias and MSE of different point estimators and it has been observed from the tables that the estimators based on auxiliary information are more efficient than the one which does not use the auxiliary information such as \bar{y}_m to overcome the imputation problems. Both the proposed classes of estimators T_P and T_g are more efficient than the estimators, \bar{y}_m , \bar{y}_{RAT} and \bar{y}_{COMP} , scrupulously, T_P has minimum MSE among all the estimators considered here.

Conclusion

Two imputation techniques are suggested using auxiliary information followed by two class of estimators for estimating the population mean in case of data values are MCAR under a SRSWR scheme. In addition, some new members are also generated from two proposed class of estimators using the suitable values of constants. The minimum biases and mean square errors of the proposed class of estimators were determined up to the first order of approximation. It was established theoretically and empirically that the proposed class of estimator performs best among the other estimators considered, and consequently the

corresponding (first proposed) method of imputation is better than the other existing methods and may be recommended for further use.

References

- Ahmed, M. S., Al-Titi, O., Al-Rawi, Z., & Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition*, 7(6), 1247-1264.
- Dass, A. K. (1988). *Contributions to the theory of sampling strategies based on auxiliary information* (Unpublished doctoral dissertation). Bidhan Chandra Agricultural University, West Bengal, India.
- Diana, G., & Perri, P. F. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics – Theory and Methods*, 39(18), 3245-3251. doi: 10.1080/03610920903009400
- Heitjan, D. F., & Basu, S. (1996). Distinguishing “missing at random” and “missing completely at random”. *The American Statistician*, 50(3), 207-213. doi: 10.1080/00031305.1996.10474381
- Kadilar, C., & Cingi, H. (2008). Estimators for the population mean in the case of missing data. *Communications in Statistics – Theory and Methods*, 37(14), 2226-2236. doi: 10.1080/03610920701855020
- Lee, H., Rancourt, E., & Sarndal, C. E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, 10(3), 231-243. Retrieved from: <http://www.jos.nu/Articles/abstract.asp?article=103231>
- Lee, H., Rancourt, E., & Sarndal, C. E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. In *Proceedings of the Survey Research Methods Section*, American Statistical Association, 384-389. Retrieved from: <http://www.amstat.org/sections/srms/Proceedings/>
- Murthy, M. N. (1967). *Sampling Theory and Methods*. Calcutta, India: Statistical Publishing Society.
- Rao, J. N. K., & Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data. *Biometrika*, 82(2), p. 453-460. doi: 10.1093/biomet/82.2.453
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581-593. doi: 10.1093/biomet/63.3.581

EXPONENTIAL TYPE IMPUTATION TECHNIQUE

Shukla, D., Thakur, N. S., Pathak, S., & Rajput, D. S. (2009). Estimation of mean under imputation of missing data using factor-type estimator in two-phase sampling, *Statistics in Transition*, 10(3), 397-414.

Singh, D., Singh, P., & Kumar, P. (1976). Handbook on Sampling Methods. New Delhi, India: Indian Agricultural Statistics Research Institute (ICAR).

Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. *Metrika*, 51(3), 266-276. doi: [10.1007/s001840000054](https://doi.org/10.1007/s001840000054)

FERROFLUID FLOW DUE TO A ROTATING DISK IN THE PRESENCE OF A NON-UNIFORM MAGNETIC FIELD

A. BHANDARI* and V. KUMAR

Department of Mathematics, College of Engineering Studies
University of Petroleum and Energy Studies
Dehradun, INDIA

E-mails: pankaj.anupam6@gmail.com; deshwal.vipin@gmail.com

The flow of a ferrofluid due to a rotating disk in the presence of a non-uniform magnetic field in the axial direction is studied through mathematical modeling of the problem. Contour and surface plots in the presence of 10 kilo-ampere/meter, 100 kilo-ampere/meter magnetization force are presented here for radial, tangential and axial velocity profiles, and results are also drawn for the magnetic field intensity. These results are compared with the ordinary case where magnetization force is absent.

Key words: magnetization force, ferrofluid, rotating disk, magnetic field.

1. Introduction

Ferrofluids are stable suspensions of colloidal ferromagnetic particles of the order of 10nm in suitable non-magnetic carrier liquids. These colloidal particles are coated with surfactants to avoid their agglomeration. Because of the industrial applications of ferrofluids, the investigation on them has fascinated the researchers and engineers vigorously for the last five decades. One of the many fascinating features of ferrofluids is the prospect of influencing the flow by a magnetic field [1, 2]. Ferrofluids are widely used in sealing of hard disc drives, rotating x-ray tubes under engineering applications. Sealing of the rotating shafts is the most known application of the magnetic fluids. The major application of ferrofluids in the electrical field is that of controlling heat in loudspeakers which makes their life longer and increases the acoustical power without any change in the geometrical shape of the speaker system. Magnetic fluids are used in the contrast medium in X-ray examinations and positioning tamponade for retinal detachment repair in eye surgery. Therefore, ferrofluids play an important role in bio-medical applications also. In the presence of a uniform magnetic field, the magnetization characteristics depend on the particle spin but do not depend on the fluid velocity. Convection of ferromagnetic fluids is gaining much importance due to their astounding physical properties.

There are rotationally symmetric flows of incompressible ferrofluids in the field of fluid mechanics, having all three velocity components; radial, tangential and vertical in space different from zero. In such types of flow, the variables are independent of the angular coordinates. Detailed accounts of magneto viscous effects in ferrofluids have been given in the monograph by Odenbach [3]. In the flow of an incompressible ferrofluid, the plate is subjected to the magnetic field $[H_r, 0, H_z]$ using, Neuringer-Rosensweig model [4]. This model was used by Verma *et al.* [5, 6, 7] for solving the paramagnetic Couette flow, helical flow with heat conduction and flow through a porous annulus. Rosensweig [8] gave an authoritative introduction to the research on magnetic liquids in his monograph and studied the effect of magnetization.

A study of flow within the boundary layer and its effect on the general flow around the body, is given in Schlichting [9]. Karman's [10] rotating disc problem is extended to the case of flow started

* To whom correspondence should be addressed

impulsively from rest and also the steady state is solved to a higher degree of accuracy than previously done by a simple analytical method which neglects the resembling difficulties in Cochran's [11] well known solution. The pioneering study of an ordinary viscous fluid flow due to an infinite rotating disc was carried by Von Karman. He introduced the famous transformation which reduces the governing partial differential equations into ordinary differential equations. Cochran obtained asymptotic solutions for the steady hydrodynamic problem formulated by Von Karman. Benton [12] improved Cochran's solutions, and solved the unsteady case. Attia [13] studied the unsteady state in the presence of an applied uniform magnetic field. The steady flow of an ordinary viscous fluid due to a rotating disc with uniform high suction was studied by Mithal [14]. Attia [15] discussed the flow due to an infinite disk rotating in the presence of an axial uniform magnetic field by taking Hall effect into consideration.

Using linear instability analysis, Venkatasubramanian and Kaloni [16] discussed the effects of rotation on the onset of convection in a horizontal layer of ferrofluids rotating about the vertical axis, heated from below and in the presence of a uniform vertical magnetic field. The effect of an alternating uniform magnetic field on convection in a horizontal layer of a ferrofluid within the framework of a quasi-stationary approach is studied by Belyaev [17].

The effect of the magnetic field along the vertical axis on thermo-convective instability in a ferromagnetic fluid saturating a rotating porous medium was studied by Sekar *et al.* [18] by using the Darcy model. Attia [19] studied the steady flow of an incompressible viscous fluid above an infinite rotating disk in a porous medium with heat transfer and also discussed the effect of porosity of the medium on the velocity and temperature distribution. Frusteri and Osalusi [20] examined the laminar convective and slip flow of an electrically conducting Newtonian fluid with variable properties over a rotating porous disk.

In general, magnetization is a function of the magnetic field, temperature and density of the fluid. This leads to convection of the ferrofluid in the presence of the magnetic field gradient. Sunil *et al.* [21] studied the effect of magnetic field-dependent viscosity on thermosolutal convection in a ferromagnetic fluid saturating a porous medium. Sunil *et al.* [22] discussed the influence of rotation on medium permeability and how MFD viscosity affects the magnetization in ferromagnetic fluid heated from below in the presence of dust particles saturating a porous medium of very low permeability using the Darcy model. The effect of MFD viscosity on thermal convection in a ferromagnetic fluid in a porous medium is studied by Sunil *et al.* [23]. Nanjundappa *et al.* [24] studied Benard-Marangoni ferroconvection in a ferrofluid layer in the presence of a uniform vertical magnetic field with magnetic field dependent (MFD) viscosity. Ram *et al.* [25] solved the non-linear differential equations under Neuringer-Rosensweig model for the ferrofluid flow by using power series approximations and discussed the effect of magnetic field-dependent viscosity on the velocity components and pressure profile. The effect of negative viscosity on the ferrofluid flow due to a rotating disk under alternating magnetic field is studied by Ram *et al.* [26-27].

In the present case, we take cylindrical coordinates r, θ, z , where the z -axis is normal to the plane and this axis is considered as the axis of rotation. The radius of the disk is taken 0.5 meter and the disk is rotating with 4 radian / second. We have studied the effects of magnetization force on the ferrofluid flow due to a rotating disk for different values of the magnetic field intensity of 10 and 100 kilo ampere per meter which is applied in the axial direction.

2. Mathematical formulation of the problem

The constitutive set of equations is as follows:

The equation of continuity

$$\nabla \cdot \mathbf{v} = 0. \quad (2.1)$$

The equation of motion

$$\rho \left[\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} \right] = -\nabla p + \mu_0 (\mathbf{M} \cdot \nabla) \mathbf{H} + \mu \nabla^2 \mathbf{v}. \quad (2.2)$$

Maxwell's equations

$$\nabla \times \mathbf{H} = \mathbf{0}, \quad \nabla \cdot (\mathbf{H} + \mathbf{M}) = 0; \quad \text{with} \quad \mathbf{M} = \chi \mathbf{H}. \quad (2.3)$$

Here ρ is the ferrofluid density, $\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla$, \mathbf{v} is the fluid velocity, p is the pressure, μ is the reference viscosity, μ_0 is magnetic permeability of free space, \mathbf{M} is the magnetization, \mathbf{H} is the magnetic field intensity, χ is the magnetic susceptibility, t is the time.

Equations (2.1) and (2.2) can be written in the cylindrical form

$$\frac{\partial v_r}{\partial r} + \frac{v_r}{r} + \frac{\partial v_z}{\partial z} = 0, \quad (2.4)$$

$$-\frac{1}{\rho} \frac{\partial p}{\partial r} + v \left[\frac{\partial^2 v_r}{\partial r^2} + \frac{\partial}{\partial r} \left(\frac{v_r}{r} \right) + \frac{\partial^2 v_r}{\partial z^2} \right] = \left[v_r \frac{\partial v_r}{\partial r} + v_z \frac{\partial v_r}{\partial z} - \frac{v_\theta^2}{r} \right], \quad (2.5)$$

$$v \left[\frac{\partial^2 v_\theta}{\partial r^2} + \frac{\partial}{\partial r} \left(\frac{v_\theta}{r} \right) + \frac{\partial^2 v_\theta}{\partial z^2} \right] = \left[v_r \frac{\partial v_\theta}{\partial r} + v_z \frac{\partial v_\theta}{\partial z} + \frac{v_r v_\theta}{r} \right], \quad (2.6)$$

$$-\frac{1}{\rho} \frac{\partial p}{\partial z} + \frac{\mu}{\rho} |\mathbf{M}| \frac{\partial}{\partial z} |\mathbf{H}| + v \left[\frac{\partial^2 v_z}{\partial r^2} + \frac{1}{r} \frac{\partial v_z}{\partial r} + \frac{\partial^2 v_z}{\partial z^2} \right] = \left[v_r \frac{\partial v_z}{\partial r} + v_z \frac{\partial v_z}{\partial z} \right] \quad (2.7)$$

where v_r , v_θ and v_z are velocity components in the radial, tangential and axial direction, respectively. Here, $v = \frac{\mu}{\rho}$ is the kinematic variable viscosity.

Here the disk is rotating with uniform angular velocity ω . The boundary conditions for the flow of the ferrofluid due to a rotating disk in the presence of a non-uniform magnetic field used by Schlichting [9] are given as follows

$$\begin{aligned} \text{at } z=0; \quad v_r &= 0, \quad v_\theta = r\omega, \quad v_z = 0, \\ \text{at } z \rightarrow \infty; \quad v_r &\rightarrow 0, \quad v_\theta \rightarrow 0. \end{aligned} \quad (2.8)$$

Here, v_z does not vanish at $z \rightarrow \infty$, but tends to a finite value.

The solution of Eqs (2.5)-(2.7) with the help of (2.8) is obtained using the Finite Element Method. All the units are taken in MKS system and the initial magnetic field is taken 10 kilo-ampere/meter and 100 kilo-ampere/meter.

3. Results and discussions

Figures 1-3 represent the radial velocity profile for different values of the magnetic field intensity. Figure 1 is plotted when the externally magnetic field is not applied. However, in the presence of the magnetic field, the radial velocity increases and depends on the strength of the field. In Fig.2, 10 kilo/ampere magnetic field is applied and in Fig.3, 100 kilo/ampere magnetic field is applied. Due to strong magnetic, polarization, the radial velocity depends on the strength of the magnetic field and the effect of the rotation of the disk diminishes. As the magnetic field increases in the axial directions, the fluid is being polarized in the axial directions and depends on the intensity of the magnetic field.

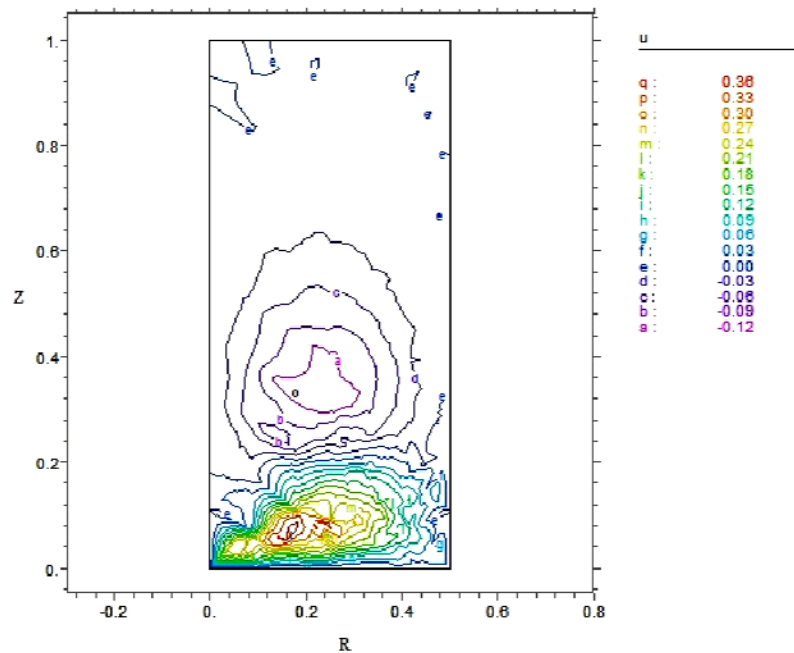


Fig.1. Radial velocity profile for $H=0$.

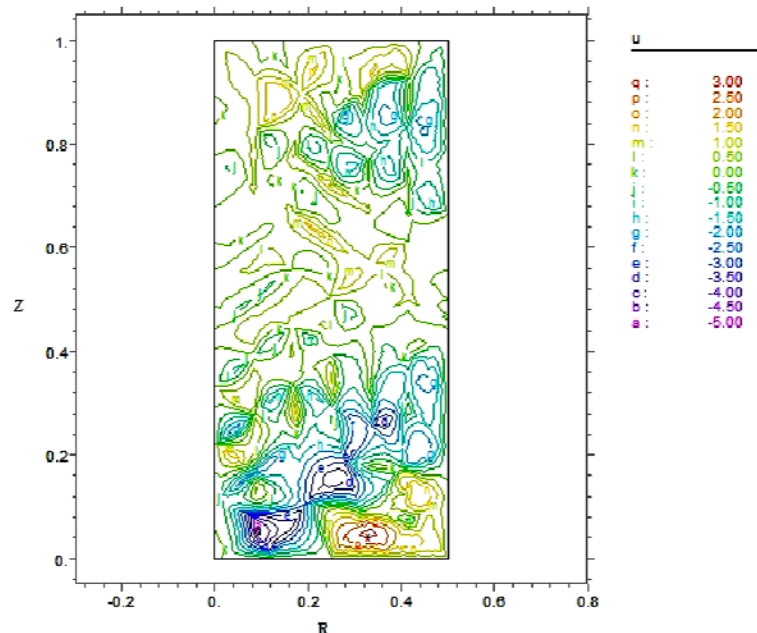


Fig.2. Radial velocity profile for $H=10$ kilo/ampere.

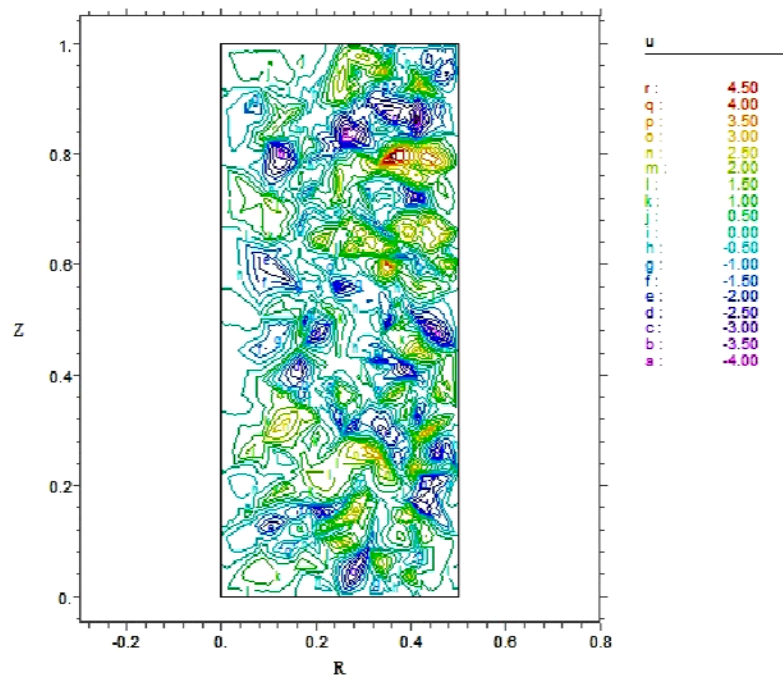


Fig.3. Radial velocity profile for 100 kilo/ampere.

Figures 4-6 represent the tangential velocity profile for different values of the magnetization force. It is clear from the figure that near the disk tangential velocity increases in Fig.4. However, in Figs 5 and 6, the tangential velocity increases along the z direction since the field is being applied in z direction. In Fig.4, the tangential velocity increases along the z -axis, however, as far as for large distance along z -axis, the tangential velocity reaches the steady state region. In Fig.5, due to the applied magnetic field in the radial direction, the velocity components get large value far from the disk. And, magnetization force is more dominant in Fig.6 in comparison with Fig.5 since the disk is being fixed in 100 kilo/ampere magnetization force.

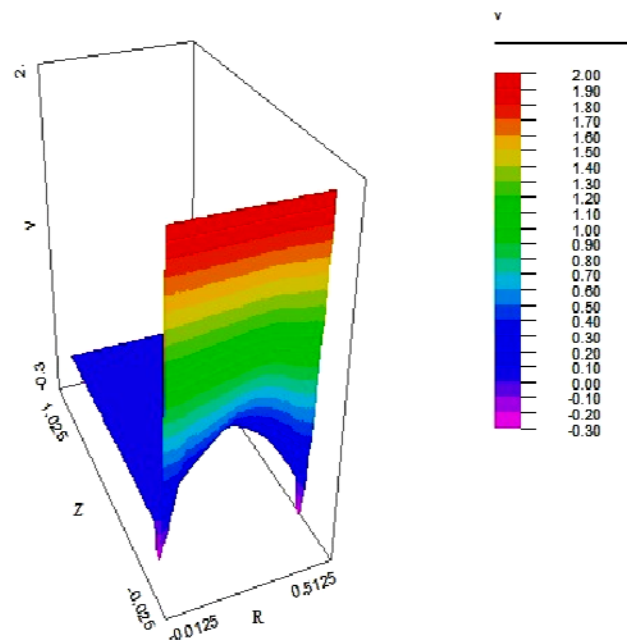


Fig.4. Tangential velocity profile for $H=0$.

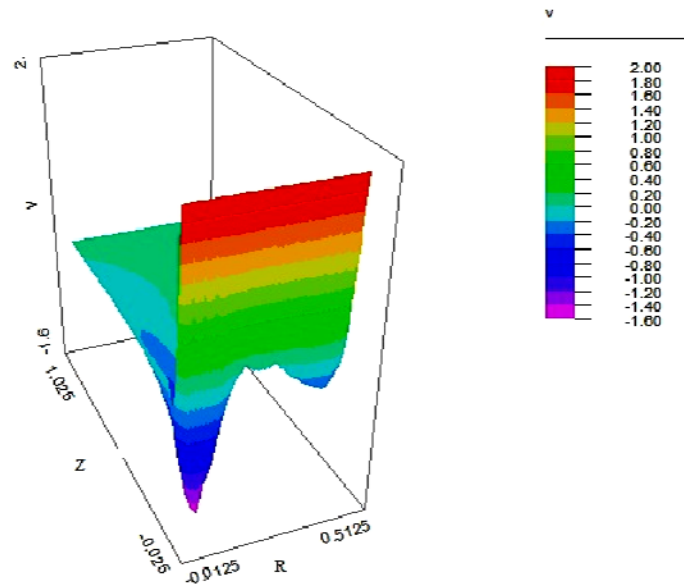


Fig.5. Tangential velocity profile for $H=10$ kilo/ampere.

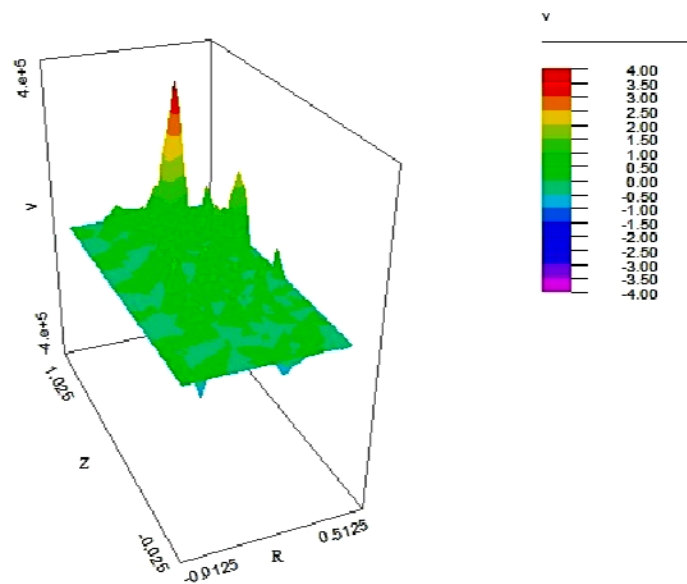


Fig.6. Tangential velocity profile for $H=100$ kilo/ampere.

Figures 6-9 depicts the axial velocity profile. The negative values of the axial velocity indicate that the flow of the ferrofluid is towards the disk. Figure 7 shows the vortex flow about the z axis and the particles follow the pattern in the flow, however, in case of 10 kilo/ampere magnetization force the path of the particles is disturbed as shown in Fig.8. However, in Fig.9, it is more disturbed in comparison with Fig.8 and the overall axial velocity is being controlled by the magnetization force as per the properties of the ferrofluid. As we increase the magnetic field intensity, it is observed from the results that the fluid velocity depends on the intensity of the field.

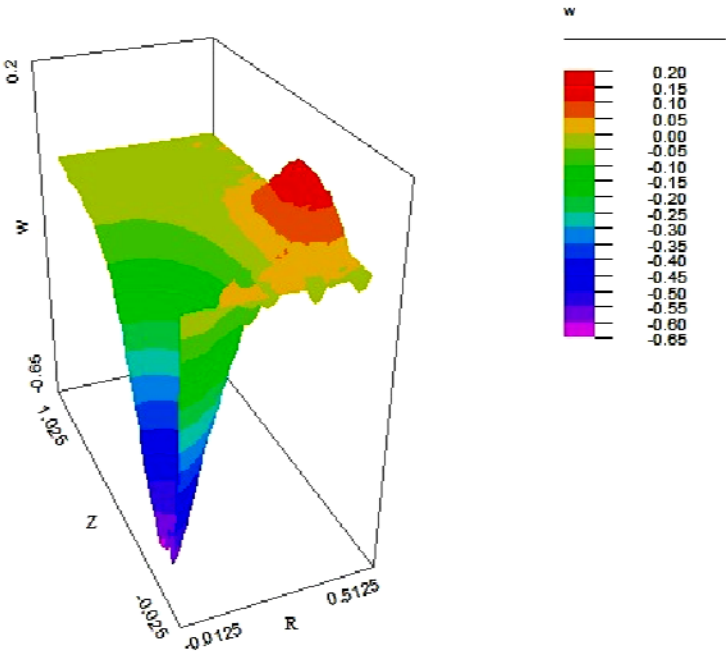


Fig.7. Axial velocity profile for $H=0$.

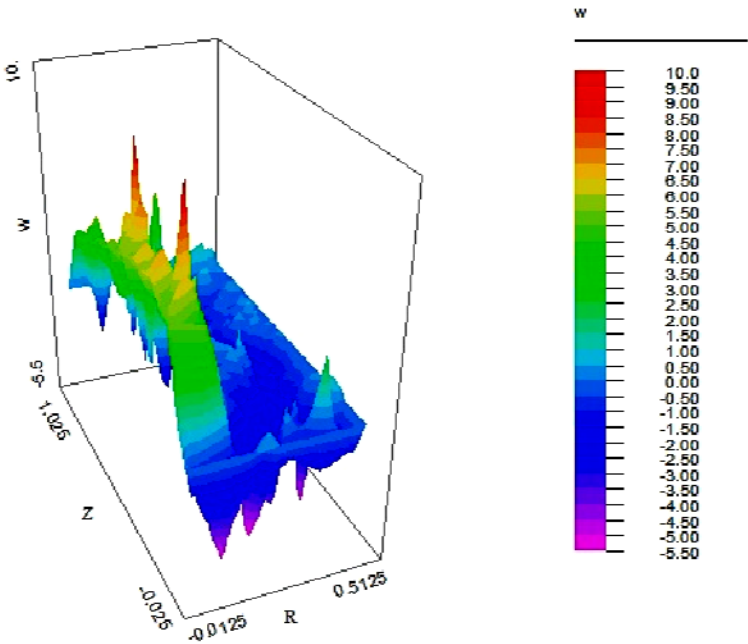


Fig.8. Axial velocity profile for $H=10$ kilo/ampere.

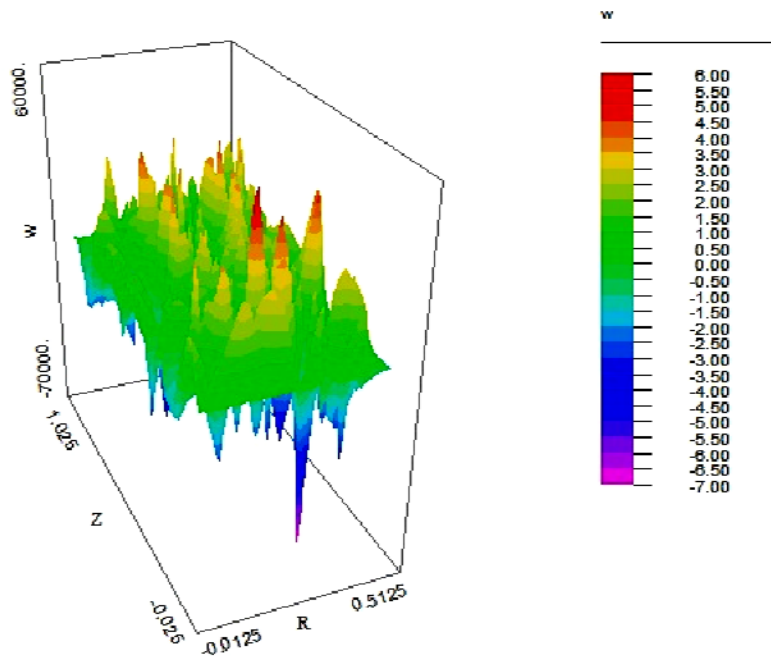


Fig.9. Axial velocity profile for $H=100$ kilo/ampere.

Figures 10-11 illustrate the behavior of the magnetic field intensity. Here the disk is being kept at 10 and 100 kilo /ampere magnetization force and that force is being applied in the axial direction. As the fluid moves along the z -axis the magnetization force reduces there due to the no-uniform distribution of the magnetization force.

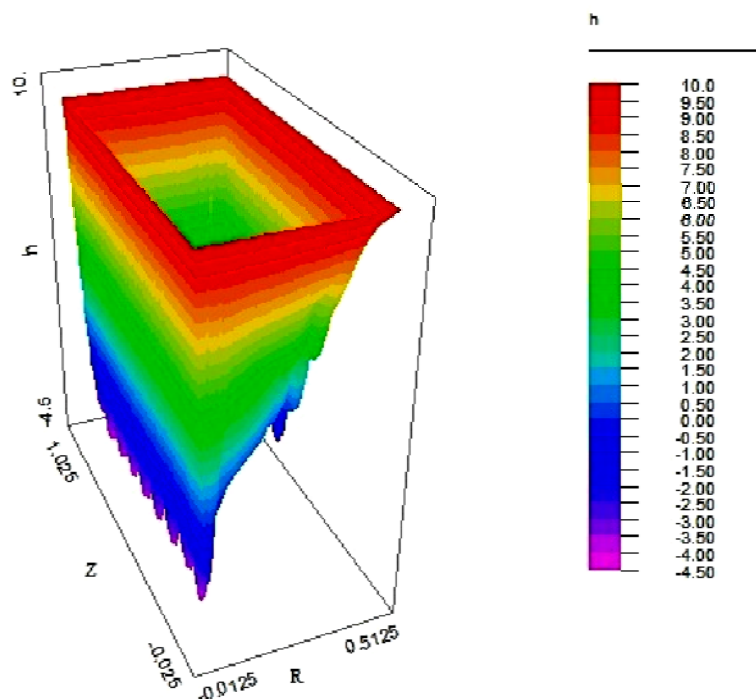


Fig.10. Variation of the magnetic field in the axial direction at $H=10$ kilo/ampere.

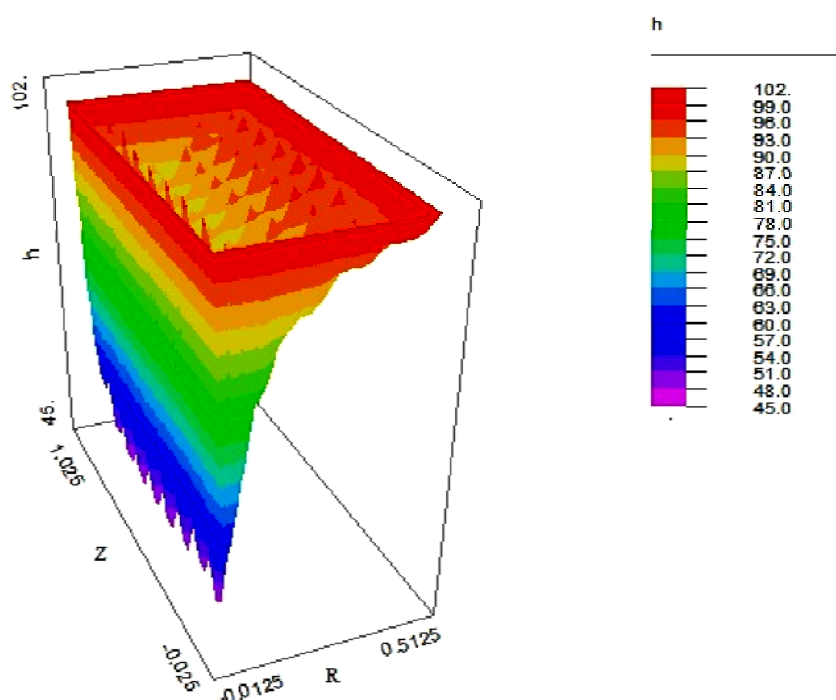


Fig.11. Variation of the magnetic field in the axial direction at $H=100$ kilo/ampere.

4. Conclusions

Present results show the effect of the magnetization force on the axisymmetric ferrofluid flow due to a rotating disk. As we increase the intensity of the magnetic field in the direction normal to the disk, the velocity of the magnetic fluid depends much on its intensity. These results indicate also the space application of this fluid since its flow can be controlled by the intensity of the magnetic field. The magnetic field in any particular direction indicates that the fluid is be polarized in the same direction and polarization depends on the strength of the magnetic field. The present results might be applicable where strong changes in the magnetic field produce an electronically controllable signal and electronically controlled damping systems.

Nomenclature

- H – magnetic field intensity
- M – magnetization
- p – fluid pressure
- r – distance from the rotational axis
- t – time
- v – velocity of ferrofluid
- μ – reference viscosity of ferrofluid
- μ_0 – magnetic permeability of free space
- v_r – radial velocity
- v_θ – tangential velocity
- v_z – axial velocity

- ρ – fluid density
- χ – magnetic susceptibility
- ν – kinematic viscosity of ferrofluid
- ω – angular velocity of the disk

References

- [1] Feynman R.P., Leighton R.B. and Sands M. (1963): *Lecturers on Physics*. – Addison-Wesley Reading, MA 1.
- [2] Shliomis M.I. (2004): *Ferrofluids as thermal ratchets*. – Physical Review Letters, vol.92, No.18, 188901.
- [3] Odenbach S. (2002): *Magneto Viscous Effects in Ferrofluids*. – Berlin: Springer-Verlag.
- [4] Neuringer J.L. and Rosensweig R.E. (1964): *Magnetic fluids*. – Physics of Fluids, vol.7, 1927.
- [5] Verma P.D.S. and Singh M. (1981): *Magnetic fluid flow through porous annulus*. – Int. J. Non-Linear Mechanics, vol.16, No.3/4, pp.371-378.
- [6] Verma P.D.S. and Vedan M.J. (1979): *Steady rotation of a sphere in a paramagnetic fluid*. – Wear, vol.52, pp.201-218.
- [7] Verma P.D.S. and Vedan M.J. (1978): *Helical flow of ferrofluid with heat conduction*. – Jour. Math. Phy. Sci., vol.12, No.4, pp.377-389.
- [8] Rosensweig R.E. (1985): *Ferrohydrodynamics*. – Cambridge University Press.
- [9] Schlichting H. (1960): *Boundary Layer Theory*. – New York: McGraw-Hill Book Company.
- [10] Karman V. (1921): *Über laminare and turbulente Reibung*. – Z. Angew. Math. Mech. I, pp.232-252.
- [11] Cochran W.G. (1934): *The flow due to a rotating disc*. – Proc. Camb. Phil. Soc., vol.30, pp.365-375.
- [12] Benton E.R. (1966): *On the flow due to a rotating disk*. – J. Fluid Mech., vol.24, No.4, pp.781–800.
- [13] Attia H.A. (1998): *Unsteady MHD flow near a rotating porous disk with uniform suction or injection*. – Journal of Fluid Dynamics Research, vol.23, pp.283-290.
- [14] Mithal K.G. (1961): *On the effects of uniform high suction on the steady flow of a non-Newtonian liquid due to a rotating disk*. – Quart J. Mech. and Appl. Math. XIV, pp.401–410.
- [15] Attia H.A. and Aboul-Hassan A.L. (2004): *On hydromagnetic flow due to a rotating disk*. – Applied Mathematical Modelling, vol.28, pp.1007-1014.
- [16] Venkatasubramanian S. and Kaloni P.N. (1994): *Effect of rotation on the thermo-convective instability of a horizontal layer of ferrofluids*. – International Journal of Engineering Sciences, vol.32, No.2, pp.237-256.
- [17] Belyaev A.V. and Simorodin B.L. (2009): *Convection of a ferrofluid in an alternating magnetic field*. – Jour. of Applied Mechanics and Technical Physics, vol.50, No.4, pp.558-565.
- [18] Sekar R., Vaidyanathan G. and Ramanathan A. (1993): *The ferroconvection in fluid saturating a rotating densely packed porous medium*. – International Journal of Engineering Sciences, vol.31, No.2, pp.241-250.
- [19] Attia H. (2009): *Steady flow over a rotating disk in porous medium with heat transfer*. – Nonlinear Analysis: Modelling and Control, vol.14, No.1, pp.21–26.
- [20] Frusteri F. and Osalusi E. (2007): *On MHD and slip flow over a rotating porous disk with variable properties*. – Int. Comm. in Heat and Mass Transfer, vol.34, pp.492-501.
- [21] Paras Ram, Anupam Bhandari, Kushal Sharma,
- [22] Sunil, Divya and Sharma R.C. (2005): *The effect of magnetic field dependent viscosity on thermosolutal convection in a ferromagnetic fluid saturating a porous medium*. – Transport in Porous Media, vol.60, pp.251-274.

- [23] Sunil, Sharma A., Shandil R.G. and Gupta U. (2005): *Effect of magnetic field dependent viscosity and rotation on ferroconvection saturating a porous medium in the presence of dust particles*. – International Communication in Heat and Mass Transfer, vol.32, pp.1387-1399.
- [24] Sunil, Bharti P.K., Sharma D. and Sharma R.C. (2004): *The effect of a magnetic field dependent viscosity on the thermal convection in a ferromagnetic fluid in a porous medium*. – Zeitschrift fur Naturforschung, vol.59a, pp.397-406.
- [25] Nanjundappa C.E., Shivakumara I.S. and Arunkumar R. (2010): *Benard-Marangoni ferroconvection with magnetic field dependent viscosity*. – Journal of Magnetism and Magnetic Materials, vol.322, pp.2256-2263.
- [26] Ram P., Bhandari A. and Sharma K. (2010): *Effect of magnetic field-dependent viscosity on revolving ferrofluid*. – Journal of Magnetism and Magnetic Materials, vol.322, No.21, pp.3476-3480.
- [27] Ram P. and Bhandari A. (2013): *Negative viscosity effects on ferrofluid flow due to a rotating disk*. – Int. Journal of Applied Electromagnetics and Mechanics, vol.41, No.3, pp.467-478.

Received: April 24, 2014

Revised: April 4, 2016



Forecasting investment and capacity addition in Indian airport infrastructure: Analysis from post-privatization and post-economic regulation era



D.P. Singh ^{a,*}, Narendra N. Dalei ^b, T. Bangar Raju ^a

^a Dept. of Transportation, University of Petroleum & Energy Studies, Kandoli Campus, Knowledge Acres, Dehradun, 248007, India

^b Dept. of Economics & International Business, University of Petroleum & Energy Studies, Kandoli Campus, Knowledge Acres, Dehradun, 248007, India

ARTICLE INFO

Article history:

Received 3 September 2015

Received in revised form

5 March 2016

Accepted 7 March 2016

Keywords:

Traffic

Cargo

Passenger

Investment

Capacity addition

Forecast

ABSTRACT

Indian economy has been on the high growth trajectory for last two and half decades and it is expected to remain a high growth economy for few more decades given the favorable demographic structure and macroeconomic fundamentals. As the per capita income increases in an open economy, the opportunity cost of time consumed for travelling goes up, thereby, the demand for civil aviation increases. Indian economy is no exception as the middle class population, which has been progressively increasing, is opting for safe and less time consuming air travel. The same is true for transportation of high value-added products through air freight. In this context, the aim of this study is to forecast air traffic, capacity addition and investment required for capacity expansion in Indian civil aviation sector for next 20 years. In order to achieve this, we have collected last 20 years' data on air traffic, Gross Domestic Product (GDP) and Index of Industrial Production (IIP) from Airport Authority of India (AAI), World Bank website and India STAT website respectively. This data has been used to work out GDP and IIP elasticity of traffic. Using these elasticities, air traffic has been projected for next 20 years, which has been used to forecast required capacity additions and investment. In a nutshell, the findings of this study are that in next 20 years 866 million passenger terminal capacity and 7.53 thousand Metric Tons (MT) cargo terminal capacities will be required at the investment of about US\$ 25.94 billion.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

With the advent of economic liberalization during 1991 there has been increase in the economic activities. Additionally with the resultant economic boom, disposable income of individuals has touched new heights. The real per capita GDP of India, which was growing at a compound annual growth rate (CAGR) of 3.9% during 1992–2001, started growing at an accelerated CAGR of over 6% during 2004–2015. Even during the recent global meltdown, India's economy was least affected, and in any case it recovered faster than the recovery of any other economy, thus, explaining the strong economic fundamentals of Indian economy. The recent trend in economic growth reveals that Indian economy is expected to be on a high growth trajectory during the next 20 years and Indian

Aviation Sector will also grow at faster pace in tandem with the economy. Thus, along with increase in growth, India would need to develop its aviation infrastructure in order to accommodate the projected demand.

Before economic liberalization and introduction of open sky policy in 1991, aviation was traditionally viewed as an elite activity. The two government airlines i.e., Air India and Indian Airlines were the only Indian carriers. With the advent of open sky policy, private airlines entered into the Indian sky, first as air taxi operators and then as scheduled operators. Indian aviation sector witnessed an unprecedented change and growth after 2003. During this period, the importance of aviation, for the development of business, trade and tourism was recognized, and the industry saw dramatic reforms across the aviation value chain.

In 2003, there were just three private carriers viz., Jet Airways, Air Sahara and Air Deccan—all operating full service models. The private carriers in those days were limited to operating domestic routes only. In 2015, there are six private carriers viz., Jet Airways, Kingfisher, Spice Jet, Indigo, Vistara and Go Air. These are operating

* Corresponding author.

E-mail addresses: dpsingh@ddn.upes.ac.in (D.P. Singh), ndalei@ddn.upes.ac.in, ndalei@gmail.com (N.N. Dalei), tbaju@ddn.upes.ac.in (T.B. Raju).

List of acronym

GDP	Gross Domestic product
IIP	Index of Industrial Product
AAI	Airports Authority of India
India	
STAT	Statistical database of India
CAGR	Compound Annual Growth Rate
LCC	Low Cost Carrier
US	United States
MOCA	Ministry of Civil Aviation
ICAO	International Civil Aviation Organization
TU	Traffic Units

RPK	Revenue Passenger Kilometers
RTK	Revenue Tones Kilometers
PPP	Public Private Partnership
FDI	Foreign Direct Investment
DGCA	Directorate General of Civil Aviation
CNS	Civil Navigation Services
ANS	Air Navigation Services
MT	Metric Tones
MPPA	Million Passengers per Annum
MMTPA	Million Metric Tons per Annum
TPA	Traffic per Annum
ATU	Airport Throughput Unit

under nine brand names, and three of them are permitted to operate on international routes.

During the 11th Five Year Plan Period, domestic carriers embraced the Low Cost Carrier (LCC) model. The market share of LCC during 2014–15 had crossed 40% of the total domestic traffic. During the current 12th Five Year Plan period, the domestic traffic for Indian carriers is growing at a healthy average annual rate of around 9%. Also, the traffic growth has resulted in increased capacity utilization of domestic carriers.

Ground handling business at Indian airports has grown to reach a size of about US\$ 308 million. This segment also witnessed increased participation of private players, such as, SATS, Celebi, Bird Group, Menzies,¹ etc. In Joint Ventures (JVs), AIR India SATS (AISATS) is a JV between national carrier Air India and Singapore Air Transport Services. In 2011, Ministry of Civil Aviation (MOCA, 2015) announced a new ground handling policy, under which only three ground handlers were allowed at each of the six metro airports in the country. One was an Air India subsidiary, the other a subsidiary of the airport operator and the third one, an entity selected through competitive bidding.

Airports Authority of India (AAI) continued its leadership in creating air connectivity across the country by incurring expenditure to the tune of US\$ 1.9 billion during the 11th Five Year Plan period and US\$ 10.8 billion has been planned for the 12th Five Year Plan period. AAI has upgraded and modernized 35 non-metro airports in the country, at an estimated cost of US\$ 692 million. AAI is enhancing air connectivity in the Northeast by way of Greenfield airports at Pakyong (Sikkim), Itanagar (Arunachal Pradesh) and Cheitu (Nagaland).

The private sector played a major role during the 11th Five Year Plan in the development of airports through Public Private Partnership (PPP) model. These include development of Greenfield International airports at Bangalore and Hyderabad, and modernization of Delhi and Mumbai international airports through consortiums and Special Purpose Vehicles (SPVs).

In view of the above, airports are being viewed as commercial enterprises rather than public service organizations. Accordingly any progressive commercial enterprises require additional investment to sustain the future growth/addition in demand. The overarching question that arises here, in the above backdrop, is how much future investment is needed in airport sector in order to meet the growing demand of its services. The forecasting of investment requirement, being the major objective of the study, can be achieved with the help of forecasting air traffic, aircraft movement, and

capacity enhancement for passenger and cargo terminals for Indian airports. Thus, forecasting of investment in airport sector will enable planners and policy makers to take correct decisions. Additionally, the government would need to allocate sufficient budget for major airports to meet growing demand of airport infrastructure services.

The remaining part of the study is presented as follows: The Section 2 presents the literature survey; Section 3 and Section 4 contain Methodology and Data source and Descriptive Data Analysis including explanation of variables respectively. Results and discussion are described in Section 5 and followed by Conclusions and Policy suggestions are given in Section 6.

2. Review of literature

A smooth-functioning air transport sector offers significant economic development benefits, particularly for landlocked, isolated, and low population-density countries (The World Bank). The ever growing demand for air travel has put pressure on airports to enhance their capacity. Understandably, it is in order to continuously provide smooth service to passengers (Zou, Kafle, Chang, & Park, 2015). The increasing demand for aviation services has led to significant pressure on existing airport infrastructure (Gelhausen et al., 2013; Saldiraner, 2013). Airports are not only locations for transport activity, but also act as hubs of commercial infrastructure (Wells and Young, 2004), that in turn offers numerous opportunities to airport owners to explore the commercial space with resultant enhanced financial gains. Aviation is a driver of economic and social development of a country. The turnover of the Indian Aviation Sector today exceeds US\$ 15.38 billion. Private sector has played an unprecedented role for developing the airport sector in the country (Damodaran, 2015). Air Transport can play a key role in economic development and in supporting long-term economic growth. It facilitates a country's integration into the global economy, thereby providing direct benefits for users. It also accrues wider economic benefits through its positive impact on productivity and economic performance (ATA).

Since privatization there has been significant increased investment in Australian airports, with over \$2.2 billion invested in new terminals, runways and other infrastructure at the leased federal airports (Australian Government, 2008). The “airport planning paradigm is shifting from the traditional pattern, being determined by high standards, established customers and long-term forecasts, to that of recognizing great uncertainty at forecasts, broad range standards and potential for a rapidly changing customer's base” (de Neufville, 2008; Magalhaes, Reis, & Maca'rio, 2015). Air transport

¹ SATS, Celebi, Bird Group, Menzies are global cargo handling companies.

demand forecasts of the aircraft industry and institutions like ICAO (International Civil Aviation Organization) use the number of passenger kilometers, counted as revenue passenger kilometers (RPK), as a unit of demand (Gelhausen, Berster, & Wilken, 2013). The demand, as measured in RPK, grew even stronger than the number of passengers during the sixteen year period from 1994 to 2010; the demand more than doubled and increased with an average growth rate of 5%.

Kasarda (2006) and Graham (2009) have a view that secondary business, such as, parking, retailing etc. in expanding airports has been increasingly evident. This is based on the fact that aircrafts evolved from its historical status as an innovative transport technology to an everyday transport mode (Stevens, 2006). In view of this the Australian Government recognizes the importance of continued investment in aeronautical infrastructure at airports, and is committed to ensuring infrastructure development (Australian Government, 2008). Crabtree et al. (2015) have forecasted that over the next 20 years, world air cargo traffic will grow by 4.7% per year; air freight, including express traffic, will average 4.8% annual growth measured in Revenue Tones Kilometers (RTKs); airmail traffic though will grow slowly, averaging 1.0% annual growth through 2033. In essence, world air cargo traffic is expected to increase from 207.8 billion RTKs in 2013 to 521.8 billion in 2033.

The literature on liberalization and foreign direct investment in the aviation sectors of India, People's Republic of China, and Thailand brings out a number of key points. Firstly, greater competition has emerged within domestic markets, including from privately owned airlines, especially from low-fare carriers. Secondly, a higher level of foreign participation in airline operations helps provide funding and management capacity that supports the adjustment process required in the incumbent carriers (Findlay and Goldstein, 2004). The long term forecasts of Boeing and Airbus, as well as that of the ICAO, have in common a continuation of the past and development over the next 20 years. Based on this they forecast further liberalization of air transport in future as one of the key drivers of growth, especially in Asian and African regions (Gelhausen, Berster, & Wilken, 2013). FDI inflows in air transport (including air cargo) during April 2000 to January 2015 stood at US\$ 562.65 million. Air Costa plans to add eight aircrafts before 2016 to its existing fleet, Boeing is planning to set up an aircraft manufacturing base in India. Vistara has signed an inter-line agreement with Singapore Airlines and Silk Air & Tata Group has launched its full-service Vistara Airline on January 9, 2015 (IBEF, 2015).

The entry of low-cost carriers, pioneered by Air Deccan, helped to greatly reduce the costs involved in flying. This helped attracting consumers for whom air travel was only a dream. Now a number of low-cost airlines are operating in India, namely; Go Airways, Spice Jet, Indigo etc., which put together, have a major share of the Indian aviation sector. Thus, domestic participation in this industry is projected to grow by 25–30% and internationally by 15%, increasing the potential customers by about 100 million in 2010. Also, by 2020 the cargo section is projected to rise to approximately three million tonnes (The World Bank). International markets contribute 16% in terms of traffic generation and 29% of all connecting passengers in the US airport network (Suau-Sanchez, Voltes-Dorta, & Rodríguez-Déniz, 2015). The number of passengers transported worldwide in air transportation has touched a volume of almost 2500 million in 2010 (ICAO, 2011).

Notwithstanding the above the biggest problem in India is the liquidity crunch. Indian aviation does not have enough funds to meet the growing demand. Therefore, the alternate is to invite FDI (Vidhusekhar, 2014). Director of Civil Aviation (DGCA) guideline suggests that in Greenfield projects, FDI up to 100% is allowed under the automatic route. In case of existing projects, FDI up to 74%

is allowed through automatic route, and beyond that and up to 100%, with prior approval of the Government (DGCA, 2013).

The policies of the Indian government encourage foreign participation. Government allows 100% FDI via the automatic route for the green field airports. Also, foreign investment up to 74% is permissible through direct approvals while special permissions are required for 100% investment. Private investors are allowed to establish general airports and captive airstrips while keeping a distance of 150 km from the existing ones. Complete tax exemption is also granted for 10 years. About 49% FDI is allowed for investment in domestic airlines via the automatic route. However, this option is not available for foreign airline corporations. Complete equity ownership is granted to NRIs (Non Resident Indians). Foreign direct investment up to 74% is allowed for non-scheduled and cargo airlines. As a result, all these policies promote foreign investment in this industry (The World Bank).

If traffic reaches levels that are close to the maximum throughput of the runway system, then the airport encounters not only problems of maintaining good quality of operations, but is faced with the fact that future traffic growth cannot be accommodated any more (Gelhausen, Berster, & Wilken, 2013). Zhang (2003) has a view that China lacks an efficient logistic-related infrastructure, which includes not just airport capacity, but road networks and technological capabilities as well. Thus, modern transportation, communications, and logistics systems are essential to support major airports like Shanghai, Beijing, Guangzhou, and Shenzhen to become regional air-cargo hubs. Some important airports, partly main hub airports have been struggling already for years with capacity constraints, among them are: London, Heathrow, Frankfurt, Paris Charles de Gaulle in Europe, and New York LaGuardia in the USA (Gelhausen, Berster, & Wilken, 2013).

Crabtree et al. (2015) have also projected that Asia will continue to lead the world in average annual air cargo growth, with domestic China and intra-Asia markets expanding 6.7% and 6.5% per year, respectively. The Asia–North America and Europe–Asia markets will grow slightly faster than the world average growth rate. Latin American markets with North America and Europe will grow at approximately the world average growth rate, as will Middle East markets with Europe. The growth of established markets is slower than the growth of the developing markets. In the light of this fact North America and Europe air cargo growth rates are below the world average rate. The Indian aviation industry is forecasted to grow phenomenally in the coming years. The Vision 2020, announced by the Civil Aviation Ministry, conceives building infrastructure to support 280 million customers. Investments to the extent of US\$ 110 billion are envisaged by 2020. About US\$ 30 billion for development and sprucing up of existing airports and US\$ 80 billion for building new fleets is being estimated (The World Bank).

Investment forecast for airport infrastructure, such as; for passenger terminal, cargo terminals and Air Navigation Services (ANS)/Communication and Navigation Services (CNS) services have not been observed in any of the literature as of now. This study is unique and one of its kind in bridging the literature gap in forecasting such investment requirement. This is attempted by forecasting air traffic, aircraft movement, and capacity addition for passenger and cargo terminals for Indian airports.

3. Methodology and Data source

The historical data collected from AAI for the period 1995–96 to 2014–15 for all Indian airports traffic (together) has been used for econometric modeling. GDP (1995–96 to 2014–15) of India collected from World Bank website has been used as explanatory variable for forecast of International passengers and domestic

passengers. Index of industrial production (IIP) for the time period 1995–96 to 2014–15 has been used as explanatory variable for forecast of Cargo traffic.

Initially, trend analysis with linear model and econometric analysis with linear regression model, double log/exponential model taking real GDP of India as independent variable and air passenger traffic as dependent variable were undertaken. However, we got some disadvantages in linear models, which start underestimating the future air traffic. Underestimates also continue to increase with increase in time horizon in long term forecast; therefore, linear models were not selected. The final double log model was selected because it gives increasing increments with the increase base of traffic which is validated statistically, based on 20 years historical data for air traffic.

The aircraft movements have been projected based on the ratios of passengers to number of aircraft movement. The detail methodologies both for passenger and cargo traffic have been explained in the following sections.

3.1. Methodology of estimating past CAGR of air traffic

3.1.1. Relationship between passenger traffic and economic growth

Economic growth affects air passenger traffic significantly. Similarly, industrial output also affects cargo traffic significantly. This has been established by all international organizations viz. International Civil Aviation Organization, Airport Council International Europe, Aircraft Manufacturers and operators of the major airports worldwide. Thus, the response of passenger air traffic due to change in economic growth, and the response of cargo traffic due to change in industrial output plays an important role in estimating the growth rate of air traffic. This can help in arriving at future forecast of investment in aviation sector with a view to meeting growing demand of aviation services.

In view of the above, we assume an exponential relationship between passenger air traffic and economic growth, and between cargo air traffic and industrial output. By this logic, the relationship between passenger traffic and economic growth can be written as:

$$P_t = f(Y_t) \quad (1)$$

Where, P: passenger traffic.
Y: GDP (proxy for economic growth).
t: Time period.

Assuming exponential relationship between passenger traffic and economic growth, we can write:

$$P_t = \beta_1 Y_t^{\beta_2} \quad (2)$$

Taking natural log to both side of eq. (2), we get,

$$\ln P_t = \ln \beta_1 + \beta_2 \ln Y_t \quad (3)$$

Differentiating both sides of eq. (3) w. r.t. Y_t ,

$$\begin{aligned} \frac{1}{P_t} \frac{dP_t}{dY_t} &= \beta_2 \frac{1}{Y_t} \\ \Rightarrow \frac{Y_t}{P_t} \frac{dP_t}{dY_t} &= \beta_2 \end{aligned} \quad (4)$$

Now in eq. (4), ceteris paribus β_2 measures the elasticity of passenger traffic due to change in economic growth. That is how much passenger traffic is responding to percentage change in economic growth keeping all other factors constant.

3.1.2. Relationship between cargo traffic and industrial output

Similarly, the relationship cargo traffic and industrial output can be written as:

$$C_t = f(Q_t) \quad (5)$$

Where, C: cargo traffic. Q: industrial output. t: Time period.

Assuming exponential relationship between cargo traffic and index of industrial production, we can write:

$$C_t = \alpha_1 Q_t^{\alpha_2} \quad (6)$$

Taking natural log to both side of eq. (2), we get,

$$\ln C_t = \ln \alpha_1 + \alpha_2 \ln Q_t \quad (7)$$

Differentiating both sides of eq. (3) w. r.t. Q_t ,

$$\begin{aligned} \frac{1}{C_t} \frac{dC_t}{dQ_t} &= \alpha_2 \frac{1}{Q_t} \\ \frac{Q_t}{C_t} \frac{dC_t}{dQ_t} &= \alpha_2 \end{aligned} \quad (8)$$

Now in eq. (4), ceteris paribus α_2 measures the elasticity of cargo traffic due to change in industrial output. That is how much cargo traffic is responding to percentage change in industrial output keeping all other factors constant.

3.1.3. Future growth of passenger traffic

We have considered the elasticity β_2 and economic growth rate (R) to compute future compound annual growth rate (CAGR) of passenger traffic. Accordingly, the projected CAGR can be written as:

$$\mu = \beta_2 R \quad (9)$$

Where, μ_2 : CAGR of passenger traffic. R: economic growth rate. β_2 : GDP elasticity of passenger traffic.

The μ arrived above is then adjusted for decreasing trend in GDP elasticity of passenger traffic and increase in air fare (passenger yield) in real terms.

3.1.4. Future growth of cargo traffic

We have considered the elasticity α_2 and index of industrial production (IIP) to compute future compound annual growth rate (CAGR) of cargo traffic. So the projected CAGR can be written as:

$$\omega = \alpha_2 \lambda \quad (10)$$

Where, ω : CAGR of cargo traffic. λ : IIP of industrial output. α_2 : IIP elasticity of cargo traffic.

The ω arrived above is then adjusted for decreasing trend in IIP elasticity of cargo traffic and increase in air fare (passenger yield) in real terms.

3.1.5. Projected aircraft movement

The aircraft movements have been projected based on forecast of the ratio of passengers to number of aircraft movement.

3.1.6. Methods of forecasting investment in airport sector

The plan period wise traffic forecast has been used to work out capacity addition for passenger and cargo terminals and ANS/CNS. These capacity additions have been used to derive investment requirement on the basis of norms used during previous five year plans.

Table 1
Descriptive statistics.

Variables	Min.	Max.	Mean	Std. Dev.
International Aircraft Movement (IAM)	92.52	345.36	196.20	95.16
Domestic Aircraft Movement (DAM)	314.73	1257.66	727.40	369.54
Total Aircraft Movement (TAM)	407.24	1603.02	923.60	464.07
International Passenger Traffic (IPT)	11.45	50.80	25.21	13.12
Domestic Passenger Traffic (DPT)	23.85	139.33	62.96	40.82
Total Passenger Traffic (TPT)	36.50	190.13	88.17	53.87
International Cargo Traffic (ICT)	452.85	1542.55	928.62	404.61
Domestic Cargo Traffic (DCT)	196.52	985.02	496.90	253.33
Total Cargo Traffic (TCT)	649.37	2527.57	1425.53	656.17
Gross Domestic Production (GDP)	448.72	1584.72	891.06	359.21
Index of industrial production (IIP)	123.3	373.2	238.13	89.70

Source: Compiled by the authors

4. Descriptive Data Analysis

The descriptive statistics at Table 1 and Fig. 1 envisaged that the Indian economy over the annual period from 1995 to 1996 to 2014–15 would grow at a rate of 6.8%, while, its industrial output would grow at 6.5% for the same period. It also has been observed from Table 1 and Fig. 1 that during the period 1995–96 to 2005–06 the international and domestic passenger traffic increased by 10.1% and 13.3% respectively, whereas the total passenger's movement registered a maximum growth of 12.4% CAGR.

The international passenger traffic had increased from 11.45 million in 1995–96 to 50.80 million in 2014–15, leading to an increase of 4.4 fold. The domestic passenger traffic had increased from 25.56 million in 1995–96 to 139.33 million in 2014–15 which works out to be 5.5 fold. This highlights that the domestic traffic had increased at a faster rate of 9.3% CAGR as against the corresponding CAGR of 8.2% for international passenger traffic during the same period.

The international passenger traffic has grown at a CAGR of 8.1%, 10.1%, 9.3% and 8.2% during 5 years, 10 years, 15 years and 19 years respectively. The corresponding growth rates for domestic passenger are 9.3%, 13.3%, 11.9% and 9.3% respectively, leading to a growth of 9%, 12.4%, 11.1% and 9% respectively in total passenger traffic.

From the Table 1 and Fig. 1, it is seen that, in absolute terms, the international cargo traffic had increased from 452.85 thousand MT in 1995–96 to 1542.55 thousand MT in 2014–15 which represents an increase of 3.4 fold or 6.7% CAGR. The domestic cargo traffic had increased from 196.52 thousand MT in 1995–96 to 985.02 thousand MT in 2014–15 representing a growth of over 5 fold or 8.9% CAGR.

The CAGR for international cargo traffic during last 5 years, 10 years, 15 years and 19 years are 4.0%, 6.5%, 7.4% and 6.7%

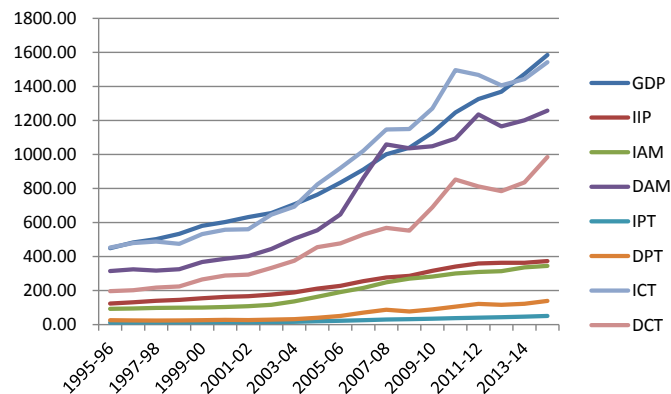


Fig. 1. Trend of economic growth, industrial growth, aircraft movement and air traffic.
Source: Compiled by the authors

respectively. The corresponding CAGR for domestic cargo traffic during last 5 year, 10 year, 15 year and 19 year are 7.4%, 8.0%, 9.1% and 8.9% respectively. In other words, domestic cargo traffic has been creasing at a faster rate as compared to growth rate of international cargo traffic due to its smaller base, speedy economic development and effort of Government of India to restrict the import in the country.

In absolute terms the international aircraft movements increased from 92.52 thousands in 1995–96 to 345.36 thousands in 2014–15 registering a CAGR of 7.2% or an increase of 3.7 fold. The domestic aircraft movement increased from 314.73 thousands to 1257.66 thousands in 2014–15 representing a CAGR of 7.6% or a growth of about 4 fold (see Fig. 1).

5. Results and discussion

5.1. Air traffic elasticities

In this study GDP elasticity and IIP elasticity have been estimated in order to have some basic understanding as to how well passenger air traffic and cargo air traffic are responding to both changes in GDP and IIP respectively. Understanding and use of responsiveness of GDP and IIP, along with their respective future growth rate, are prerequisites in forecasting air traffic which we have attempted in this study. GDP will continue to grow at the rate of 7%–8% up to 2030–31 as per revised forecast of GDP by Government of India. Indian IIP will continue to grow at 7–9% as projected by Government of India.

Table 2 also elucidates that the GDP elasticity of domestic passenger traffic is 1.64 which indicates that domestic passenger traffic is highly responsive to economic growth i.e. 10% increase in GDP stimulates 16.4% growth in domestic passengers traffic at 1% level of significance. The GDP elasticity of international passenger traffic is 1.28 which indicates that international passenger traffic is also highly responsive to economic growth i.e. 10% increase in GDP stimulate 12.8% growth in domestic passengers traffic significantly at 1% level of significance.

Similarly, the IIP elasticity of domestic cargo traffic is 1.38. This indicates that domestic cargo traffic is also highly responsive to growth in industrial production, i.e. 10% increase in GDP stimulate 13.8% growth in domestic cargo traffic significantly at 1% significant level. The IIP elasticity of international cargo traffic is 1.17 which indicates that international cargo traffic is also highly responsive to growth in industrial production, i.e. 10% increase in GDP stimulates 11.7% growth in domestic cargo traffic significantly at 1% level of significance (see Table 2).

5.2. Air traffic forecast

Based on the above discussion, the traffic forecast has been prepared making use of elasticities for passenger and cargo traffic

Table 2
Air traffic elasticities (1995–96 to 2014–15).

Variables	lnDPT	lnIPT	lnDCT	lnICT
lnGDP ^a	1.64*** (0.082)	1.28*** (0.039)	–	–
lnIIP ^b	–	–	1.38*** (0.038)	1.17*** (0.028)

***significant at 1% level ($p < 0.01$).

Note: 1. ^aGDP of India, ^bIIP of India.

2. Standard errors are given in the parentheses.

3. "ln" stands for natural log.

Source: Author's estimation

Table. 3

Traffic Forecast and Capacity addition up to 2031–32.

Forecast period	Passengers(in millions)			Cargo (in '000 MT)			Aircraft Movement (in '000)		
	IPT	DPT	TPT	ICT	DCT	TCT	IAM	DAM	TAM
2014–15	50.8	139.3	190.1	1542.49	986.37	2528.86	345.46	1257.56	1603.02
2024–25	109.67	329.77	439.45	3034.31	2129.5	5163.81	679.57	2714.98	3394.55
2031–32	176.11	565.17	741.29	4872.44	3649.59	8522.03	1021.83	4359.66	5381.49
2014–15 to 2031–32 (Capacity Addition)	125.31	425.87	551.19	3329.95	2663.22	5993.17	676.37	3102.1	3778.47

Source: Author's estimation

and ratio trends projection for aircraft movements, which are presented in Table. 3.

Table. 3 Suggests that international passenger traffic is projected to grow at the rate of 7–8% and will reach 176.11 million at the end of 2031–32, whereas domestic passenger traffic is projected to grow at 8–9% and will reach 565.17 million at the end of 2031–32. Thus, the total passenger traffic is expected to touch 741.29 million by the end of 2031–32.

The international cargo has been projected to grow at the rate of 7%, and is meant to become 4872.44 thousand MT by 2031–32 and the domestic cargo traffic has been projected to grow at a relatively higher growth rate of 8% and is expected to hover around 3649.59 thousand MT by the end of 2031–32. The international aircraft movement is projected to grow between 6 and 7% and will reach 1021.83 thousands at the end of 2031–32. With this projection the domestic aircraft moment is meant to touch 4359.66 thousand in 2031–32 with a projected growth of 7–8%.

To handle the above projection there will be a need for capacity addition in major 17 Indian airports viz. Delhi, Mumbai, Bangalore, Hyderabad, Cochin, Nagpur, Chennai, Kolkata, Trivandrum, Ahmedabad, Goa, Calicut, Guwahati, Jaipur, Srinagar, Amritsar and Port Blair to accommodate the projected traffic. The projected traffic can be accommodated in major 17 Indian airports, provided adequate amount of capacity addition with the required amount of projected investment is undertaken by Government of India.

5.3. Capacity addition and investment forecast

As mentioned above, there is a need for capacity addition in Aviation Sector in India. To meet the future air traffic demand, capacity addition is required, which in turn requires the investment in the airport infrastructure, and that includes investment in passenger terminal, cargo terminal, Air Navigation Services (ANS)/Communication and Navigation Services (CNS), safety & security etc. During the next 17–20 years, an additional capacity of about 866 MPPA will be required besides the existing capacity of 233 MPPA. Out of 866 MPPA capacities, the 366 MPPA is envisaged to be added by end of 2021–22 and another 500 MPPA by end of 2031–32. This will require the augmentation of the capacity by expanding the existing terminals, creating new terminals at Brownfield airports² and creation of 30–35 Greenfield airports.

During the next 17–20 years, 7.53 million metric tonnes per annum (MMTPA) cargo capacity is projected to be added. Out of the 7.53 MMTPA, 3.18 MMTPA cargo capacities are envisaged to be added by the end of 2021–22. Another 4.37 MMTPA will be added by end of 2031–32.

Table. 4 shows that investment of US\$ 26 million per million passengers is required in passenger terminals and investment of US\$ 158 million per million metric tones of cargo is required, which will amount to US\$ 22.57 billion for passenger terminals and US\$

1.2 billion for cargo terminals. Beside this US\$ 2.2 billion is forecasted for upgraded ANS/CNS services. This will result into a total investment of US\$ 25.94 billion by end of 2030–31.

India is an emerging economy, and it does have a potential to grow in future as mentioned above. So far as the trade, tourism, culture, health services and education services are concerned India is highly recognized worldwide. All these sectors as of now have well defined domestic and international air connectivity. But as has already been mentioned our projected air traffic needs major amount of capacity addition in passenger terminals, cargo terminals, air navigation services and air communication services. Therefore there will be a need of total investment of around US\$25.94 billion by end of 2031–32, which will generate capacity addition in passenger terminals of around 419% more, in cargo terminals of around 327% more and in ANS/CNS of around 372% more. These will be over and above their respective existing capacity. In order to meet these capacity additions, there will be need of investment of around 263% more in passenger terminals, 328% more in cargo terminals and only 14% more in ANS/CNS by end of 2031–32. These too will be over and above their respective projected investments during 2012 to 2016–17. The target of capacity addition clearly shows that for every five years, the growth rate of capacity addition in passenger terminals must be around 43.5%, for cargo terminals around 40% and for ANS/CNS around 50%. In order to meet these targeted growth rates in capacity addition, the growth rate of investment in every prospective five years span must be around 42.25% in passenger terminals, 47.2% in cargo terminals and 13.24% in ANS/CNS.

The total projected investment of US\$ 25.94 billion must be allocated among 17 major airports based upon their Airport Through-put Units (ATU). These are computed considering revenue earned, air craft movement, passenger traffic and cargo traffic of respective air ports. The ATU of India ranges between 1.5 million to 70.3 million with lowest being Port-Blair and highest being Delhi. The percentage allocation of projected investment of US\$ 25.94 billion is shown in Fig. 2, which indicates that a maximum amount of US\$ 6.51 billion and a minimum amount of US\$ 0.14 billion needs to be invested in Delhi and Port-Blair airports respectively.

6. Conclusion and policy recommendations

India, being an emerging economy, has potential to grow along with its trade, tourism, culture, health services and education services. All these sectors have a direct and indirect link with air connectivity domestically and internationally, and the same has already been well defined. However, our projected air traffic needs major amount of capacity addition in passenger terminals, cargo terminals, air navigation services and air communication services to support these sectors. Thus, a sound forecasting of air traffic is a pre-requisite to forecast investment and capacity addition in airport infrastructure, which we have done in our present study.

Our analysis finds that domestic and international passenger traffics are responding well to economic growth. Domestic and

² Brownfield airports are existing airports, Greenfield airports are airports developed at new sites.

Table 4

Forecast of Capacity addition and Investment (US\$1 = INR65).

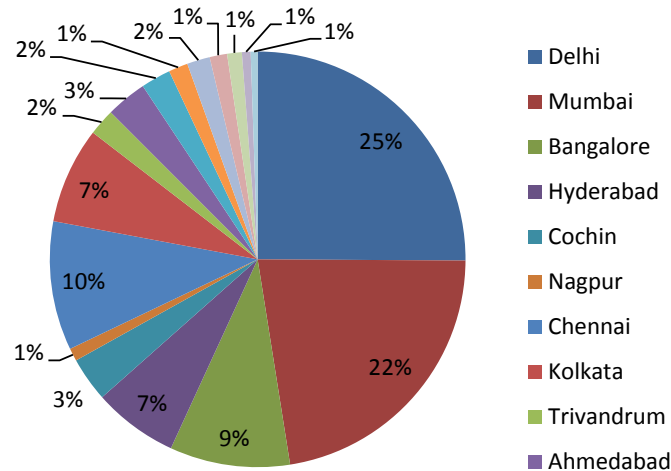
Forecasting period	Passenger terminal		Cargo terminal		ANS/CNS		Total
	Capacity addition (in million)	Investment (US\$ in million)	Capacity addition (in '000 MT)	Investment (US\$ in million)	Capacity addition (in '000 Movements)	Investment (US\$ in million)	Investment (US\$ in million)
2012–13 to 2016–17	64	2486	0.6	115	0.32	677	3278
2017–18 to 2021–22	302	4545	2.58	240	0.85	204	4989
2022–23 to 2026–27	168	6526	1.79	344	1.18	523	7394
2027–28 to 2031–32	332	9012	2.56	492	1.51	769	10,274
Total	866	22,570	7.53	1192	3.86	2173	25,935

Source: Author's estimation

international cargo traffic is also responding well to growth in industrial production. However, in order to meet the growing demand of airport infrastructure services in India and to support the trade, tourism, health, education, manufacturing and other service sectors, there is an urgent requirement of capacity addition in aviation sector. This in turn requires the investment in the airport infrastructure that includes investment in passenger terminal, cargo terminal, Air Navigation Services (ANS)/Communication and Navigation Services (CNS), safety & security etc. In this regard our study establishes that out of 866 MPPA capacities, the 366 MPPA is envisaged to be added by end of 2021–22 and another 500 MPPA by end of 2031–32. During next 17–20 years, 7.53 MMTPA Cargo Capacity is projected to be added, out of which, 3.18 MMTPA cargo capacity is envisaged to be added by the end of 2021–22 and 4.37 MMTPA by end of 2031–32. Thus, by end of 2031–32, Indian Government would invest US\$ 25.94 billion in major 17 airports, in proportion of traffic handled by each airport, with a minimum of 1% in Port-Blair airport and a maximum of 25% in Delhi airport in order to meet the growing demand of airport infrastructure services.

One can safely say that by adopting scientific forecasting methods, discussed in the paper, will minimize the risk involved and maximize the achievement of planned target. One of the major problems in present planning system was that the cost overruns were indicating increasing trend due to overshoot of actual resource required as compared to planned resources. Thus, this study will help airport infrastructure planner to reduce such cost to a greater extent.

However, the major shortcoming of this study is the

**Fig. 2.** Allocation of Projected Investment among 17 major Indian Airports.

Source: Compiled by the authors

FIVE YEAR PLAN PERIODS REFERRED IN THE PAPER

Plan	Plan period	Date
11th Plan	2007–08 to 2011–12	1–4–2007 to 31–3–2012
12th Plan	2012–13 to 2016–17	1–4–2012 to 31–3–2017
13th Plan	2017–18 to 2021–22	1–4–2017 to 31–3–2022
14th Plan	2022–23 to 2026–27	1–4–2022 to 31–3–2027
15th Plan	2027–28 to 2031–32	1–4–2027 to 31–3–2032

NOTE: India follows financial year for example 2011–12 represents the Period from 1st April 2011 to 31st March 2012.

unavailability of yield data for passenger and cargo traffic for past 20 years. Consequently, due to data deficiency, price (yield) elasticity could not be worked out, which would have further improved the quality and accuracy of the forecast. Since the study was a long term forecast, therefore, confidence interval has not been worked out. Since long term planning is indicative planning, accordingly, pessimistic and optimistic scenario have not been worked out required for short term and medium term planning.

This study is a novel attempt which will guide airport infrastructure planners and policy makers of India, and also rest of the world to take correct future capacity addition and investment decisions in order to meet growing demand of airport infrastructure services. Timely decision(s) on future capacity addition and investment decision(s) will reduce cost overruns; avoid traffic congestion and also tackle miscellaneous problems pertaining to passenger terminals, cargo terminals Air Navigation Services, Communication and Navigation Services, safety and security etc.

Acknowledgment

The authors acknowledge Airport Authority of India and its member staffs viz. Mr. S. Raheja, Member (Planning); Ms. Madhu Batra (GM&CP); Mr. V. Krishnan; Joint GM (CP & MS); Ms. Renu Grover, Senior Manager (EP) and Mr. Gopal Chand, Manager (EP) for providing data related to this study. The authors also acknowledge Mr. Anshuman Behera, MA Economics student of UPES for help in data compilation and statistical analysis. The authors gratefully acknowledge Prof. SPS Narang, UPES and Dr. Pravakar Sahoo, Institute of Economic Growth, Delhi for their valuable comments on earlier version of this paper.

References

- ATA. (n.d.). Aviation Economic Benefits. ATA.
- Australian Government, 2008. National Aviation Policy Green Paper: Flight Path to the Future. Canberra ACT 2601. Department of Infrastructure, Transport, Regional Development and Local Government, Commonwealth of Australia.
- Crabtree, T., Hoang, T., Edgar, J., Tom, R., 2015. World Air Cargo Forecast. The Boeing Company. Seattle, Washington 98124–2207 USA.
- de Neufville, R., 2008. Low-cost Airports for low-cost airlines: flexible design to

- manage the risks. *J. Transp. Plan. Technol.* 31 (1), 35–68.
- DGCA., 2013. DGCA Guideline. DGCA, New Delhi.
- Damodaran, K., 2015. The economics of indian aviation sector. *AKG J. Technol.* 6–11.
- Findlay, C., & Goldstein, A. (n.d.). Liberalization and Foreign Direct Investment in Asian Transport Systems: The Case of Aviation. *Asian Development Review*, 21(1), 37–65.
- Gelhausen, M., Berster, P., Wilken, D., 2013. Do airport capacity constraints have a serious impact on the future development of air traffic? *J. Air Transp. Manag.* 3–13.
- Graham, A., 2009. How important are commercial revenues to today's airports? *J. Airtransport Manag.* 15 (3), 106–111.
- IBEF., 2015. Indian Aviation Industry. Retrieved 7 8, 2015, from. www.ibef.org: www.ibef.org.
- International Civil Aviation Organization (ICAO), 2011. ICAO Traffic Statistics. International Civil Aviation Organization, Montreal.
- Kasarda, J., 2006. The rise of the aerotropolis. *Next Am. City* 35–37.
- Magalhaes, L., Reis, V., Maca'rio, R., 2015. Can flexibility make the difference to an airport's productivity? an assessment using cluster analysis. *J. Air Transp. Manag.* 47, 90–101.
- MoCA. (2015).
- Saldıraner, Y., 2013. Airport master planning in Turkey; planning and development problems and proposals. *J. Air Transp. Manag.* 32, 71–77. September.
- Stevens, N., 2006. City airports to airport cities. *Qld. Plan.* 37–40.
- Suau-Sanchez, P., Voltes-Dorta, A., Rodríguez-Déniz, H., 2015. Regulatory airport classification in the US: the role of international markets. *Transp. Policy* 157–166.
- The World Bank. (n.d.). The Role of the Public and Private Sector. The World Bank.
- Vidhusekhar, P., 2014. Foreign Direct Investment Role in Indian Aviation Market. KU, Kerala.
- Wells, A.T., Young, S.B., 2004. *Airport Planning and Management*, fifth ed. McGraw Hill, New York.
- Zhang, A., 2003. Analysis of an international air-cargo hub: the case of Hong Kong. *J. Air Transp. Manag.* 123–138.
- Zou, B., Kafle, N., Chang, Y.-T., Park, K., 2015. US airport financial reform and its implications for airport efficiency: an exploratory investigation. *J. Air Transp. Manag.* 66–78.



Fuzzy-frequency ratio model for avalanche susceptibility mapping

Satish Kumar, Snehmani, Pankaj Kumar Srivastava, Akshay Gore & Mritunjay Kumar Singh

To cite this article: Satish Kumar, Snehmani, Pankaj Kumar Srivastava, Akshay Gore & Mritunjay Kumar Singh (2016): Fuzzy-frequency ratio model for avalanche susceptibility mapping, International Journal of Digital Earth, DOI: [10.1080/17538947.2016.1197328](https://doi.org/10.1080/17538947.2016.1197328)

To link to this article: <http://dx.doi.org/10.1080/17538947.2016.1197328>



Published online: 29 Jun 2016.



Submit your article to this journal [↗](#)




View related articles [↗](#)



View Crossmark data [↗](#)



Fuzzy–frequency ratio model for avalanche susceptibility mapping

Satish Kumar^{a,b}, Snehmani^b, Pankaj Kumar Srivastava^a, Akshay Gore^b and Mritunjay Kumar Singh^c 

^aCollege of Engineering Studies, University of Petroleum & Energy Studies (UPES), Dehradun, Uttarakhand, India; ^bSnow and Avalanche Study Establishment, Defence Research & Development Organisation (DRDO), Chandigarh, India; ^cSpatial Strokes Technologies Private Limited, Zirakpur, Punjab, India

ABSTRACT

Avalanche activities in the Indian Himalaya cause the majority of fatalities and responsible for heavy damage to the property. Avalanche susceptibility maps assist decision-makers and planners to execute suitable measures to reduce the avalanche risk. In the present study, a probabilistic data-driven geospatial fuzzy–frequency ratio (fuzzy–FR) model is proposed and developed for avalanche susceptibility mapping, especially for the large undocumented region. The fuzzy–FR model for avalanche susceptibility mapping is initially developed and applied for Lahaul-Spiti region. The fuzzy–FR model utilized the six avalanche occurrence factors (i.e. slope, aspect, curvature, elevation, terrain roughness and vegetation cover) and one referent avalanche inventory map to generate the avalanche susceptibility map. Amongst 292 documented avalanche locations from the avalanche inventory map, 233 (80%) were used for training the model and remaining 59 (20%) were used for validation of the map. The avalanche susceptibility map is validated by calculating the area under the receiver operating characteristic curve (ROC-AUC) technique. For validation of the results using ROC-AUC technique, the success rate and prediction rate were calculated. The values of success rate and prediction rate were 94.07% and 91.76%, respectively. The validation of results using ROC-AUC indicated the fuzzy–FR model is appropriate for avalanche susceptibility mapping.

ARTICLE HISTORY

Received 9 February 2016
Accepted 31 May 2016

KEYWORDS

Snow avalanche; avalanche susceptibility mapping; GIS; fuzzy–frequency ratio (fuzzy–FR) model; Lahaul & Spiti

1. Introduction

Snow avalanche is the most destructive phenomenon in the snow bound mountainous region of Western Indian Himalaya, which poses a serious threat to human, economic, communication, property and other infrastructure. Due to heavy snowfall, active tectonic setting, snowstorm, weather and topographical conditions, Western Indian Himalaya experience high rates of avalanche activity with an average of 30–40 deaths each year (Ganju, Thakur, and Rana 2002; Ganju and Dimri 2004; Gardner and Saczuk 2004; Sharma, Mathur, and Snehmani 2004). Due to climate change and weather variability, such hazard may increase in the future. Taking into consideration the present state of affairs and the future view, the avalanche threat is likely to assume greater proportions in the near future. Therefore, there is growing demand for avalanche susceptibility mapping over large and undocumented mountainous region. Through scientific analysis of avalanche activities, avalanche workers can assess and predict avalanche susceptible areas and hence minimize the avalanche

risk through appropriate preparation. Geospatial modelling is recognized as a significant approach for the assessment of avalanche susceptibility (Selçuk 2013).

In past years, a number of avalanche runoff modelling studies are performed with statistical models such as alpha-beta (α - β) and runoff ratio models (Sinickas and Jamieson 2014), dynamic models such as Voellmy-Salm model (Salm 1993), Voellmy-fluid model (Bartelt, Salm, and Gruber 1999), two-parameter model (Perla, Cheng, and McClung 1980), particle model (Perla 1984), leading edge model (McClung and Mears 1995), AVAL-1D (Oller et al. 2010), erosion and deposition model (Naaïm, Faung, and Naaïm-Bouvet 2003), Snow Avalanche MOdelling and Simulation (SAMOS) model (Sailer, Rammer, and Sample 2002), the modified version of SAMOS, that is, SamosAT (Sailer et al. 2008) and rapid mass movement simulation (Christen, Kowalski, and Bartlett 2010). Extensive research has been performed on the identification and probabilistic assessment of weather and snow cover conditions influencing avalanche activity (Hendrikx et al. 2005; McClung and Schaerer 2006; Jomelli et al. 2007; Schweizer, Mitterer, and Stoffel 2009). On the other hand, relatively little research has been conducted on the identification of potential avalanche release areas (Bühler et al. 2013; Pistocchi and Notarnicola 2013) and susceptibility mapping of snow avalanches (Selçuk 2013; Snehmani et al. 2014).

A heuristic (expert judgement) method, the analytical hierarchy process (AHP) model (Saaty 1980) based on weighting and rating has been successfully attempted for avalanche susceptibility mapping (Selçuk 2013; Snehmani et al. 2014). The modified-AHP model has been introduced by Nefeslioglu et al. (2013) for natural hazard assessment, especially for avalanches. The primary benefit of this method is the values incorporated in the model depending on the expert judgements. Sometimes, this process could lead to the main drawbacks, particularly in pair-wise comparisons and results the imperfection of the model. The weight of evidence and logistic regression models based on data-driven technique were also attempted for avalanche mapping (Pistocchi and Notarnicola 2013).

On the other hand, modelling of avalanches using fuzzy approach are limited (Jaccard 1990; Barpi 2004; Ghinai and Chung 2005; Zischg et al. 2005). Integration of fuzzy logic (Zadeh 1965) with AHP and frequency ratio approach (Bonham-Carter 1994) (fuzzy-FR) are widely applied for natural hazard and geo-environmental problems (Aksoy and Ercanoglu 2012; Huang, Li, and Wang 2012; Kavzoglu, Sahin, and Colkesen 2014; Shahabi, Hashim, and Ahmad 2015). As per knowledge, these models are not ever developed for avalanche susceptibility mapping. Hence, the fuzzy-FR model provides novelty to the present investigation.

Fuzzy logic (Zadeh 1965) is commonly utilized to solve complicated and decision-making problems. The fuzzy set theory has been successfully and widely used in different disciplines of the complex degree of uncertainties (Barpi 2004; Bui et al. 2012). In fuzzy set theory, an entity is a member of an entity set if it has a membership degree of 1. On the other hand, if the entity has a membership degree of 0, then it is not a member of the entity set (Hines 1997). This theory is idealized to map the spatial entities as the members of an entity set (Pourghasemi et al. 2013).

In natural hazard assessment, the FR value is the ratio of the probabilities of hazard occurrence to the non-occurrence factors' attributes of the total area (Bonham-Carter 1994; Pradhan and Lee 2010). The FR method provides the capability to calculate the level of the relationship between the dependent and independent variables (Oh et al. 2011). The theoretical background of fuzzy logic and FR method is described in the next sections.

Avalanche susceptibility mapping is most effective if detailed topographical and meteorological information can be combined with an avalanche inventory map. However, these types of homogeneous input data set are frequently not available over large areas. Six main terrain factors of snow avalanche occurrence used in this study are slope, aspect, curvature, elevation, terrain roughness and vegetation cover, respectively. An avalanche inventory map was used to exploit the information of the avalanche occurrence locations. Through fuzzy-FR model, we investigated the relationship between terrain factors and the spatial distribution of observed avalanche locations: using the variables showing highest association with avalanches, we generated maps of the

favourability to avalanche susceptibility on the basis of observed avalanche events. These maps were subsequently tested using independent observations, enabling the assessment of their predictive ability. The analysis identifies the most relevant factors that explain observed avalanches, and conclusions are drawn about the advantages and limitations of the fuzzy–FR method in avalanche susceptibility mapping.

This study was also intended by considering that the avalanche occurrence factors are commonly affected by a degree of uncertainty highly, due to complex mountainous topography. The fuzzy–FR model is effective in avalanche mapping by assuming the uncertainty degree of the avalanche occurrence factors.

2. Theoretical background of fuzzy set theory and FR method

2.1. Fuzzy set theory

The fuzzy set theory (Zadeh 1965) is extensively applied to complex problems, which are difficult to state precisely in crisp values. It is also capable of permitting the vague information (Feizizadeh et al. 2014). Fuzzy logic allows to handle the notion of partial truth in which the factors can be characterized with the degrees of truth and false (Barpi 2004). The fuzzy logic is helpful in the development of expert knowledge based and naturally vague large complex systems. In geospatial mapping and management process, the fuzzy logic is able to utilize the spatial entities on a map as memberships of a spatial entity or a crisp set (Feizizadeh et al. 2014). However, a spatial object of the set can assign to membership values between 0 and 1, which defines a degree of membership.

A fuzzy set is defined as a set of systematic pairs in Equation (1) (Feizizadeh et al. 2014):

$$A\{z, MF(z)\}, \quad z \in Z, \quad (1)$$

where $MF(z)$ is a membership function (MF) of the set A , Z indicates a space of entities. Figure 1 shows a linear membership function (LMF) which comprises the basis for the MFs. This function has four parameters such as a , b , c and d , which determine the shape of the function. These parameters

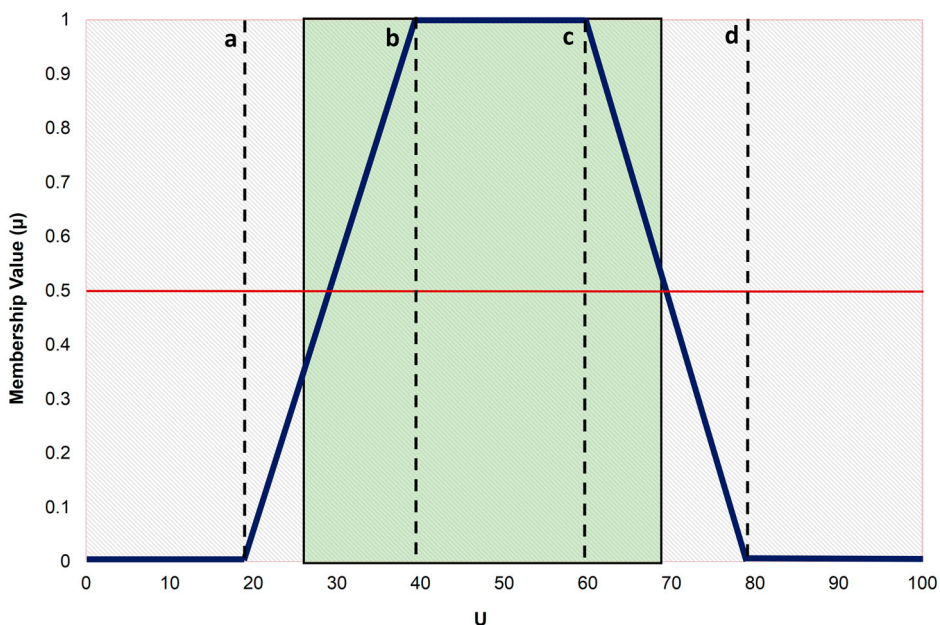


Figure 1. Linear membership function.

signify the smallest, the most promising and the largest favourable values to define a fuzzy entity (Kahraman, Cebeci, and Ulukan 2003). The LMF is defined using this concept, which is given as (Kainz 2008)

$$\mu_A(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & b < x < c, \\ \frac{d-x}{d-c} & c \leq x \leq d, \\ 0 & x > d. \end{cases} \quad (2)$$

The different shaped membership functions such as the trapezoidal, S & L-shaped and triangular can also be created for the appropriate purpose by using the proper values of a , b , c and d , respectively (Kainz 2008).

2.2. FR method

FR is the probability of the occurrence of a particular event (Bonham-Carter 1994). For avalanche susceptibility mapping, this method is based on the calculation of the relationship between observed avalanche occurrences and each avalanche occurrence factor. Suppose, for an event X and related event factors attributed to Y , the FR of Y is given in Equation (3) (Oh et al. 2011):

$$P\{X|Y\} = \frac{P\{X \cap Y\}}{P\{Y\}}. \quad (3)$$

The formula behind FR can also be expressed by the given Equation (4):

$$FR_{ij} = \frac{N_p(SX_i) / \sum_{i=1}^m SX_i}{N_p(SX_j) / \sum_{j=1}^n SX_j}, \quad (4)$$

where $N_p(SX_i)$ is used to represent the total number of avalanche occurrence pixels in class i of avalanche occurrence factor X ; $N_p(SX_j)$ is used to represent the total number of pixels in avalanche occurrence factor X ; m is the number of classes in the avalanche occurrence factor X ; n is the number of avalanche occurrence factors.

In the present study, avalanche occurrence locations are considered as dependent variables and terrain factors with vegetation cover values (which influence the avalanche occurrence) are considered as independent variables.

3. Study area

Lahaul-Spiti region of Himachal Pradesh, Indian Himalaya is selected for implementation of the fuzzy-FR model because this region is frequently affected by avalanche activities and documented avalanche locations data set is available for a few parts of this region. The present study area is one of the most avalanche-prone areas in the Western Himalaya and most parts of Lahaul region are prone to avalanche. The spatial extent of the study area is $76^\circ 21' 29.73''$ to $77^\circ 47' 57.44''$ East longitude and $32^\circ 05' 30.09''$ to $33^\circ 15' 20.85''$ North latitude (Figure 2) of highly complex mountainous zone having a mean altitude of 4701 m. The study area covers the area of approximately 6651.17 km² and encloses the Lahaul-Spiti district. The study area is mostly covered by the snow in late winter season. In late summer, all the seasonal snow melted and the ground features are exposed with permanent snow or ice-cover areas. The permanent snowline is positioned above 4260 m asl, while the lower limit of the snowline prolongs to 1500 m asl (Owen et al. 1995; Laxton and Smith 2009).

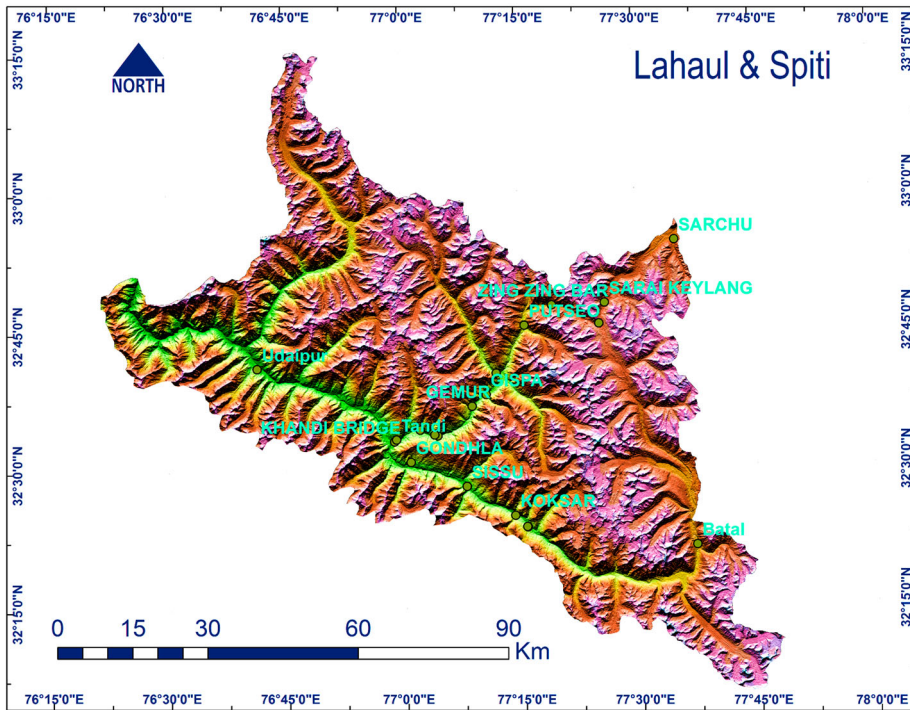


Figure 2. Location map of the study area.

4. Data set and methodology

4.1. Data set

Development of a model for avalanche susceptibility mapping required the identification of most prominent avalanche occurrence factors and collection of related data set. The analysis of avalanche susceptibility areas requires data of known avalanche locations and conditions that might explain their distribution (Pistocchi and Notarnicola 2013). Factually, the region-based avalanche susceptibility mapping should be practical and appropriate. Knowledge of internal and external stresses on the snowpack, which depends on a complex set of thermal and mechanical factors (McClung and Schaerer 2006), is essential in avalanche modelling, but information for this purpose is not systematically available over large regions. Snowpack stress conditions vary significantly with time. For the purposes of the spatial analysis targeted here, it would be necessary to define an indicator of the snowpack stress conditions in a map form. However, to the best of our knowledge, no such simple indicator has been identified so far. On the other hand, it may be argued that snow conditions, under the given weather over a region, are essentially controlled by the morphology of avalanche release areas (slope, aspect, curvature, elevation and terrain roughness) and vegetation cover. Therefore, we limited our consideration to these six factors.

Schweizer, Jamieson, and Schneebeli (2003) stated three types of parameters (1: terrain, 2: meteorological and 3: Snowpack) on which avalanche occurrence depends. Amongst these three parameters, terrain is the most influencing parameter, and easier to map due to its temporal stability (Snehmani et al. 2014). In this study, ASTER GDEM V2 (30 m spatial resolution) has been utilized for the extraction of terrain parameters (slope, aspect, curvature and elevation) because as per the recent studies by Singh et al. (2015, 2016), ASTER GDEM V2 is the most accurate digital elevation model (DEM) amongst all the freely available moderate resolution DEMs for this region.

Table 1. Landsat 8 OLI bands and attributes.

No.	Band	Spatial resolution (m)	Wavelength (μm)
1	Band 1-Blue	30	0.43–0.45
2	Band 2-Blue	30	0.45–0.51
3	Band 3-Green	30	0.53–0.59
4	Band 4-Red	30	0.64–0.67
5	Band 5-Near-infrared (NIR)	30	0.85–0.88
6	Band 6-Shortwave infrared (SWIR)	30	1.57–1.65
7	Band 7-SWIR	30	2.11–2.29
8	Band 8-Panchromatic (PAN)	15	0.50–0.68
9	Band 9-Cirrus	30	1.36–1.38
10	Band 10-Thermal infrared (TIR)	100	10.60–11.19
11	Band 11-TIR	100	11.50–12.51

The six avalanche occurrence factors such as slope, aspect, elevation, curvature, terrain roughness and vegetation cover (normalized difference vegetation index [NDVI]) were obtained from the ASTER GDEM V2 (30 m \times 30 m) and Landsat 8 operational land imager (OLI) imagery (30 m \times 30 m). The attributes of Landsat 8 OLI bands are shown in Table 1.

4.2. Methodology

The overall methodology is summarized in flow chart presented in Figure 3. The flow chart shows the construction of avalanche-related database and procession stages of avalanche susceptibility mapping using the fuzzy-FR model. In the first part of the methodology, the snow-avalanche-occurrence-related data set were gathered and transformed into a database, including topographical factors and documented data of past avalanche events. An avalanche inventory map was prepared; subsequently the avalanche occurrence factors were analysed and reclassified. FR coefficients for each of these factor categories are computed and normalized in [0, 1] to express their FM degrees and termed as ‘Fuzzy membership functions’ (Bonham-Carter 1994). Finally, the fuzzy overlay analysis combines the avalanche occurrence factors to obtain an avalanche susceptibility map using Fuzzy OR operator. The results of avalanche susceptibility mapping using the fuzzy-FR model were compared with documented avalanche occurrence locations by calculating the area under the receiver operating characteristic curve (ROC-AUC) analysis technique. The detailed methodology is presented in the next subsections.

4.2.1. Preparation of avalanche inventory map

The first and main fundamental part of the avalanche susceptibility mapping is to acquire the information about the avalanches that have occurred in the past. The scenario of avalanche occurrence in the past and present is the source to the prediction of avalanches; an avalanche inventory map plays an important role in such study. Furthermore, the mapping of documented avalanche locations is important to define the relationship between the avalanche and occurrence factors. An avalanche inventory map gives the essential information for assessing avalanche susceptibility. Precise recognition of the location of snow avalanches is imperative for avalanche susceptibility mapping.

In order to obtain a comprehensive and detailed avalanche inventory map, exhaustive field surveys and interpretations are accomplished in the avalanche-prone region with well-supported high-resolution satellite imagery and Survey of India mapsheets. A total number of 292 avalanche locations were demarcated in the avalanche inventory map. Afterwards, the avalanche inventory map was divided into two groups: (i) training data and (ii) validation data. In training data, 233 (80%) number for avalanche locations were randomly selected and utilized for training the model. In validation data, the remaining 59 (20%) number of avalanche locations were kept and used for the purpose of testing and validation of the results, respectively.

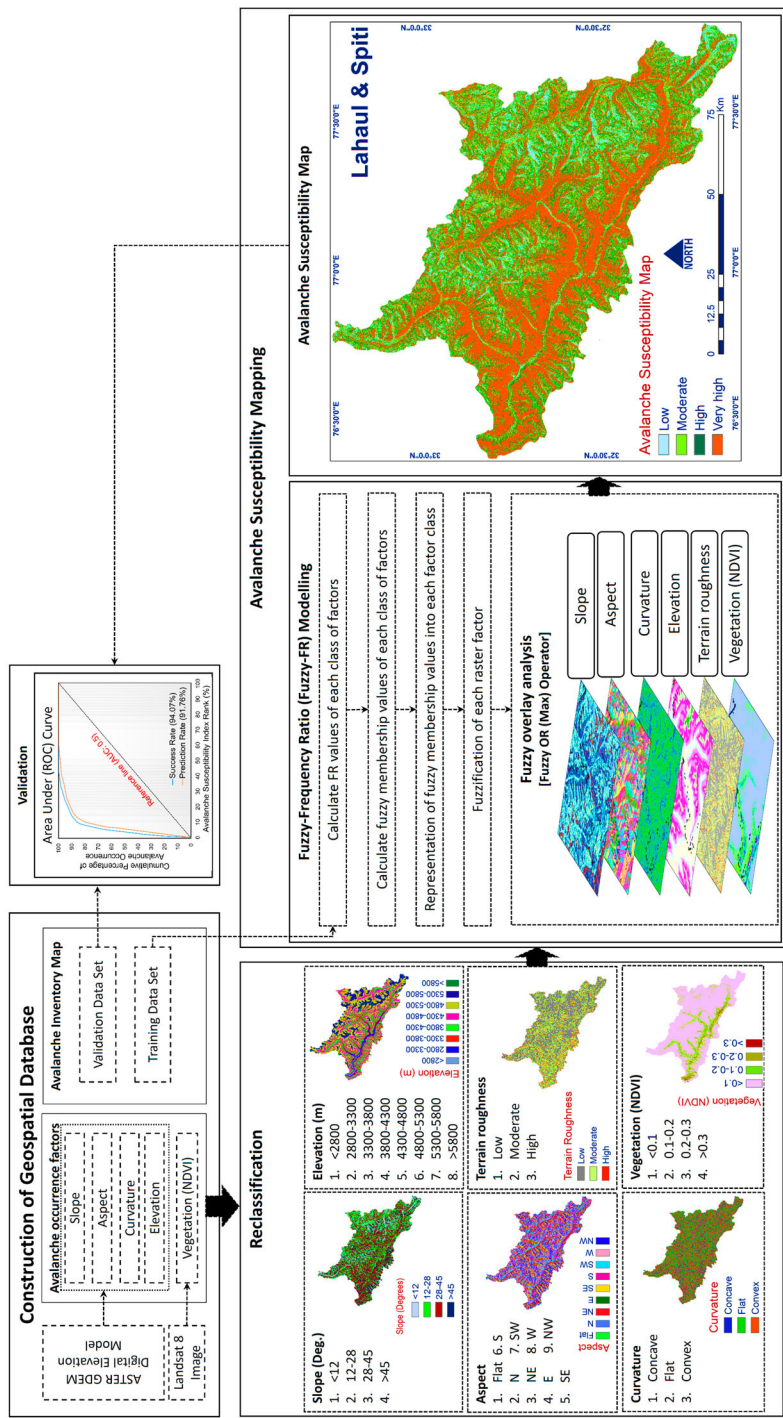


Figure 3. Flow chart of the methodology.

4.2.2. Analysis of avalanche occurrence factors and generation of thematic GIS layers

Avalanche formation conditions are crucial for the analysis of avalanche occurrence factors. Various significant avalanche occurrence conditions are meteorological, snowpack stability, topographical, natural triggers and anthropogenic activities. Understanding the snowpack stability is significant for avalanche susceptibility modelling, but the data are not thoroughly available for the present study area. The frequency of naturally triggered avalanches is more strongly correlated to topographical factors than meteorological factors (Smith and McClung 1997). Moreover, meteorological factors and snowpack stability depend on the weather condition, which keeps on changing continuously. However, the topographical factors are only constant factors in the avalanche susceptibility mapping. Six most prominent avalanche occurrence topographical factors such as slope, aspect, curvature, elevation, terrain roughness and ground cover are used for avalanche susceptibility mapping and summarized as follows.

The slope is an important and the primary avalanche occurrence factor (Ghinoi and Chung 2005; Bühler et al. 2013; Snehmani et al. 2014). Statistically, it is generally accepted that most of avalanches activities occurred in a region where the slope angle ranges from 25° to 45° and rarely at the slope angle less than 25° (McClung and Schaerer 2006; Selçuk 2013). The ASTER GDEM V2 was used to generate slope values using the maximum transformation in neighbours' cells values technique (Burrough and McDonnell 1998). Subsequently, the reclassification method was used to discriminate the slope values into four categories such as slope angle below 12°, 12–28°, 28–45° and above 45° (Figure 4(a)). While choosing the four slope classes, statistics presented in Table 2 were considered to modify the classes given by Selçuk (2013). As in the present study, none of the avalanche accident was observed in the area where the slope angle is less than 12° (Table 2), so instead of using 0–10° threshold as considered by Selçuk (2013) for no avalanche class, threshold limit 0–12° was used. Threshold limits for the rest of the slope classes were taken as considered by Selçuk (2013). In present study area, the slope angle ranges from 0° to 83°. The slope category that ranges from 28 to 45° has 78% documented avalanche slope followed by 18% in 12–28°. The slope class less than 12° showed no avalanche clues in the past. The attributes of all slope categories are mentioned in Table 2.

Aspect is a principal parameter associated with the avalanche occurrence (Ghinoi and Chung 2005; Selçuk 2013). The past avalanche accidents in Indian Himalaya has shown (Ganju, Thakur, and Rana 2002) that northern, eastern and southern slopes contribute almost equally in the release of avalanches. Since the dominant wind direction during winter months is westerly, southerly, easterly and northerly slopes get additional drift snow. The ASTER GDEM V2 was used to generate aspect values using an algorithm that incorporates the values of the cell's eight neighbours (Burrough and McDonnell 1998) and further reclassified into nine principal categories (Figure 4(b)). Six of nine aspect categories showed significant association with avalanche occurrence locations such as north, northeast, south, east, southeast and southwest, respectively (Table 2).

The curvature has deemed as an important terrain factor in the release of avalanches (Maggioni and Gruber 2003; Ghinoi and Chung 2005; Snehmani et al. 2014). Generally, convex slopes are more avalanche susceptible than flat and concave slopes (Nefeslioglu et al. 2013). The areas of sudden change in the curvature also contribute to the release of avalanches. In the present study, the curvature values have been derived from ASTER GDEM using the second derivative value of the input cell \times cell of surface (Moore, Grayson, and Landson 1991) and reclassified into three principal categories such as concave, convex and flat (Figure 4(c)).

Convex curvature showed maximum relationship with documented avalanche locations (43.78%) than concave (33.48%) and flat (22.75%) curvature (Table 2).

Elevation is an important terrain factor in the snow avalanche occurrence because snowpack frequently varied due to varying snowfall, wind and temperature at different elevations (McClung and Schaerer 2006; Guy and Birkeland 2013; Selçuk 2013). The wind speed increases at high elevation and helps in increasing the snow sliding. Higher elevation area receives heavy snowfall most of the time in the winter season and thus increasing the avalanche activity. The conditions on higher elevation area are suitable for avalanche occurrence because this area generally exposed by wind, sun,

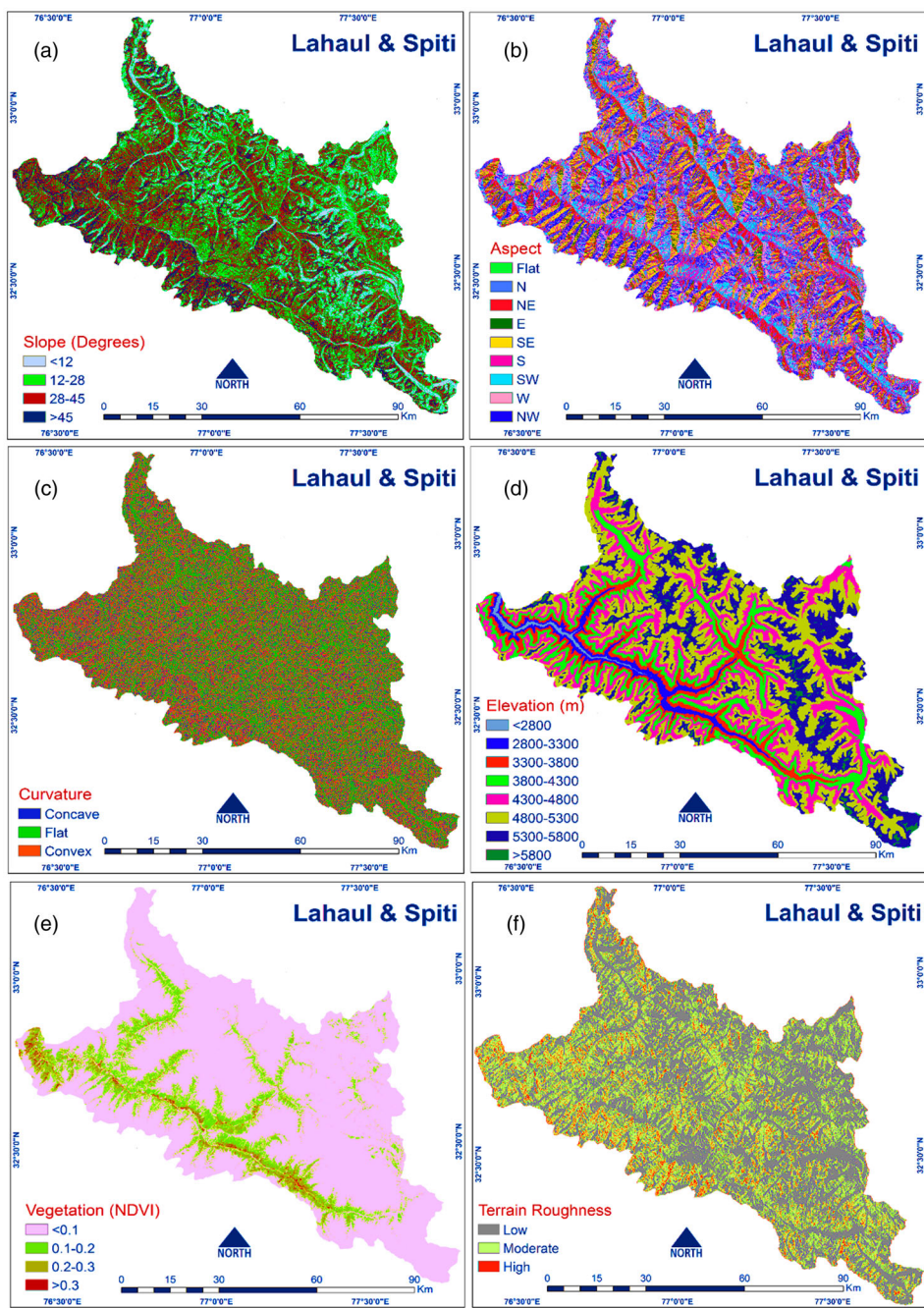


Figure 4. Avalanche occurrence factors: (a) slope, (b) aspect, (c) curvature, (d) elevation, (e) vegetation and (f) terrain roughness.

different snowpack conditions and minimum forest/vegetation cover (McClung and Schaerer 2006). In general, the avalanche activity at lower elevation is minor because the snow at minor elevation often melts and changes to rain. The topographical characteristics of Lahaul region are seemly favourable for avalanche occurrence. The study area has complex terrain with the altitude ranging from 2302 to 6443 m. In Western Indian Himalaya, maximum avalanches generally start at the elevation ranges from 2700 m and terminate at about 6000 m. All slopes beyond 5500 m elevation are generally glaciated. In Lahaul region, 98% of total documented avalanche sites have existed

Table 2. Spatial relationship between each avalanche conditioning factor, avalanche occurrence and fuzzy membership values.

Factor	Class	Class no.	No. of pixels in class		Avalanche occurrence		FR	Fuzzy membership function
			Number	%	Number	%		
Slope	0–12	1	1063674	14.45	0	0.00	0.00	0.00
	12–28	2	2424234	32.93	44	18.88	0.57	0.28
	28–45	3	2820771	38.31	182	78.11	2.04	1.00
	>45	4	1053553	14.31	7	3.00	0.21	0.10
Aspect	Flat	1	394	0.01	0	0.00	0.00	0.00
	N	2	885958	12.03	38	16.31	1.36	1.00
	NE	3	866318	11.77	30	12.88	1.09	0.80
	E	4	974675	13.24	38	16.31	1.23	0.91
	SE	5	972011	13.20	35	15.02	1.14	0.84
	S	6	872758	11.85	34	14.59	1.23	0.91
	SW	7	933351	12.68	26	11.16	0.88	0.65
	W	8	940738	12.78	16	6.87	0.54	0.40
	NW	9	916029	12.44	16	6.87	0.55	0.41
Curvature	Concave	1	2529797	34.36	78	33.48	0.97	0.46
	Flat	2	2170362	29.48	53	22.75	0.77	0.00
	Convex	3	2662073	36.16	102	43.78	1.21	1.00
Elevation	<2800	1	54807	0.74	0	0.00	0.00	0.00
	2800–3300	2	284256	3.86	4	1.72	0.44	0.22
	3300–3800	3	553951	7.52	36	15.45	2.05	1.00
	3800–4300	4	1004255	13.64	60	25.75	1.89	0.92
	4300–4800	5	1641751	22.30	52	22.32	1.00	0.49
	4800–5300	6	2318163	31.49	48	20.60	0.65	0.32
	5300–5800	7	1399584	19.01	33	14.16	0.75	0.36
	>5800	8	105465	1.43	0	0.00	0.00	0.00
NDVI	<0.1	1	6035049	81.97	194	83.26	1.02	1.00
	0.1–0.2	2	997811	13.55	31	13.30	0.98	0.92
	0.2–0.3	3	277598	3.77	7	3.00	0.80	0.46
	>0.3	4	51774	0.70	1	0.43	0.61	0.00
Terrain Roughness	Low	1	4854080	65.93	148	63.52	0.96	0.65
	Moderate	2	2011430	27.32	77	33.05	1.21	1.00
	High	3	496722	6.75	8	3.43	0.51	0.00

between the elevation ranges from 3200 to 5800 m. In the present study, the elevation range was reclassified into eight categories as shown in Figure 4(d). The attributes of all elevation categories are mentioned in Table 2. The elevation class of 3800–4300 m has maximum number of avalanche activities, which was followed by the elevation class of 4300–4800 m. The elevation classes of less than 2800 m and above 5800 m showed no avalanche occurrence/activity in the past.

Vegetation cover is another significant factor in the assessment of avalanche susceptibility. Avalanches generally released on barren slopes. Dense trees coverage hold the snow and protects against snow avalanches (Selçuk 2013). In general, vegetation cannot stop the snow avalanches, but highly vegetation coverage controls the volume of snow that contributes in the release of an avalanche. Grassy slopes release avalanches differently than those with shrubs. Based on the experienced judgement in Indian Himalaya, the forested areas would contribute to about only by 10% towards avalanche initiation. Shrubs, tall grass would contribute to about 30% towards avalanche initiation and the maximum of 60% is from barren slopes. The application of vegetation cover in the preparation of the GIS-based map has been achieved with the help of Landsat 8 OLI satellite image. The NDVI method was used for extraction of vegetation cover values.

The NDVI values were calculated using the formula from the Equation (5) (Ke et al. 2015)

$$NDVI = \frac{NIR - R}{NIR + R}, \tag{5}$$

where NIR and R are the energy reflected in the near-infrared and red portions of the electromagnetic spectrum.

The values of NDVI are further reclassified into four categories such as <0.1 , $0.1-0.2$, $0.2-0.3$ and >0.4 (Figure 4(e)). The vegetation cover increases with the increase in the NDVI values. The vegetation cover is dense with the NDVI values <0.2 . Snow avalanches showed a significant relationship with <0.1 NDVI values because maximum avalanches have occurred (83.26%) in this category of NDVI (Table 2).

Terrain roughness is a significant avalanche occurrence component. Terrain roughness reports a rugged and sporadic surface, which blocks the snowpack in the downward movement and prevent the formation of a consistent weak layer, important for fracture in the snowpack (McClung 2001; Schweizer, Jamieson, and Schneebeli 2003). In the present study, the terrain roughness is calculated using the ruggedness method of Sappington, Longshore, and Thompson (2007). The study area with ruggedness values are classified into three classes of roughness (Figure 4(f)). The spatial relationship between each class of terrain roughness with avalanche occurrences is shown in Table 2.

4.2.3. Fuzzy-FR modelling for avalanche susceptibility mapping

In fuzzy-FR modelling for avalanche susceptibility mapping, the pixel of GIS layer of avalanche occurrence factor is considered as susceptible to avalanches. The pixel values are ranges from 0 to 1. The value 0 signifies 'not susceptible to avalanche' and the value 1 signifies 'highly susceptible to avalanche', respectively. The values range from 0 to 1 can be chosen with the degree of membership of an entity set. The values of the degree of membership of an entity set can be derived on the basis of subjective judgement (Bonham-Carter 1994). These values can also be derived from FR or multi criteria decision analysis methods such as an AHP. Herein, the FR method is applied to derive the FM values.

4.2.3.1. FR values for each class of avalanche occurrence factors. The FR values for each class of the factor are calculated by using the formula in Equation (4), respectively. The avalanche occurrence location data set was considered as dependent variables and terrain factors with vegetation cover values (which influence the avalanche occurrence) were considered as independent variables. Using the FR method, the relationship between avalanche occurrence locations and each class of the occurrence factors were calculated and shown in Table 2.

4.2.3.2. Fuzzification of FR-based avalanche occurrence factors. To fuzzify the avalanche occurrence factors, the FR values for each class of avalanche occurrence factor are normalized in $[0, 1]$ to express their membership degrees using the formula

$$\mu(C_{ij}) = \frac{FR_{ij} - \text{Min}(FR_{ij})}{\text{Max}(FR_{ij}) - \text{Min}(FR_{ij})} [\text{Max}(\mu(C_{ij})) - \text{Min}(\mu(C_{ij}))] + \text{Min}(\mu(C_{ij})), \quad (6)$$

where $\mu(C_{ij})$ is the FM value; FR_{ij} is the FR value of a class of avalanche occurrence factor; $\text{Min}(FR_{ij})$ is the minimum FR value of an occurrence factor; $\text{Max}(FR_{ij})$ is the maximum FR value of an occurrence factor; $\text{Max}(\mu(C_{ij}))$ and $\text{Min}(\mu(C_{ij}))$ are the maximum and minimum normalization limits.

Therefore, for all factors, the membership functions are obtained and each class of avalanche occurrence factor was assigned a membership function. The membership functions were used to obtain a relationship between the selected factors and the degree of the membership of avalanche and no-avalanche locations. The membership function layers for avalanche susceptibility are generated using the fuzzy LMF, that is, 'FuzzyLinear ({min}, {max})' for avalanche occurrence factor maps (Figure 4(a-f)).

4.2.3.3. Fuzzy overlay analysis to obtain fuzzified avalanche susceptibility index. Fuzzy overlay analysis has the capability to combine various reclassified occurrence factors to calculate the natural

hazard (Kirschbaum, Stanley, and Yatheendradas 2015). Bonham-Carter (1994) described five fuzzy operators for combining the occurrence factors such as 'Fuzzy OR', 'Fuzzy AND', 'Fuzzy Product', 'Fuzzy Sum' and 'Fuzzy Gamma'. In present study, the 'Fuzzy OR' operator functions was used to combine the membership functions of the occurrence factors to obtain the fuzzified index to express the possibility of belonging to avalanche or not. The reason for using the 'Fuzzy OR' operator to combine the membership functions was to take the maximum value at each point to estimate the effectiveness factors representing the possibility of the avalanche occurrence. The remaining operators such as 'Fuzzy AND', 'Fuzzy Product', 'Fuzzy Sum' and 'Fuzzy Gamma' can also be utilized to combine the membership functions of the occurrence factors to generate the avalanche susceptibility map. In this study, all the membership functions of the avalanche occurrence factors are combined using the 'Fuzzy OR' operator given in Equation (7):

$$\mu_{\text{combination}} = \text{MAX}(\mu_A, \mu_B, \mu_C, \dots), \quad (7)$$

where $\mu_{\text{combination}}$ is the combination of the maximum values of each membership function of the factor, μ_A is the membership function for factor A at a particular location, μ_B is the membership function for factor B, μ_C is the membership function for factor C and so on.

The final fuzzified avalanche susceptibility index was obtained using Equation (7) and expressed in Equation (8):

$$\mu_{\text{ASI}} = \text{MAX}(\mu_{\text{Slope}}, \mu_{\text{Aspect}}, \mu_{\text{Curvature}}, \mu_{\text{Elevation}}, \mu_{\text{Terrain roughness}}, \mu_{\text{Vegetation}}), \quad (8)$$

where μ_{ASI} denotes the calculated avalanche susceptibility index; μ_{Slope} , μ_{Aspect} , $\mu_{\text{Curvature}}$, $\mu_{\text{Elevation}}$, $\mu_{\text{Terrain roughness}}$ and $\mu_{\text{Vegetation}}$ are the membership values for slope, aspect, curvature, elevation, terrain roughness and vegetation.

Subsequently, the performance of the avalanche susceptibility map was discussed in the next section.

5. Results and discussion

To create the fuzzy membership functions of the avalanche occurrence factors, the FR values were calculated for each class of the factors on the basis of Equations (3) and (4). FR values calculated for each class of the avalanche occurrence factors are shown in Table 2. These FR values depict the spatial relationship between the avalanche occurrence and avalanche occurrence factors. In slope classes, the class of the angle ranging from 28°– to 45° has the highest FR value of 2.04 then other slope classes (Table 2) and indicated a very high possibility of the avalanche. In the aspect factor, the avalanche activities were abundant in the north, northeast, east and southeast facing and lower in the case of west and northwest aspect angles (Table 2). The curvature factor represents the topographical morphology. A zero curvature value indicates the flat surface, negative curvature values indicates concave surface and positive curvature values indicates convex surface. The FR value for flat curvature indicated a lowest probability of avalanches than concave surface and convex surface (Table 2). In the case of elevation factor, the FR value of the elevation classes of the value below 2800 m and above 5800 m indicated a very low probability of avalanches. The FR values of the elevation classes 3300–3800 m and 3800–4300 m indicated very high probability of the avalanches. The FR values of the elevation class ranges 4300–4800 m, 4800–5300 m and 5300–5800 m indicated high to moderate probability of the avalanches (Table 2). The FR value for NDVI indicated the probability of the avalanche occurrence increases with a decrease in the NDVI values (Table 2). In case of the terrain ruggedness factor, the FR values for low and moderate ruggedness class indicated high probability of the avalanches than high ruggedness class (Table 2).

These FR values for each class of the occurrence factors were normalized to a scale from 0 to 1 to express their fuzzy membership functions using the formula given in Equation (6). The fuzzy membership functions of each class of the occurrence factors are shown in the Table 2. The value 0 is

allocated to the lowest observed avalanche occurrence classes and the value 1 allocated to the highest observed avalanche occurrence classes of the factors. The representation of fuzzy set for each avalanche occurrence factors were obtained and shown in Table 3. Subsequently, the thematic GIS layers were created on the basis of fuzzy set of the avalanche occurrence factors. To construct fuzzified index maps for each avalanche occurrence factors, a fuzzy linear membership ‘FuzzyLinear ({min}, {max})’ function was applied. A fuzzy overlay technique using Fuzzy OR (Max) operator in fuzzy mathematics was applied to combine all the fuzzified maps of the avalanche occurrence factors to create an avalanche susceptibility map.

Finally, the avalanche susceptibility map was classified into four classes/zones such as low, moderate, high and very high susceptibility based on the natural breaks (Jenks) method and shown in Figure 5. The Jenks method is a data grouping scheme intended to determine the best organization of values by interactively comparing sum of the squared difference between observed values within each class and class means (Jenks 1967). This method is also known as Jenks optimization method.

According to avalanche susceptibility zones, the high susceptibility zones are covering 32.09% (1405.91 km²). The very high susceptibility zones are covering 14.29% (626.03 km²). The moderate susceptibility zones are contributing 26.49% (1160.71 km²) area. The remainders moderate and low susceptibility zones are covering 20.35% (891.67 km²) and 6.78% (296.91 km²) of the study area.

The final avalanche susceptibility map indicates that the north, northeast and northwest parts of the Lahaul-Spiti region have the highest avalanche risk. The higher susceptibility zones concentrate on the middle of the study area and along the river and its torrents. The moderate susceptibility zones are in the middle and lower reaches and some in the upper reaches, most of which are located on both sides of high susceptibility zones.

The avalanche susceptibility map was verified by comparing the map with both the training data of avalanche occurrence locations that were used for building the models and with the avalanche occurrence locations that were not used during the model building process from the avalanche inventory map.

A ROC-AUC analysis technique was used to validate the results. For validation of the results by ROC-AUC technique, the success rate and the prediction rate were calculated. The training data of the documented avalanche occurrence locations was used to calculate the success rate of the avalanche susceptibility map. The validation data of the documented occurrence locations was used to calculate the prediction rate of the avalanche susceptibility map.

In this study, 233 (80%) of randomly selected avalanche locations from inventory map were used for training the model and remaining 59 (20%) were used for validation of the model results. The results show that the values of AUC for success rate and prediction rate were 0.9407 and 0.9176. Hence, the accuracy of success rate and prediction rate was 94.07% and 91.76%, respectively.

As per the results of the prediction rate, this is suggested that in higher susceptibility zones, human-related activities should avoid and not carry out projects in these susceptibility zones. In the moderate susceptibility zones, human activity should avoid these zones and engineering control standards should be implemented if necessary. Risk assessment must also be performed in the low susceptibility zones to ensure the threat of avalanche occurrence to humankind and other related infrastructure. This is also suggested that the dynamic avalanche models should be further utilized to quantify the avalanche risk and run-out areas of potential avalanche sites. The avalanche

Table 3. Fuzzy set of avalanche occurrence factors.

Occurrence factor	Fuzzy set
μ_s Slope	(0/1, 0.28/2, 1/3, 0.10/4)
μ_s Aspect	(0/1, 1/2, 0.80/3, 0.91/4, 0.84/5, 0.91/6, 0.65/7, 0.40/8, 0.41/9)
μ_s Curvature	(0.46/1, 0/2, 1/3)
μ_s Elevation	(0/1, 0.22/2, 1/3, 0.92/4, 0.49/5, 0.32/6, 0.36/7, 0/8)
μ_s Terrain roughness	(0.65/1, 1/2, 0/3)
μ_s Vegetation (NDVI)	(1/1, 0.92/2, 0.46/3, 0/4)

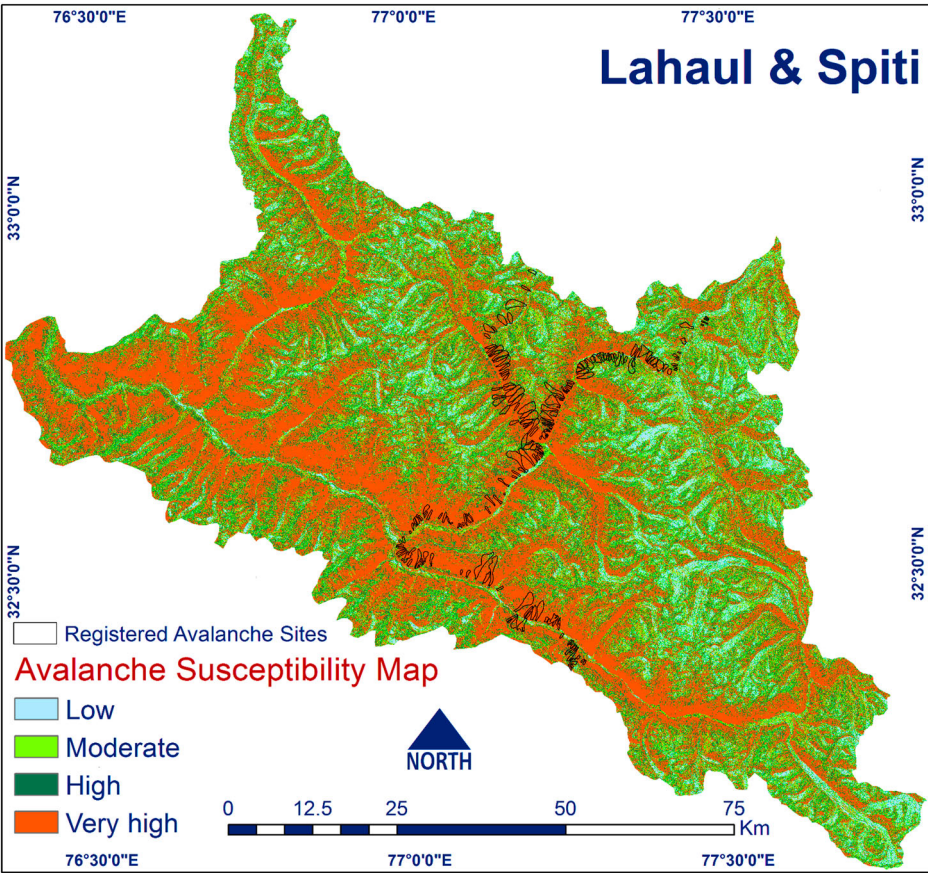


Figure 5. Avalanche susceptibility map.

susceptibility map obtained in this study can be a better reference for planners and engineers to design and develop control structures. This avalanche susceptibility map presents helpful information so that consciousness can be accorded to higher avalanche susceptibility zones for any kind of infrastructure development in the near future. The avalanche susceptibility map will be useful for concerned experts in disaster management, planning and mitigation purposes.

6. Conclusion

In this study, a GIS-based data-driven fuzzy–FR model is introduced for finding and spatial distribution of the avalanche susceptibility in Lahaul-Spiti region. Avalanche susceptibility mapping was performed using various topographical and satellite image in GIS in order to prepare the susceptibility maps. Six avalanche occurrence factors were selected for avalanche susceptibility mapping such as slope, aspect, curvature, terrain roughness, elevation and ground cover, which were generated from ASTER GDEM V2. The selection of these six avalanche occurrence factors were based on the easily availability of data and the highly relevance with respect to avalanche occurrences. However, more avalanche occurrence factors such as meteorological factors can be considered for avalanche susceptibility mapping.

The fuzzy–FR method is appropriate for avalanche susceptibility mapping, where documented avalanche locations’ data are limited or unavailable. This method should be considered when documentation data are limited and experts cannot go on the field to check all avalanche sites in detail.

This approach may be useful in undocumented or poorly documented areas, for the screening of avalanche potentials.

The model results are assessed using ROC-AUC technique. It is found that the fuzzy-FR model is applicable for avalanche susceptibility mapping over the mountainous region of Lahaul-Spiti. The resultant avalanche susceptibility map could be easily utilized efficiently to design control structures, to prepare avalanche risk maps and for planning and mitigation purposes. In such cases, the knowledge of the topographical and meteorology is beneficial.

The membership functions of avalanche occurrence factors used in the present model can also be inferred from expert judgement-based method, that is, AHP. In this approach, an expert knowledge can be easily included in the fuzzy approach.

The fuzzy-FR model reveals its applicability for avalanche susceptibility mapping over large undocumented mountainous region. Moreover, the model with fuzzy membership functions calculated in this study may be used in other study areas or valleys of Himachal Pradesh with similar geomorphological, meteorological and snowpack conditions, where documented avalanche occurrences data are limited or does not exist. The fuzzy-FR model would need revision when applied in areas with different conditions. In such cases, the model could be easily revised by modifying the fuzzy membership functions.

Acknowledgements

The authors are thankful to all the officers and other staffs members of SASE Chandigarh, India for their direct and indirect support for this research.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

Funding: The authors are thankful to Defence Research & Development Organisation (DRDO), Ministry of Defence, Government of India for providing funds under Project IABP/HimUdaan, No. IF-10/SAS-42, dated 30 September 2011 to carry out this research work.

ORCID

Mritunjay Kumar Singh  <http://orcid.org/0000-0003-2492-8842>

References

- Aksoy, B., and M. Ercanoglu. 2012. "Landslide Identification and Classification by Object-based Image Analysis and Fuzzy Logic: An Example from the Azdavay Region (Kastamonu, Turkey)." *Computers and Geosciences* 38 (1): 87–98. doi:10.1016/j.cageo.2011.05.010.
- Barpi, F. 2004. "Fuzzy Modelling of Powder Snow Avalanches." *Cold Regions Science and Technology* 40 (3): 213–227. doi:10.1016/j.coldregions.2004.08.003.
- Bartelt, P., B. Salm, and U. Gruber. 1999. "Calculating Dense Flow Avalanche Runout Using a Voellmy-fluid Model with Active/Passive Longitudinal Straining." *Journal of Glaciology* 45 (150): 242–254. doi:10.3198/1999JoG45-150-242-254.
- Bonham-Carter, G. F. 1994. *Geographic Information Systems for Geoscientists: Modeling with GIS*. Ottawa: Pergamon Press.
- Bui, D. T., B. Pradhan, O. Lofman, I. Revhaug, and O. B. Dick. 2012. "Spatial Prediction of Landslide Hazards in Hoa Binh Province (Vietnam): A Comparative Assessment of the Efficacy of Evidential Belief Functions and Fuzzy Logic Models." *Catena* 96: 28–40. doi:10.1016/j.catena.2012.04.001.
- Burrough, P. A., and R. A. McDonell. 1998. *Principles of Geographical Information Systems*. New York: Oxford University Press.

- Bühler, Y., S. Kumar, J. Veitinger, M. Christen, A. Stoffel, and Snehamani. 2013. "Automated Identification of Potential Snow Avalanche Release Areas based on Digital Elevation Models." *Natural Hazards and Earth System Sciences* 13: 1321–1335. doi:10.5194/nhess-13-1321-2013.
- Christen, M., J. Kowalski, and P. Bartlett. 2010. "RAMMS: Numerical Simulation of Dense Snow Avalanches in Three-dimensional Terrain." *Cold Regions Science and Technology* 63 (1–2): 1–14. doi:10.1016/j.coldregions.2010.04.005.
- Feizizadeh, B., M. S. Roodposhti, P. Jankowski, and T. Blaschke. 2014. "A GIS-based Extended Fuzzy Multi-criteria Evaluation for Landslide Susceptibility Mapping." *Computers and Geosciences* 73: 208–221. doi:10.1016/j.cageo.2014.08.001.
- Ganju, A., and A. P. Dimri. 2004. "Prevention and Mitigation of Avalanche Disasters in Western Himalayan Region." *Natural Hazards* 31: 357–371. doi:10.1023/B:NHAZ.0000023357.37850.aa.
- Ganju, A., N. K. Thakur, and V. Rana. 2002. "Characteristics of Avalanche Accidents in Western Himalayan Region, India." International Snow Science Workshop, Penticton, B.C..
- Gardner, J. S., and E. Sazuk. 2004. "Systems for Hazards Identification in High Mountain Areas: An Example from the Kullu District, Western Himalaya." *Journal of Mountain Science* 1 (1): 115–127. doi:10.1007/BF02919334.
- Ghini, A., and C. J. Chung. 2005. "STARTER: A Statistical GIS-based Model for the Prediction of Snow Avalanche Susceptibility using Terrain Features' Application to Alta val Badia, Italian Dolomites." *Geomorphology* 66 (1–4): 305–325. doi:10.1016/j.geomorph.2004.09.018.
- Guy, Z. M., K. W. Birkeland. 2013. "Relating Complex Terrain to Potential Avalanche Trigger Locations." *Cold Regions Science and Technology* 86: 1–13. doi:10.1016/j.coldregions.2012.10.008.
- Hendriks, J., I. Owens, W. Carran, and A. Carran. 2005. "Avalanche Activity in an Extreme Maritime Climate: the Application of Classification Trees for Forecasting." *Cold Regions Science and Technology* 43 (1–2): 104–116. doi:10.1016/j.coldregions.2005.05.006.
- Hines, J. W. 1997. *Fuzzy and Neural Approaches in Engineering*. New York: Wiley.
- Huang, S., X. Li, and Y. Wang. 2012. "A New Model of Geo-environmental Impact Assessment of Mining: A Multiple-criteria Assessment Method Integrating Fuzzy-AHP with Fuzzy Synthetic Ranking." *Environmental Earth Science* 66: 275–284. doi:10.1007/s12665-011-1237-z.
- Jaccard, C. 1990. "Fuzzy Factorial Analysis of Snow Avalanches." *Natural Hazards* 3 (4): 329–340. doi:10.1007/BF00124391.
- Jenks, G. F. 1967. "The Data Model Concept in Statistical Mapping." *International Yearbook of Cartography* 7: 186–190.
- Jomelli, V., C. Delval, D. Grancher, S. Escande, D. Brunstein, B. Hetu, L. Filion, and P. Pech. 2007. "Probabilistic Analysis of Recent Snow Avalanche Activity and Weather in the French Alps." *Cold Regions Science and Technology* 47 (1–2): 180–192. doi:10.1016/j.coldregions.2006.08.003.
- Kahraman, C., U. Cebeci, and Z. Ulukan. 2003. "Multi-criteria Supplier Selection using Fuzzy AHP." *Logistics Information Management* 16 (6): 382–394. doi:10.1108/09576050310503367.
- Kainz, W. 2008. *Fuzzy Logic and GIS*. Vienna: University of Vienna. http://homepage.univie.ac.at/wolfgang.kainz/Lehrveranstaltungen/ESRI_Fuzzy_Logic/File_2_Kainz_Text.pdf.
- Kavzoglu, T., E. K. Sahin, and I. Colkesen. 2014. "Landslide Susceptibility Mapping using GIS-based Multi-criteria Decision Analysis, Support Vector Machines, and Logistic Regression." *Landslides* 11 (3): 425–439. doi:10.1007/s10346-013-0391-7.
- Ke, Y., J. Im, J. Lee, H. Gong, and Y. Ryu. 2015. "Characteristics of Landsat 8 OLI-derived NDVI by Comparison with Multiple Satellite Sensors and in-Situ Observations." *Remote Sensing of Environment* 164: 298–313. doi:10.1016/j.rse.2015.04.004.
- Kirschbaum, D., T. Stanley, and S. Yatheendradas. 2015. "Modeling Landslide Susceptibility over Large Regions with Fuzzy Overlay." *Landslides* 1–12. doi:10.1007/s10346-015-0577-2.
- Laxton, S. C., and D. J. Smith. 2009. "Dendrochronological Reconstruction of Snow Avalanche Activity in the Lahul Himalaya, Northern India." *Natural Hazards* 49: 459–467. doi:10.1007/s11069-008-9288-5.
- Maggioni, M., and U. Gruber. 2003. "The Influence of Topographic Parameters on Avalanche Release Dimension and Frequency." *Cold Regions Science and Technology* 37: 407–419. doi:10.1016/S0165-232X(03)00080-6.
- McClung, D. M. 2001. "Characteristics of Terrain, Snow Supply and Forest Cover for Avalanche Initiation caused by Logging." *Annals of Glaciology* 32: 223–229.
- McClung, D. M., and A. I. Mears. 1995. "Dry-flowing Avalanche Run-up and Run-out." *Journal of Glaciology* 41: 359–372.
- McClung, D. M., and P. Schaerer. 2006. *The Avalanche Handbook*. 3rd ed. Seattle, WA: The Mountaineers Books.
- Moore, I. D., R. B. Grayson, and A. R. Landson. 1991. "Digital Terrain Modelling: A Review of Hydrological, Geomorphological, and Biological Applications." *Hydrological Processes* 5 (1): 3–30. doi:10.1002/hyp.3360050103.
- Naaim, M., T. Faung, and F. Naaim-Bouvet. 2003. "Dry Granular Flow Modelling Including Erosion and Deposition." *Surveys in Geophysics* 24 (5): 569–585. doi:10.1023/B:GEOP.0000006083.47240.4c.
- Nefeslioglu, H. A., E. A. Sezer, C. Gokceoglu, and Z. Ayas. 2013. "A Modified Analytical Hierarchy Process (M-AHP) Approach for Decision Support Systems in Natural Hazard Assessments." *Computers and Geosciences* 59: 1–8. doi:10.1016/j.cageo.2013.05.010.

- Oh, H. J., Y. S. Kim, J. K. Choi, and S. Lee. 2011. "GIS Mapping of Regional Probabilistic Groundwater Potential in the Area of Pohang City, Korea." *Journal of Hydrology* 399 (3–4): 158–172. doi:10.1016/j.jhydrol.2010.12.027.
- Oller, P., M. Janeras, H. D. Buen, G. Arno, M. Christen, C. Garcia, and P. Martinez. 2010. "Using AVAL-1D to Simulate Avalanches in the Eastern Pyrenees." *Cold Regions Science and Technology* 64: 190–198. doi:10.1016/j.coldregions.2010.08.011.
- Owen, L. A., D. I. Benn, E. Derbyshire, D. J. A. Evans, W. A. Mitchell, D. Thompson, S. Richardson, M. Lloyd, and C. Holden. 1995. "The Geomorphology and Landscape Evolution of the Lahul Himalaya, Northern India." *Z Geomorphol* 39: 145–174.
- Perla, R. 1984. "Particle Simulation of Snow Avalanche Motion." *Cold Regions Science and Technology* 9: 191–202. doi:10.1016/0165-232X(84)90066-1.
- Perla, R., T. T. Cheng, and D. M. McClung. 1980. "Two-parameter Model of Snow-avalanche Motion." *Journal of Glaciology* 94: 197–207.
- Pistocchi, A., and C. Notarnicola. 2013. "Data-driven Mapping of Avalanche Release Areas: A Case Study in South Tyrol, Italy." *Natural Hazards* 65 (3): 1313–1330. doi:10.1007/s11069-012-0410-3.
- Pourghasemi, H. R., B. Pradhan, C. Gokceoglu, and K. D. Moezzi. 2013. "A Comparative Assessment of Prediction Capabilities of Dempster-Shafer and Weights-of-evidence Models in Landslide Susceptibility Mapping using GIS." *Geomatics, Natural Hazards and Risk* 4 (2): 93–118. doi:10.1080/19475705.2012.662915.
- Pradhan, B., and S. Lee. 2010. "Delineation of Landslide Hazard Areas on Penang Island, Malaysia, by using Frequency Ratio, Logistic Regression, and Artificial Neural Network Models." *Environmental Earth Sciences* 60 (5): 1037–1054. doi:10.1007/s12665-009-0245-8.
- Saaty, T. L. 1980. *The Analytical Hierarchy Process*. New York: McGraw-Hill.
- Sailer, R., W. Felling, R. Fromm, P. Jorg, L. Rammer, P. Sampl, and A. Schaffhauser. 2008. "Snow Avalanche Mass-balance Calculation and Simulation-model Verification." *Annals of Glaciology* 48: 183–192. doi:10.3189/172756408784700707.
- Sailer, R. L., L. Rammer, and P. Sample. 2002. "Recalculation of an Artificially Released Avalanche with SAMOS and Validation with Measurements from a Pulsed Doppler Radar." *Natural Hazards and Earth System Sciences* 2: 211–216.
- Salm, B. 1993. "Flow, Flow Transition and Runout Distances of Flowing Avalanches." *Annals of Glaciology* 18: 221–226.
- Sappington, J. M., K. M. Longshore, and D. B. Thompson. 2007. "Quantifying Landscape Ruggedness for Animal Habitat Analysis: A Case Study using Bighorn Sheep in the Mojave Desert." *The Journal of Wildlife Management* 71 (5): 1419–1426. doi:10.2193/2005-723.
- Schweizer, J., J. B. Jamieson, and M. Schneebeli. 2003. "Snow Avalanche Formation." *Review of Geophysics* 41: 10–16. doi:10.1029/2002RG000123.
- Schweizer, J., C. Mitterer, and L. Stoffel. 2009. "On Forecasting Large and Infrequent Snow Avalanches." *Cold Regions Science and Technology* 59 (2–3): 234–241. doi:10.1016/j.coldregions.2009.01.006.
- Selçuk, L. 2013. "An Avalanche Hazard Model for Bitlis Province, Turkey, using GIS Based Multicriteria Decision Analysis." *Turkish Journal of Earth Sciences* 523–535. doi:10.3906/yer-1201-10.
- Shahabi, H., M. Hashim, and B. B. Ahmad. 2015. "Remote Sensing and GIS-based Landslide Susceptibility Mapping using Frequency Ratio, Logistic Regression, and Fuzzy Logic Methods at the Central Zab Basin, Iran." *Environmental Earth Sciences* 73 (12): 8647–8668. doi:10.1016/j.catena.2013.11.014.
- Sharma, S. S., P. Mathur, and Snehmami. 2004. "Change Detection Analysis of Avalanche Snow in Himalayan Region using Near Infrared and Active Microwave Images." *Advances in Space Research* 33: 259–267. doi:10.1016/S0273-1177(03)00472-1.
- Singh, M. K., R. D. Gupta, Snehmami, A. Bhardwaj, and A. Ganju. 2016. "Scenario-based Validation of Moderate Resolution DEMs Freely Available for Complex Himalayan Terrain." *Pure and Applied Geophysics* 173 (2): 463–485. doi:10.1007/s00024-015-1119-5.
- Singh, M. K., R. D. Gupta, Snehmami, S. Kumar, A. Ganju. 2015. "Assessment of Freely Available CartoDM V1 and V1.1R1 with respect to High Resolution Aerial Photogrammetric DEM in High Mountains." *Geocarto International*. doi:10.1080/10106049.2015.1094524.
- Sinickas, A., and B. Jamieson. 2014. "Comparing Methods for Estimating β Points for use in Statistical Snow Avalanche Runout Models." *Cold Regions Science and Technology* 104–105: 23–32. doi:10.1016/j.coldregions.2014.04.004.
- Smith, M. J., and D. M. McClung. 1997. "Avalanche Frequency and Terrain Characteristics at Rogers' Pass, British Columbia, Canada." *Journal of glaciology* 43: 165–171.
- Snehmami, A. Bhardwaj, A. Pandit, and A. Ganju. 2014. "Demarcation of Potential Avalanche Sites using Remote Sensing and Ground Observations: A Case Study of Gangotri Glacier." *Geocarto International* 29 (5): 520–535. doi:10.1080/10106049.2013.807304.
- Zadeh, L. A. 1965. "Fuzzy Sets." *Information and Control* 8 (3): 338–353. doi:10.1016/S0019-9958(65)90241-X.
- Zischg, A., S. Fuchs, M. Keiler, and G. Meißl. 2005. "Modelling the System Behaviour of Wet Snow Avalanches using an Expert System Approach for Risk Management on High Alpine Traffic Roads." *Natural Hazards and Earth System Sciences* 5: 821–832. SRef-ID: 1684-9981/nhess/2005-5-821.

Numerical Study of the Cavity Geometry on Supersonic Combustion with Transverse Fuel Injection

M.F. Khan, R. Yadav, Z.A. Quadri and S.F. Anwar

Abstract The small residence time, in designing the engine of a supersonic aircraft, presumed to have a very important character in combustion. At hypersonic flight, the flow is supersonic while entering in the combustor to avoid excessive heating and fuel is essentially to be injected, mixed and combusted entirely within a short residence time of the order of millisecond. In order to resolve the restrictions given by short residence time, numerous studies have been carried out to suggest the concepts of injection, among which, the transverse fuel injection in a combustor with a cavity is being used in several engines. This paper describes the numerical study of the combustion enhancement with hydrogen fuel injection in a transverse aperture nozzle into a supersonic hot air stream. Several cavities with single and dual steps with different cavity wall angles are analyzed. Eddy dissipation concept model with detailed hydrogen-air combustion with 21 reactions and 9 species transport has been applied to numerically simulate the reacting flow of hydrogen fuel scramjet combustor. The mixing enhancement, static pressure and combustion efficiencies are analyzed for different cavity geometries.

Keywords Cavity flame holder · Hydrogen injection · EDC model · Scramjet combustor

1 Introduction

The supersonic combustion is a complex phenomenon, which consists of shock interaction, heat released and turbulent mixing in the supersonic airflow. As the residence time of the incoming airflow in the combustor is extremely short i.e. of

M.F. Khan (✉) · R. Yadav · Z.A. Quadri
Department of Aerospace Engineering, University of Petroleum & Energy Studies,
Dehradun, India
e-mail: faisalmechster@gmail.com; faisal.khan@stu.ddn.ac.in

S.F. Anwar
Department of Mechanical Engineering, Aligarh Muslim University, Aligarh, India

the order of milliseconds [1], the combustor equipped with such design features, shall enhance adequate mixing of the air and fuel so as to promote the preferred chemical transport reaction and hence heat released can take place in the permissible limitations of residence time of the air-fuel mixture. For promoting this procedure, it needs a strong interpretation of fuel injection process which governs the combustion and supersonic mixing of air-fuel along with the elements [2], which influences the losses inside the combustion chamber. The mixing process inside the combustor cannot be independent from the effect of heat released in a practical flow field of interest in the scramjet and chemical composition deviates because the factors responsible for mixing, including transport properties and density are directly governed by chemical and thermal changes [3]. Many strategies have been proposed [4–6] to improve rapid air-fuel mixing. Apart from these, the generation of axial vortices was suggested as a mean of better far-field mixing characteristic as being propagated to a considerable distance, with the suppressing characteristics of the supersonic flow [7, 8]. As a result of vortex generation, the mixing is enhanced at micro and macro levels by allowing larger amount of air into fuel and increasing the fuel-air interface area respectively [9]. However, using the cavity inside the combustor is one of the simplest methods of vortex generation while causing a recirculation region behind the step with hot gases, acts as a permanent source of ignition [7, 10]. A new model to forecast fuel-air mixing length in the vicinity of the scramjet combustor was derived by Doolan et al. [11] and investigated that the combustion efficiency of supersonic flames is sturdily affected by the existence of shock waves.

2 Numerical Modeling

2.1 Governing Equations

The governing equation consists of equation of continuity, momentum, energy and chemical species. The conservation form of the continuity equation in reactive flow.

For individual species is written as follows:

$$\rho \frac{\partial Y_k}{\partial t} + \rho(\mathbf{v} \cdot \nabla)Y_k - \nabla \cdot (\rho D_{mk} \nabla Y_k) = \omega_k \quad (1)$$

where Y is the mass fraction of k th species, \mathbf{v} is the mixture velocity, D_m is the mass diffusion constant and ω is the reaction rate.

The momentum equation may be written in the conservation form as

$$\frac{\partial}{\partial t} (\rho v_j) + \frac{\partial}{\partial x_i} (\rho v_i v_j) + \frac{\partial}{\partial x_i} \left[p \delta_{ij} - \mu \left(\frac{\partial v_i}{\partial x_j} + \frac{1}{3} \frac{\partial v_j}{\partial x_i} \right) \right] = \rho \sum_{k=1}^N Y_k f_{kj} \quad (2)$$

where p is the pressure, μ is the dynamic viscosity and f is the mixture fraction.

The energy conservation equation takes the following form

$$\frac{\partial}{\partial t}(\rho E) + \frac{\partial}{\partial x_i}(\rho E v_i + p v_i) - \frac{\partial}{\partial x_i} \left(k T_i + \sum_k \rho H D_{mk} Y_{ki} + \tau_{ij} v_j \right) = S - \sum_k H_k^0 \omega_k \quad (3)$$

where E is the total energy, H is the enthalpy, H_k^0 is enthalpy of formation and S is the source term i.e. energy due to body forces and can be expressed by

$$S = \rho \sum_{k=1}^N Y_k f_k \cdot v \quad (4)$$

It indicates the existence of N species equations having conservation of mass, momentum and energy for the individual species, where only $N-1$ equations of Y_k species are independent.

The viscous stress tensor τ_{ij} can be written as follows

$$\tau_{ij} = \mu \left(\frac{\partial v_i}{\partial x_j} \frac{\partial v_j}{\partial x_i} - \frac{2}{3} \frac{\partial v_k}{\partial x_k} \delta_{ij} \right) \quad (5)$$

The equation of state for calculating pressure is given by

$$p = \rho R^0 T \sum_{k=1}^N \frac{Y_k}{W_k} \quad (6)$$

where R^0 is the universal gas constant (8.3143 J/g-mol K), W_k being molecular weight. The mixture enthalpy is described by

$$H_k = H_k^0 + \int_{T_0}^T c_{pk} dT \quad (7)$$

where H_k^0 is the enthalpy of formation and T_0 is the reference temperature for thermodynamic calculations. A general form of specific heat for k species is given by

$$c_{pk} = A_k + B_k T + C_k T^2 + D_k T^3 + E_k T^4 \quad (8)$$

The turbulent viscosity is calculated by incorporating 2 equations, shear stress transport k - ω model by taking the transport of principal turbulent shear stress into account with wall Prandtl number as 0.85 and Turbulent Schmidt number as 0.7.

Table 1 Hydrogen-air combustion model [13]

	Reactions	A_i	α	E_i
1	$H + O_2 = O + OH$	$1.04E + 14$	0.00	$1.531E + 04$
2	$O + H_2 = H + OH$	$3.82E + 12$	0.00	$7.948E + 04$
3	$H_2 + OH = H_2O + H$	$2.16E + 08$	1.51	$3.430E + 03$
4	$OH + OH = O + H_2O$	$3.34E + 04$	2.42	$-1.930E + 03$
5	$H_2 + M = H + H + M$	$4.58E + 19$	-1.40	$1.040E + 05$
6	$O + O + M = O_2 + M$	$6.16E + 15$	-0.50	$0.000E + 00$
7	$O + H + M = OH + M$	$4.71E + 18$	-1.00	$0.000E + 00$
8	$H_2O + M = H + OH + M$	$6.06E + 27$	-3.32	$1.208E + 05$
9	$H + O_2(M) = HO_2(+M)$	$4.65E + 12$	0.44	$0.000E + 00$
10	$HO_2 + H = H_2 + O_2$	$2.75E + 06$	2.09	$-1.451E + 03$
11	$HO_2 + H = OH + OH$	$7.08E + 13$	0.00	$2.950E + 02$
12	$HO_2 + O = O_2 + OH$	$2.85E + 10$	1.00	$-7.239E + 02$
13	$HO_2 + OH = H_2O + O_2$	$2.89E + 13$	0.00	$-4.970E + 02$
14	$HO_2 + HO_2 = H_2O_2 + O_2$	$4.20E + 14$	0.00	$1.200E + 04$
15	$HO_2 + HO_2 = H_2O_2 + O_2$	$1.30E + 11$	0.00	$-1.630E + 03$
16	$H_2O_2(+M) = OH + OH(+M)$	$2.00E + 12$	0.90	$4.875E + 04$
17	$H_2O_2 + H = H_2O + OH$	$2.41E + 13$	0.00	$3.970E + 03$
18	$H_2O_2 + H = HO_2 + H_2$	$4.82E + 13$	0.00	$7.950E + 03$
19	$H_2O_2 + O = OH + HO_2$	$9.55E + 06$	2.00	$3.970E + 03$
20	$H_2O_2 + OH = HO_2 + HO_2$	$1.74E + 12$	0.00	$3.180E + 02$
21	$H_2O_2 + OH = HO_2 + H_2O$	$7.59E + 13$	0.00	$7.270E + 03$

M = Third body

2.2 Combustion Modeling

Generalized finite rate chemistry model with 21 reactions and 9 species (N_2 , H_2 , H , O_2 , O , OH , H_2O , HO_2 and H_2O_2) transport is applied, incorporating eddy dissipation concept (EDC) model (Table 1).

Chemical reactions for the present combustion model are characterized by the chemical reaction equation of the following form:

$$\sum_{k=1}^N v'_{ki} M_k \xrightleftharpoons[k_b]{k_f} \sum_{k=1}^N v''_{ki} M_k \quad (i = 1, 2, 3, \dots, M) \quad (9)$$

where v_{ki} is the stoichiometric coefficient of the species k for the reaction step i , with the prime and double prime representing the reactant and product, respectively. M_k is the chemical symbol for the species k and k_f and k_b denote the specific reaction rate constants for the forward and backward reactions respectively. These reactions are governed by the law of mass action related by reaction rate ω_k

$$\omega_k = W_k \sum_{i=1}^M (v''_{ki} - v'_{ki}) \left[k_{fi} \prod_{j=1}^N C_j^{v'_{ij}} - k_{bi} \prod_{j=1}^N C_j^{v''_{ij}} \right] \quad (10)$$

where C_j is the molar concentration. The Arrhenius expression for forward and backward reaction rate is given by

$$k_{fi} = A_{fi} T^{\alpha_{fi}} \exp\left(-\frac{E_{fi}}{R^0 T}\right) \quad (11)$$

$$k_{bi} = A_{bi} T^{\alpha_{bi}} \exp\left(-\frac{E_{bi}}{R^0 T}\right) \quad (12)$$

Here A_i is the pre exponent factor, α is the temperature constant, E_i is the activation energy and R^0 is the universal gas constant.

2.3 Solution Strategy and Method Validation

The governing equations are solved with commercial code Ansys Fluent 14.5, using a complete structured mesh generation by ICEM CFD. Density based solver with absolute velocity formulation, cell based Green-Gauss theorem is applied for spatial discretization and flux is calculated using Roe-FDS scheme.

The numerical method is validated with the previous experimental work done for two cases, namely cold flow and reacting flow condition. Figure 1a shows a free

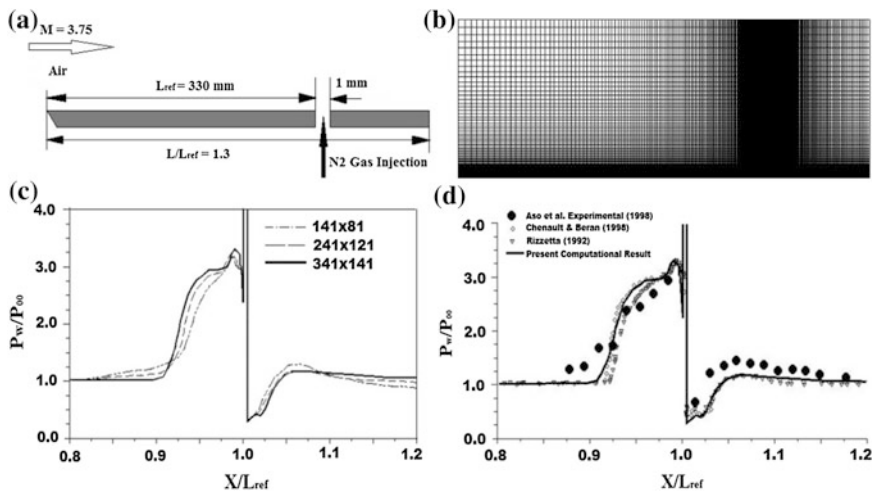


Fig. 1 Normalized static pressure plot for injection on the flat plate. **a** Flat plate geometry, **b** Grid generation (341 × 141), **c** Grid independency plot, **d** Normalized static pressure plot

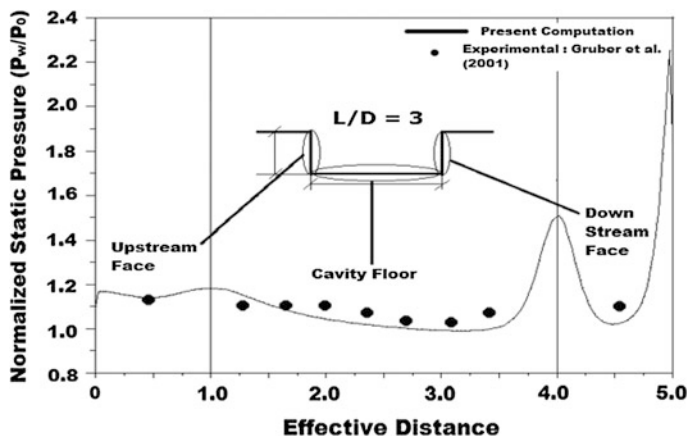


Fig. 2 Normalized static pressure distribution for cavity $L/D = 3$

stream air flows over flat plate with initial conditions of having free stream Mach no 3.75, total pressure of 1.2 MPa and total temperature of 299 K. Nitrogen gas is injected through a slot nozzle at sonic velocity and at a total pressure of $p_j = 10.29p_\infty$. The wall pressure distribution is plotted against the dimensionless distance as shown in Fig. 1d, which is compared with the experimental results of Aso et al. [14] and the results of previous work done by Rizzetta et al. and Chenault et al. [15, 16].

In the second validation, we have considered the configuration of $L/D = 3$ from the experimental work done by Gruber et al. [12]. The initial boundary conditions of free stream stagnation pressure as 690 kPa and free stream stagnation temperature as 300 K are used. The wall static pressure distribution is plotted against the effective distance in Fig. 2, which is closer to the experimental results.

3 Results

3.1 Simulated Results

Figure 3 shows the cavity geometries with single and double angled step at the cavity walls for $L/D = 3$ with flow conditions, for $D = 15$ mm. The free stream conditions at inlet are considered as Mach 2.5, static temperature 1000 K, static pressure 1 atm, mass fraction of O_2 0.2329, mass fraction of N_2 0.7671 and at the Hydrogen injection as Mach 1.0, stagnation pressure 7.57 atm and stagnation temperature 600 K.

Figure 4 shows streamline pattern and shock wave interaction for the case without cavity. A recirculation zone is formed upstream of the injection point Fig. 4a, which is a low pressure region (barrel shock) and helps in partial air-fuel

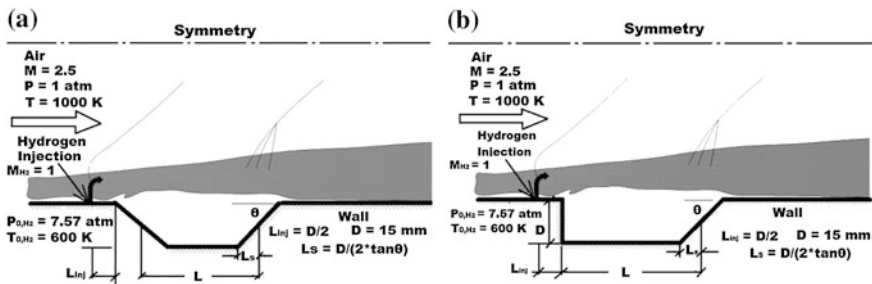


Fig. 3 Geometry. **a** Double step angled cavity, **b** Single step angled cavity

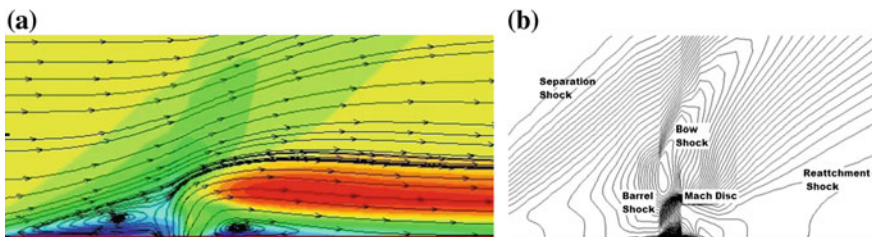


Fig. 4 Present results without cavity. **a** Streamlines, **b** Pressure contours

mixing and thus enhancing the combustion efficiency. There exists an expansion fan around the leading edge of the jet exit and a reattachment shock downstream due to the shear layer growth Fig. 5a. Due to heat released in chemical reaction, the gas temperature approaches 1900 K Fig. 5b, as adiabatic wall condition is imposed to the walls. When the strong shock impinges on the boundary layers, flow separation occurs over the entire span of cavity length as shown in Fig. 5c, d and the flow reattached at the rear wall of the cavity resulting a strong trailing edge shock wave.

It can be seen that this shock is stronger in case of single 30° rear wall angle cavity as compared to double step cavity Fig. 5c, d. As the injected fuel travels over the cavity length, the trailing edge shock interfaces with the fuel jet within the shear layer significantly affect the chemical reaction to cause shock induced combustion. The high temperature gas inside the cavity acts as a flame-holder and enhances the combustion efficiency.

In order to evaluate the completeness of the combustion with and without cavities, the consumption of fuel injected upstream at different given location of x has to be calculated and further combustion efficiency can be defined as

$$\eta_c = \frac{m_{fuel,in} - m_{fuel,x}}{m_{fuel,in}} = 1 - \frac{\int \alpha \rho u dA}{\sum m_{fuel,in}} \quad (13)$$

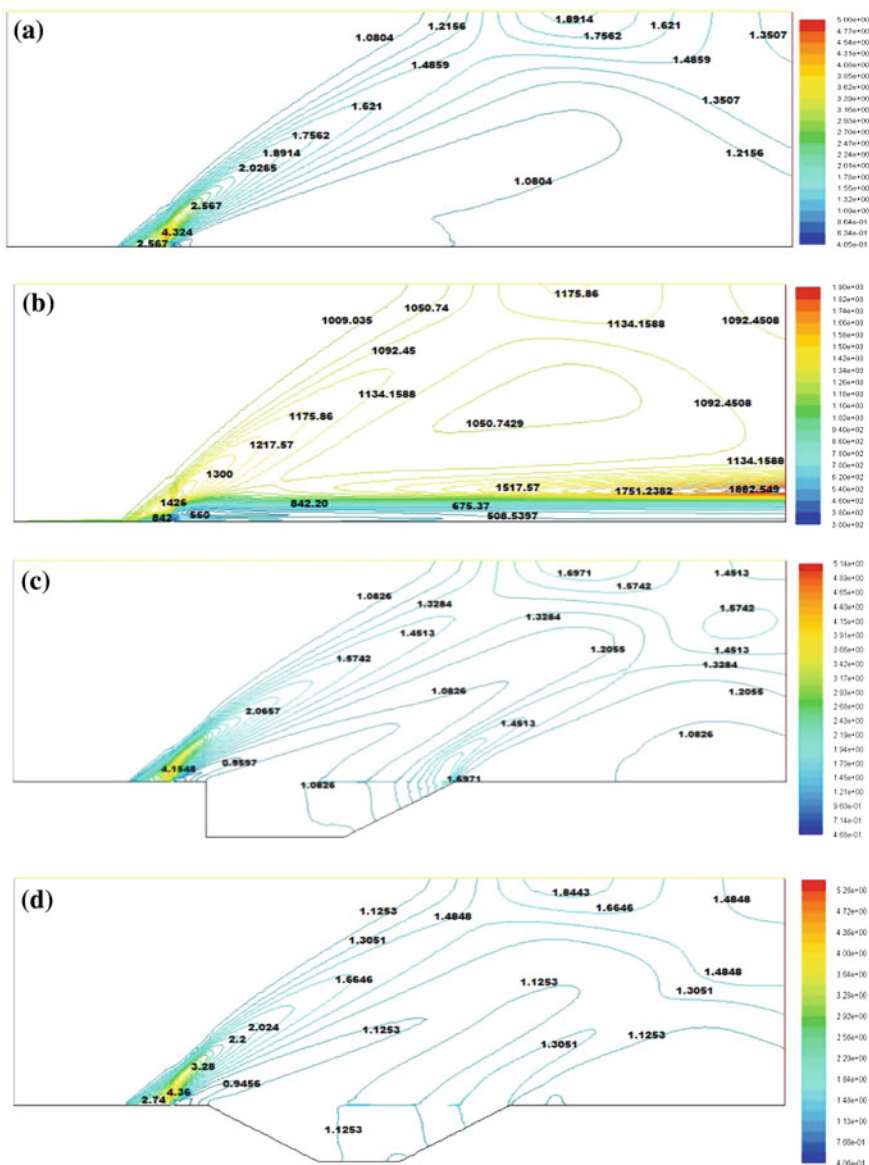


Fig. 5 **a** Pressure contour without cavity, **b** Temperature contour without cavity, **c** Pressure contour single step cavity, **d** Pressure contour double step cavity

where $m_{fuel,in}$ is the mass of the fuel consumed in the chemical reaction. The mass fraction of H_2 in Fig. 6 reflects that the least amount of unburnt H_2 at the end of the chemical reaction is found to be in case of single step cavity which is 0.125 and is 0.625 in case without cavity. Based on the Eq. (13), a graph of combustion

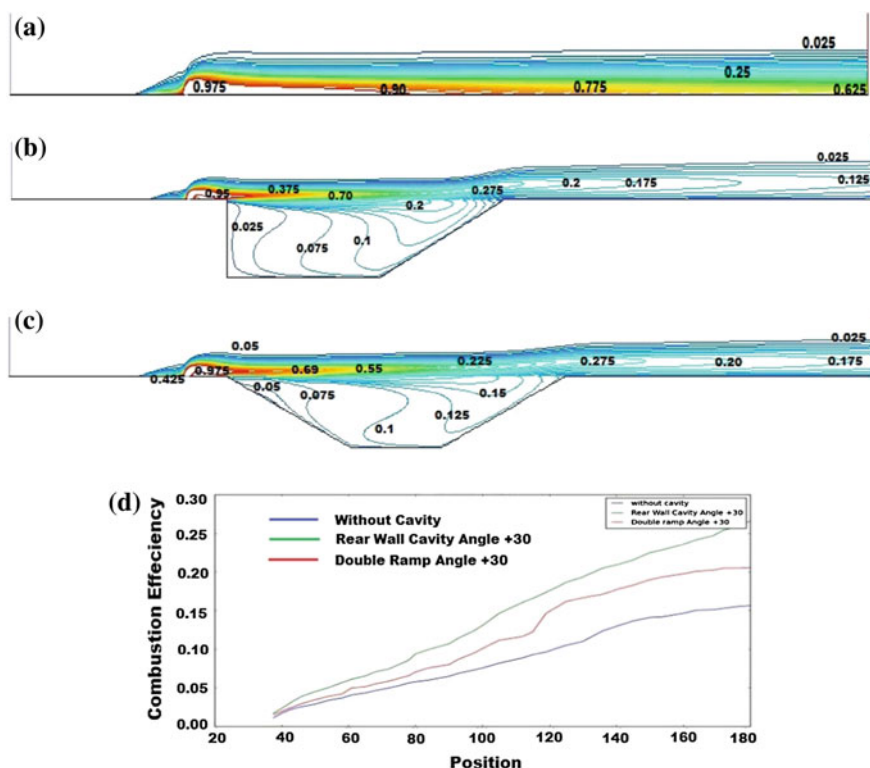


Fig. 6 Mass fraction of H_2 **a** Without cavity, **a** $+30^\circ$ Cavity, **c** Double step $+30^\circ$ cavity
d Combustion efficiency comparison for cavity angles

efficiency is plotted Fig. 6d. It can be observed that introducing a cavity with double angle at the front and rear wall does not have a considerable effect as single rear wall angle cavity has.

4 Conclusions

The effect of single and dual step cavities are analyzed and compared in the performance without cavity case, considering the physics of air-fuel mixing enhancement, flame holding phenomenon and shock induced combustion for a supersonic airflow. Single step cavity for the same angle is more efficient as compared with the other two cases. Cavity depth has a significant effect on the residence time, hence the cavity dimensions should be chosen, considering the flame holding and auto ignition.

References

1. Huang, W., Luo, S.-B., Pourkshnian, M., Lin, M.A., Derek, B.I., Jun L., Wang, Z.: Numerical simulations of a typical hydrogen fueled scramjet combustor with a cavity flameholder. In: Proceedings of the World Congress on Engineering, vol. 2. London, UK, 30 June–2 July
2. Ladeinde, F.: A Critical Review of Scramjet Combustion Simulation (Invited), 47th AIAA Aerospace Science Meeting, 5–8 January. Oriando, Florida (2009)
3. Segel, C.: The Scramjet Engine, Cambridge University Press (2009)
4. Fuller, R.P., Wu, P.K., Nejad, A.S., Schetz, J.A.: Comparison of physical and aerodynamic ramps as fuel injectors in supersonic flow. *J. Prop. Power*, vol. 14, pp. 135–145 (1998)
5. Riggins, D.W., McClinton C.R.: A computational investigation of mixing and reacting flows in supersonic combustors. AIAA 92–0626 (1992)
6. Liskinsky, D.S., True B., Holdeman, M.A.: Crossflow mixing of non-circular jet. AIAA 95–0732 (1995)
7. Kim, K.M., Beak, S.W., Han, C.Y.: Numerical study on supersonic combustion with cavity based fuel injection. *Int. J. Heat Mass Trans.* **47**, 271–286 (2004)
8. Gruber, M.R., Nejad, A.S., Chen, T.H., Dutton, C.J.: Mixing and penetration studies of sonic jets in a mach 2 freestream. *J. Prop. Power* **11**(2), 315–323 (1995)
9. Yakar, A.B., Hanson, R.K.: Cavity flame-holders for ignition and flame stabilization in scramjets: an overview. *J. Prop. Power* **17**, 869–877 (2001)
10. Yakar, A.B., Mungal, M.G., Hanson, R.K.: Time evolution and mixing characteristics of hydrogen and ethylene transverse jets in supersonic crossflow. *Phys. Fluid* **18**(2), 026101 (2006)
11. Doolan, C.J.: A Supersonic Combustion Model for Scramjet Vehicle Performance, 5th Asia Pacific Conference on Combustion, Adelaide, Australia, 18–20 July 2005
12. Gruber, M.R., Baurle, R.A., Mathur, T., Hsu, K.Y.: Fundamental studies of cavity based flame-holder concepts for supersonic combustors. *J. Prop. Power* **17**(1) (2001)
13. Burke, M.P., Chaos, M., Ju, Y., Dryer, F.L., Klippenstein, S.J.: Comprehensive H₂/O₂ kinetic model for high pressure combustion. *Int. J. Chem. Kinet.* (2011)
14. Aso, S.: Experimental study on mixing phenomena in supersonic flows with slot injection. AIAA-91-0016 (1991)
15. Rizzetta, D.P.: Numerical simulation of slot injection into a turbulent supersonic stream. AIAA **30**, 2434–2439 (1992)
16. Chenault, C.F., Beran, P.S.: k- ϵ and reynolds stress turbulence model comparisons for two-dimensional injection flows. AIAA **36**, 1401–1412 (1998)

Plasmonic Ladder-Like Structure for NIR Polarizer

P. Mandal¹

Received: 19 February 2016 / Accepted: 9 May 2016
© Springer Science+Business Media New York 2016

Abstract Polarization-dependent light transmission property is investigated in two-dimensional plasmonic ladder-like structure in the Near-infrared (NIR) regime of 900 to 1600 nm. The plasmonic ladder-like structures are fabricated using cost-effective laser interference lithography. Optical transmission studies reveal that in the stated NIR regime, the structure has nearly 30 % absolute transmission with respect to air when the long axis is aligned parallel to the polarization axis of the incident excitation and has negligible transmission at the crossed polarization state. The findings have potential implications in designing large area flat NIR polarizers.

Keywords Plasmonic ladder-like structures · Near-infrared frequency · Plasmonic polarizer · Optical transmission · Laser interference lithography

Introduction

Subwavelength metal-dielectric plasmonic structures (MDPSs) have been very interesting, and promising optical systems where light-matter interaction at such length scale gives rise to novel optical phenomena. MDPSs exhibit enhanced light transmission [1–5] due to surface plasmon reso-

nant excitation, size and shape-dependent resonant reflection and transmission [6, 7], and polarization-controlled color tunability and filtering. The structures have also been investigated for wavelength selecting switching applications [7–10] and for designing various plasmonic and photonic devices in the visible and NIR regimes [11–18]. Fundamentally, it is interesting to see how a simple plasmonic system consisting of circular elements, elliptic holes, slits, different anisotropic structures of varied shapes, such as, G [19, 20], H [21–23], L [24, 25], T [26–28] etc. can be used to manipulate optical properties in the spectral domain of interests.

Polarization state of output radiation gets affected when an incident radiation interacts with an anisotropic structure. Structural anisotropy can easily be achieved by introducing shape anisotropy in the fabricated plasmonic structures. Metal-dielectric gratings patterned in two-dimensional periodic arrays of elongated near-elliptic elements are promising anisotropic media [29]. Light transmission through such type of plasmonic media greatly depends on the polarization state of the input radiation which enables the surface plasmon polariton excitation [7, 9, 30]. Therefore, there is a great possibility that the structural anisotropy can be useful for optical manipulation [20]. Furthermore, planar structures have added advantage of designing compact and slim optical devices [31–38]. To this context, fabrication of low cost broadband NIR polarizers with wide active area is challenging. The present report demonstrates fabrication of a broadband NIR plasmonic polarizer using simple cost-effective laser interference lithography (LIL) technique, and its polarization dependent optical transmission is studied in the NIR spectral domain. The experimental results show that the fabricated ladder-like plasmonic structures can act as flat polarizers in the wavelength range from 900 to 1400 nm. The parallel-to-crossed (x-pol to y-pol) polarized transmission ratio is observed to be about 8:1.

✉ P. Mandal
pmandal@ddn.upes.ac.in

¹ Department of Physics, University of Petroleum and Energy Studies, Dehradun 248007, India

Experimental

Plasmonic NIR polarizers were fabricated as follows [29]: positive tone photoresist (maP-1205 from micro-resist technology, Germany) was spin coated on to cleaned glass substrates (10 mm × 10 mm × 1 mm) at 4000 rpm (revolution per minute) for 45 s. The photoresist-coated substrates were pre-baked at 80 °C for 1 min. Sustained interference patterns were obtained using a coherent laser source (He–Cd Kimmon blue laser: wavelength 442 nm, laser power of 30 mW). Laser beam was focused to an aperture (20 μm) by a 20× objective. The expanded beam was then split into two beams using 50:50 beam splitter. The component beams were then adjusted to overlap onto each other using mirrors. The angle between the interfering beams was adjusted to get desired periodic structure. The pre-baked sample was mounted on to sample stage holder and exposed to the interference pattern for 1 min. After first exposure, the sample stage was rotated by 15°, and a second exposure was performed for another 1 min. Finally, the exposed samples were developed using standard developer (maD335). The development of the exposed film resulted in periodically patterned photoresist. A thin layer of gold (20 nm) was thermally evaporated onto these dielectric patterned structures to obtain final metal-dielectric plasmonic ladder-like structures. The surface topography was obtained using a scanning electron microscope (ZEISS, EVO-18), and typical topological images are shown in shown in Fig. 1. Two-dimensional (2D) ladder-like structures have period of about 6400 nm along major axis (long axis, say, x-axis) and of about 850 nm along minor axis (short axis, say, y-axis). The typical modulation depth is of about 200 nm. The strip length (x-

direction) is about 2600 nm, and gap between two strips is about 600 nm. The top gold acts like a modulated surface.

Broadband zeroth-order optical transmission spectra in the NIR regime of 900 to 1600 nm were recorded using a BX51 microscope through 100× objective. The transmitted signal was fed to the optical fiber fitted at the trinocular. The other end of the fiber was connected to the CCD (charged coupled devices) spectrometer (BWTEK NIR spectrometer: Model-Sol1.7, TE: BTC261P). NIR polarizer was used to adjust the polarization states of input excitation beam. The sample orientation was kept at a fixed orientation (say short axis vertical), and the input polarization axis was rotated to align it with respect to sample axis to acquire polarization dependent transmission spectra.

Results and Discussions

Incident beam polarization dependent broadband transmitted spectra (0th order) of the fabricated sample are shown in Fig. 2. Polarization angle of excitation beam varied from 0° to 90° (y-pol to x-pol) with an angle step of 15°. The spectra show a systematic change in the transmitted intensity with the incident beam polarization angle θ (with respect to the minor axis of the sample structure, black arrows in Fig. 2). The spectra show highest transmitted intensity ~30 % for $\theta = 90^\circ$ (electric field is orientated along the major axis, x-pol) and about 4 % for $\theta = 0^\circ$ (electric field is along the minor axis, y-pol) in the wavelength range of about 900 to 1400 nm. The ratio of the transmitted intensity for parallel (x-pol) to perpendicular (y-pol) polarization is about 8:1 in the stated spectral

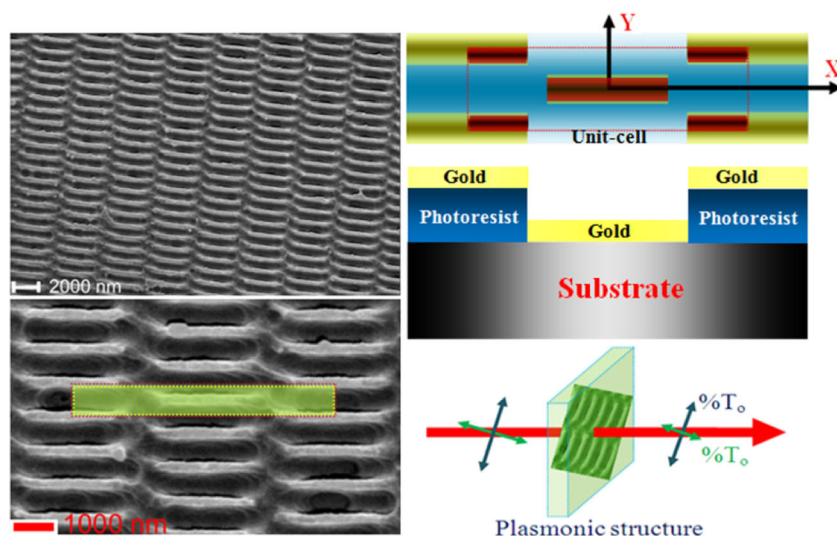


Fig. 1 Typical scanning electron microscope images of 2D plasmonic ladder-like structure (complementary layer pair) of period 850 nm along y-axis and 6400 nm along x-axis at different magnification. The yellow-colored shaded region (left column: bottom image) indicates a unit cell

which is schematically shown in the right column: top figure, for clarity. The two-dimensional unit cell cut along the y-axis is shown in middle figure (of right column). Bottom figure (of right column) represents the schematic view of the structure acting as polarizer (for better visibility)

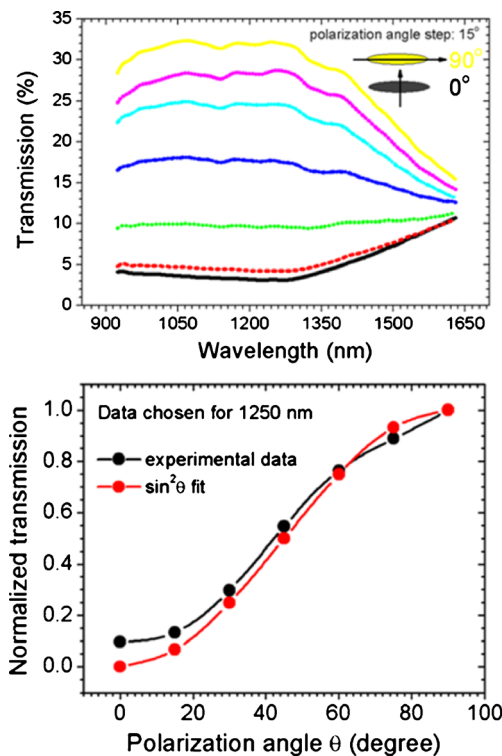


Fig. 2 Broadband NIR transmission spectra of composite plasmonic structure at various polarization state of input beam in the range of 900 to 1600 nm, showing a change in the absolute transmission from around 30 to 4 % in the case of parallel to perpendicular polarizations (x-pol to y-pol), respectively. The polarization angle dependent normalized transmitted intensity plot (for 1200 nm) and corresponding $\sin^2\theta$ fit are shown in *bottom* figure

range. The large band width from 900 to 1400 nm has reasonably large transmission ratio and would be suitable in designing NIR polarizer. The polarization dependent transmission plot (bottom figure of Fig. 2) follows $\sin^2\theta$ fit (transmission values picked at 1200 nm for the plot). The $\sin^2\theta$ fit is applied because of dual modes co-existing in the resonance phenomena, and in this case the transmission intensity will follow Malus' law [39] according to

$$T(\theta, \lambda) = T_{\perp}(\lambda)\sin^2(\theta) + T_{\parallel}(\lambda)\cos^2(\theta) \quad (1)$$

where, $T_{\perp}(\lambda)$ ($\theta=0^\circ$) is the transmission corresponds to resonance at 900 to 1400 nm and $T_{\parallel}(\lambda)$ ($\theta=90^\circ$) is that beyond 1600 nm. The observed low transmission value in the frequency band from 900 to 1400 nm arises due to plasmon excitation (caused by Bragg scattering along the short axis) of the ladder-like structure. Excitation due to Bragg scattering along the long axis is expected to fall beyond 1600 nm, therefore, the energy corresponding to 900 to 1400 nm will not be absorbed.

The effect of structural anisotropy on the light transmission is investigated further by investigating the light transmission through a more symmetric structure, such as, square lattice having 4-fold symmetry. The transmission spectra in the case

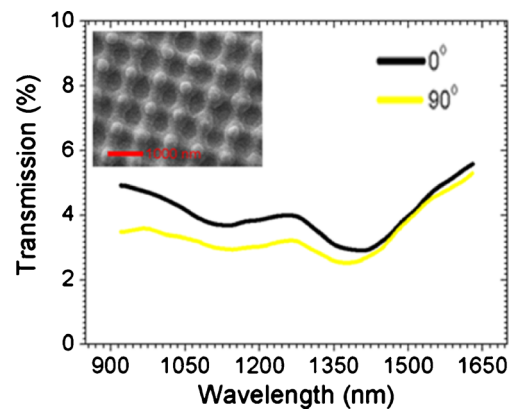


Fig. 3 Observed transmission spectra of plasmonic square lattice structure at two crossed polarization states (x-pol: 90° and y-pol: 0°), showing almost no change in the spectral transmission. Inset shows the SEM image of the fabricated square lattice structure. Scale bar represents 1000 nm

of parallel and perpendicular polarizations are shown in Fig. 3. Almost no change in the transmitted spectra is observed when the polarization angle changes from 90° deg to 0° deg (x-pol to y-pol). The overall transmission is only about 5 % in both the cases. It is noteworthy that the dimension of the plasmonic particles becomes smaller, and the period in either direction is about 850 nm. Bragg scattering in reciprocal space can provide the additional momentum required for the plasmonic excitation, which can lead to higher value of absorption or dissipation within the structure, as explained above.

Numerical Modeling

For in-depth understanding of the optical transmission at crossed polarization states, the plasmonic ladder-like structure is computed using finite difference time domain simulation (FDTD) method (OPTIWAVE). The structure is designed as follows: on a 200-nm thick SiO_2 supportive layer ladder-like photoresist pattern. The resist pattern is considered to have a height of 200 nm to match experimentally measured value. On the top of a 20-nm thin gold layer, which results in a complementary layer pair (not simple slit structure or metal bars), similar to as shown in Fig. 1. The supportive SiO_2 layer has constant refractive index 1.5. Lorentz–Drude dispersion is considered for the gold layer. The periodicities of the ladder-like structure along the x- and y-directions are 6400 and 850 nm, respectively. Periodic boundary conditions (PBC) are applied along x- and y-directions, whereas anisotropic perfectly matched layers (APML) is applied along z-direction. A linearly polarized light (x-pol or y-pol) is allowed to fall normally onto the structure, and the output reflection ($R(\lambda)$), transmission ($T(\lambda)$), and absorption ($A(\lambda)$) spectra are obtained.

Simulated transmission, reflection, and absorption spectra for x-pol and y-pol are plotted in Fig. 4. Transmission

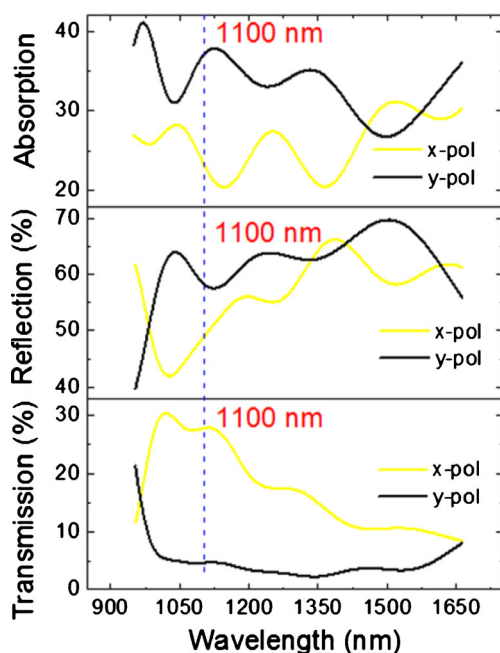


Fig. 4 Simulated transmission (*bottom row*), reflection (*middle row*), and absorption (*top row*) spectra of plasmonic ladder-like structure for two crossed polarization states. *Blue dotted vertical line* represents the selected wavelength at which the structure is simulated in order to analyze near-field distribution at the resonating wavelength

spectrum for x-pol shows broad transmission in the range of 975 to 1500 nm, showing transmission of about 30 % between 1000 to 1200 nm. For y-pol, the transmission is within 5 % in the stated range. This is consistent with the experimental observations. Reflection and absorption spectra show, however, opposite behavior. For the y-pol case, the reflection and absorption are seen to be higher compared to the corresponding reflection and absorption for x-pol excitation. Higher absorption indicates the plasmon excitation for the y-pol incident light. Plasmon excitation along the short axis is possible due to grating coupling which provides the additional momentum. The detailed discussion on the plasmonic excitation in this type of ladder-like complementary structure is reported in the article previously published by our group [29].

The electromagnetic near-field (E field) distribution obtained by simulating the ladder-like structure at a resonating wavelength 1100 nm (selected on the basis of Fig. 4 transmission spectra) is shown in Fig. 5. The field distribution shown here is just taken on the top of the structure surface. The left and right columns show the electric field components E_x , E_y , and E_z , respectively, for the x-pol and y-pol excitations. The E_x field distribution (x-pol) shows the concentration of near-field at the air-gap region. The E_y and E_z fields are weakly confined at the metallic edges. For the y-pol case, the E_x field distribution shows weak confinement at the metallic corners. However, E_y field distribution shows a different behavior. A very strong confinement of the near-field is observed at the metallic edges. The E_z fields for both the cases show similar

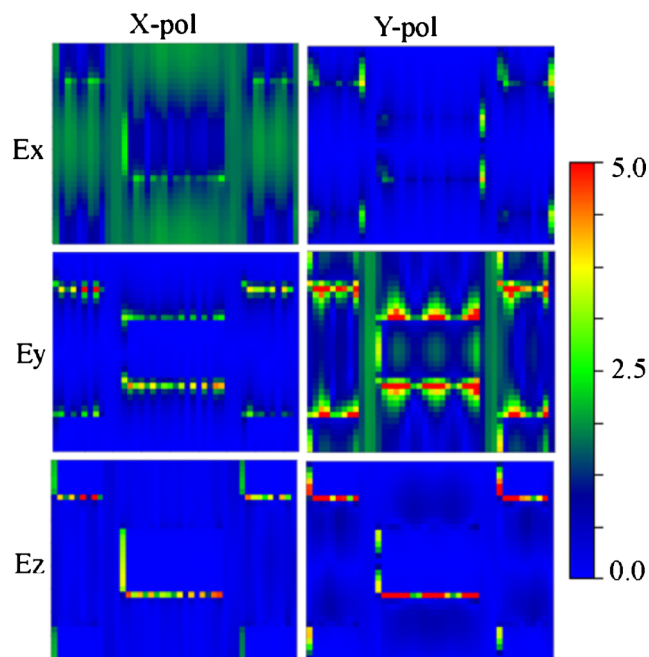


Fig. 5 Simulated electromagnetic near-field (E field) distribution for two crossed polarization states of incident radiation. x-pol case (*left column*) and y-pol (*right column*). Respective near-field components are labeled as E_x , E_y , and E_z

distribution, except a slight stronger confinement for the y-pol case. Overall, the y-pol excitation shows stronger confinement compared to x-pol excitation. Strong confinement leads to higher absorption as has been seen from the absorption spectrum plotted in Fig. 4. As the structure is more dissipative and reflective for the y-pol excitation, the spectral transmission is expected to be less. The computed results thus support the experimental observations.

Conclusions

It is demonstrated through experimental and computational studies that a ladder-like plasmonic structure can be used in designing broadband NIR polarizer in the spectral range from 900 nm to about 1400 nm. Spectral transmission studies reveal about 33 % transmission for x-pol excitation and about 4 % transmission for y-pol excitation. This significant transmission ratio (~ 8) demonstrates the polarizing action in the stated NIR regime and the suitability of the plasmonic ladder-like structure for the realization of broadband NIR polarizer.

Acknowledgments The author is thankful to Prof. S. A. Ramakrishna, Department of Physics, IIT Kanpur for computation and experimental support.

References

1. Ebbesen TW, Lezec HJ, Ghaemi HF, Thio T, Wolff PA (1998) *Nature* 391:667
2. Hu WQ, Liang EJ, Ding P, Cai GW, Xue QZ (2009) *Opt Express* 17:21843
3. Ding P, Liang EJ, Hu WQ, Zhou Q, Zhang L, Yuan YX, Xue QZ (2009) *Opt Express* 17:2198
4. Rodrigo SG, Martín-Moreno L, Nikitin AY, Kats AV, Spevak IS, García-Vidal FJ (2009) *Opt Lett* 34:4
5. Fan W, Zhang S, Minhas B, Malloy KJ, Brueck SRJ (2005) *Phys Rev Lett* 94:033902
6. Jiang YW, Tzuang LDC, Ye YH, Wu YT, Tsai MW, Chen CY, Lee SC (2009) *Opt Express* 17:2631
7. Mandal P, Anantha Ramakrishna S, Patil R, Venu Gopal A (2013) *J Appl Phys* 114:224303
8. McCrindle IJH, Grant J, Drysdale TD, Cumming DRS (2014) *Adv Opt Mater* 2:149
9. Sai G, Zhao Y, Liu H, Teo S, Zhang M, Huang TJ, Danner AJ, Teng J (2011) *Appl Phys Lett* 99:033105
10. Xu T, Wu Y-K, Luo X, Guo LJ (2010) *Nat Commun* 1:59
11. Song Z, Zhang L, Liu QH (2016) *Plasmonics* 11:61
12. Yang YM, Wang WY, Moitra P, Kravchenko II, Briggs DP, Valentine J (2014) *Nano Lett* 14:1319
13. Mandal P (2016) *Plasmonics* 11:223
14. Mandal P (2015) *Plasmonics* 10:439
15. Mandal P, Ramakrishna SA (2011) *Opt Lett* 36:3705
16. Gansel JK, Thiel M, Rill MS, Decker M, Bade K, Saile V, Gvon F, Linden S, Wegener M (2009) *Science* 325:1513
17. Fang Z, Zhen Y-R, Fan L, Zhu X, Nordlander P (2012) *Phys Rev B* 85:245401
18. Tittl A, Michel AKU, Schaferling M, Yin X, Gholipour B, Cui L, Wuttig M, Taubner T, Neubrech F, Giessen H (2015) *Adv Mater* 27:4597
19. Valev VK, Smisdom N, Silhanek AV, De Clercq B, Gillijns W, Ameloot M, Moshchalkov VV, Verbiest T (2009) *Nano Lett* 9:3945
20. Kauranen M, Zayats AV (2012) *Nat Photonics* 6:737
21. Li G, Chen X, Ni B, Li O, Huang L, Jiang Y, Hu W, Lu W (2013) *Nanotechnology* 24:205702
22. Hou B, Liao XQ, Joyce KSP (2001) *Opt Express* 18:3946
23. Yuan B, Zhou W, Wang J (2014) *J Opt* 16:105013
24. Black L-J, Wiecha PR, Wang Y, de Groot CH, Paillard V, Girard C, Muskens OL, Arbouet A (2015) *ACS Photonics* 2:1592
25. Hsu H, Makitalo J, Laukkanen J, Kuittinen M, Kauranen M (2010) *Opt Express* 18:16601
26. Abbas MN, Chang Y-C, Shih MH (2010) *Opt Express* 18:2509
27. Gupta B, Pandey S, Nahata A (2014) *Opt Express* 22:2868
28. Wang C-M, Chang Y-C, Abbas MN, Shih M-H, Tsai DP (2009) *Opt Express* 17:13526
29. Behera G, Mandal P, Ramakrishna SA (2015) *J Appl Phys* 118:063104
30. DiMaio JR, Ballato J (2006) *Opt Express* 14:2380
31. Alam MZ, Aitchison JS, Mojahedi M (2012) *Opt Lett* 37:55
32. Zhang X, Liu H, Tian J, Song Y, Wang L (2008) *Nano Lett* 8:2653
33. Dai D, Wang Z, Julian N, Bowers JE (2010) *Opt Express* 18:27404
34. Azzam S, Obayya SS (2015) *IEEE photon Tech Lett* 28:1
35. Mahros AM, Tharwat MM, Ashry I (2015) *Appl Optics* 54:4464
36. Hamidi SM, Zamani M (2015) *Opt Eng* 54:107104
37. Han C, Tam WY (2015) *Appl Phys Lett* 106:081102
38. Zhou J, Guo LJ (2014) *Sci Rep* 4:3614
39. Ellenbogen T, Seo K, Crozier KB (2012) *Nano Lett* 12:1026

Precompound emission in low-energy heavy-ion interactions from recoil range and spin distributions of heavy residues: A new experimental method

Manoj Kumar Sharma,^{1,*} Pushpendra P. Singh,² Vijay Raj Sharma,³ Mohd. Shuaib,³ Devendra P. Singh,⁴ Abhishek Yadav,⁵ Unnati,⁶ R. Kumar,⁵ B. P. Singh,^{3,†} and R. Prasad³

¹*Department of Physics, Shri Varsheny (Postgraduate) College, Aligarh, Uttar Pradesh 202 001, India*

²*Department of Physics, Indian Institute of Technology, Ropar, Punjab 140001, India*

³*Department of Physics, A.M.U., Aligarh 202002, India*

⁴*Department of Physics, University of Petroleum and Energy Studies, Dehradun, India*

⁵*Inter University Accelerator Centre, New Delhi, 110067, India*

⁶*Department of Physics, Delhi University, Delhi, India*

(Received 17 April 2016; revised manuscript received 16 August 2016; published 26 October 2016)

Recent investigations of heavy-ion reactions at low incident energies have indicated the presence of precompound emission component in considerable strength. In most cases the strength of the precompound component is estimated from the difference in forward-backward distributions of emitted light fast particles and also from the analysis of the measured excitation functions. This paper reports a new method of deciphering the relative contributions of compound and precompound components associated with fusion of ^{16}O with ^{159}Tb , ^{169}Tm , and ^{181}Ta targets by measuring the recoil ranges of heavy residues in an absorbing medium along with the online measurement of the spin distributions in reaction residues produced in the fusion ^{16}O beam with ^{159}Tb and ^{169}Tm targets. Analysis of recoil range and spin distributions of the residues shows two distinct linear momentum-transfer components corresponding to precompound and compound nucleus processes. The input angular momentum associated with precompound products is found to be relatively lower than that associated with compound nucleus process. The precompound components obtained from the present analysis are consistent with those obtained from the analysis of excitation functions.

DOI: [10.1103/PhysRevC.94.044617](https://doi.org/10.1103/PhysRevC.94.044617)

I. INTRODUCTION

The experimental observation of emission of light fast particles (LFP), particularly in the heavy-ion reactions at relatively low energies below 6 MeV/nucleon [1], has regenerated interest in the precompound nucleus (PCN) emission process, since it is expected to occur at high energies $\approx 10\text{--}15$ MeV/nucleon [2]. Generally, the relative strengths of compound and precompound components in such reactions are estimated from the enhancement in the flux of emitted LFP in the forward direction over the backward direction. Another method often employed is to analyze the measured excitation functions (EFs) for deviations from the statistical predictions and to attribute them to the PCN emission process. The understanding of the PCN and the compound nucleus (CN) emission in light-ion reactions has been well studied during the past few decades but in heavy-ion reactions it needs to be further explored particularly for those associated with the loss of particles in the primary stage in a very short reaction time (10^{-21} s) prior to the establishment of equilibrated CN [3–8]. The emission of such PCN particles reduces the momentum of the product residues. As such, the measurements of the momentum transfer during the interaction may provide a promising tool for the characterization of the reaction mechanism involved.

Although information about the momentum transfer in heavy-ion reactions may be obtained by several methods [2,9,10], in the present measurements this information has been obtained from the study of the recoil range distributions (RRDs) and the spin distributions (SDs) of the reaction residues. Since loss of particles emitted via the PCN process takes away a significant part of angular momentum, the angular momentum associated with the PCN products is relatively lower than that associated with the CN process. Therefore, in PCN reactions, the residues are populated with relatively less high spin states as compared to the spin states of the residues populated via CN process.

II. MEASUREMENTS AND ANALYSIS OF THE DATA

In order to investigate the role of PCN emission in heavy-ion reactions, three self-consistent measurements i.e., the RRDs, the SDs, and the EFs, have been performed. This paper reports on the measurements of the following: (i) the RRDs for the reactions $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ at incident energy 88 MeV; $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$; and $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$ at 90 MeV; and $^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Tl}$ reaction at 81, 90, and 96 MeV, respectively; (ii) the SDs for reactions $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ and $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$; and (iii) the EFs for the reactions $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$, $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$, and $^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Tl}$. Though the experimental details of RRDs [11,12], EFs [11,13], and SDs [14] are given in some

*Corresponding author: manojamu76@gmail.com

†bpsinghamu@gmail.com

of our earlier publications [11–14], a brief description on each is given in the following.

A. Recoil range distributions

An experiment has been carried out at the Inter University Accelerator Centre (IUAC), New Delhi, India, to measure the distribution of ranges of recoiling residues produced both by the CN and the PCN emission processes in $^{16}\text{O} + ^{159}\text{Tb}$, $^{16}\text{O} + ^{169}\text{Tm}$, and $^{16}\text{O} + ^{181}\text{Ta}$ systems. In the RRD experiments, the target followed by a stack of nearly 15 thin Al catcher foils of varying thickness ($\approx 16\text{--}45\ \mu\text{g}/\text{cm}^2$ prepared by the vacuum evaporation technique) was mounted in the irradiation chamber normal to the beam direction. Depending on the momentum carried away by the product residues, the recoiling residues were trapped at different ranges in the stack of thin Al foils. The duration of irradiation was about 15 h. The thickness of the catcher foils was measured precisely prior to their use, by measuring the energy loss suffered in each catcher foil by 5.485-MeV α particles from ^{241}Am source. The code SRIM was used for determining the thickness from the energy-loss measurements. The activities induced in each thin catcher were followed off line for about two weeks using a precalibrated high-resolution (2 keV for 1.33-MeV γ ray of ^{60}Co) HPGe detector.

In order to obtain the RRDs, the experimental values of cross section (σ) for the reaction products in different catcher foils were measured by the activation method and then these measured cross sections (σ) of the product residues in each catcher were divided by their respective thicknesses to give the resulting yields. The resulting yields plotted against cumulative catcher thicknesses give the experimental RRDs. The experimental RRD for reaction $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ is shown in Fig. 1. Solid curve in this figure guides the eye to the experimental RRD data. The bumps (peaks) in the experimental RRD arise due to the overlap of the heavy residues produced via two different reaction mechanisms, i.e., the CN and the PCN emissions of particles. As can be seen from this figure, the experimental RRD data for the reaction $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ has two peaks, one at a relatively lower value ($\approx 225\ \mu\text{g}/\text{cm}^2$) of cumulative catcher thickness and the other at $\approx 330\ \mu\text{g}/\text{cm}^2$. The peak at $\approx 330\ \mu\text{g}/\text{cm}^2$ corresponds to the fraction of residue produced through the CN process and is consistent with the full momentum transfer events in complete fusion reactions. The peak at relatively smaller range $\approx 225\ \mu\text{g}/\text{cm}^2$ may be attributed to the fact that the residues ^{183}Ir are produced via the PCN process when emission of neutrons takes place prior to the establishment of thermodynamical equilibrium. The emission of two neutrons through the PCN process takes a significant part of momentum as compared to the CN process. Thus, it reduces the momentum of the composite system and hence decreases the range of recoiling residues in the stopping medium. The theoretical simulations of the experimental RRD data for reaction $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ have also been performed using the code SRIM and the fitting program ORIGIN. A Gaussian peak (shown by blue dotted curve in Fig. 1) corresponding to the CN emission is constructed with ORIGIN

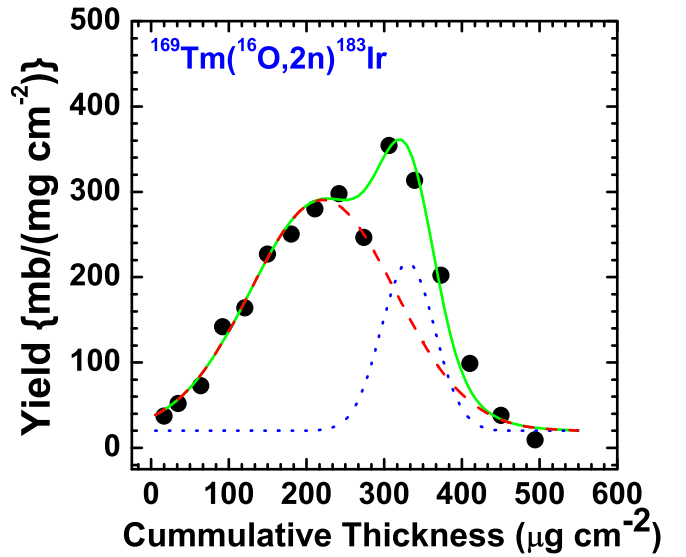


FIG. 1. The experimental RRD for the reaction $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$ at energy $\approx 88\text{ MeV}$. In this figure a Gaussian peak (blue dotted curve) at higher cumulative thickness represents the CN contribution while the other Gaussian (red dashed curve) at lower thickness represents the PCN contribution. The procedure to decipher the PCN and the CN contributions is discussed in the text.

program using the following expression;

$$y = y_0 + \frac{A}{w\sqrt{\pi/2}} e^{-2\frac{(x-x_c)^2}{w^2}}, \quad (1)$$

where y_0 = base, x_c = center, A = area under the peak, and w = width (FWHM) are the parameters required for fitting the data for a Gaussian peak of CN emission. The CN Gaussian peak reproduces the experimental data at higher cumulative thicknesses corresponding to the full momentum transfer. In order to find out the contribution of PCN emission, the CN Gaussian peak so constructed by the above procedure has been subtracted from the experimental RRD data. The Gaussian peak (shown by a red dashed curve) at lower cumulative catcher thickness of larger width is obtained, which shows the contribution of PCN emission. By using a similar procedure, the RRDs for reactions $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, and $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$ have also been measured at $\approx 90\text{ MeV}$. The experimental RRDs with their theoretical simulations for these reactions are shown in Figs. 2(a)–2(c), respectively. As can be seen from these figures, the experimental RRD data for these reactions have two peaks, one at a relatively lower value of cumulative catcher thickness corresponding to the PCN and the other for the CN at higher cumulative thickness.

In order to judge the reliability (or applicability) of the present method to decipher the CN and the PCN processes and to justify it, the RRDs for the $2n$ channel in the reaction $^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Ti}$ at three different energies, 81, 90, and 96 MeV, respectively, have also been measured and are shown in Figs. 2(d)–2(f). The following conclusions may be drawn from Figs. 2(d)–2(f): (i) At each energy, the measured RRD may be resolved in two distinctly different peaks. (ii) Peak corresponding to higher range (full momentum transfer) is

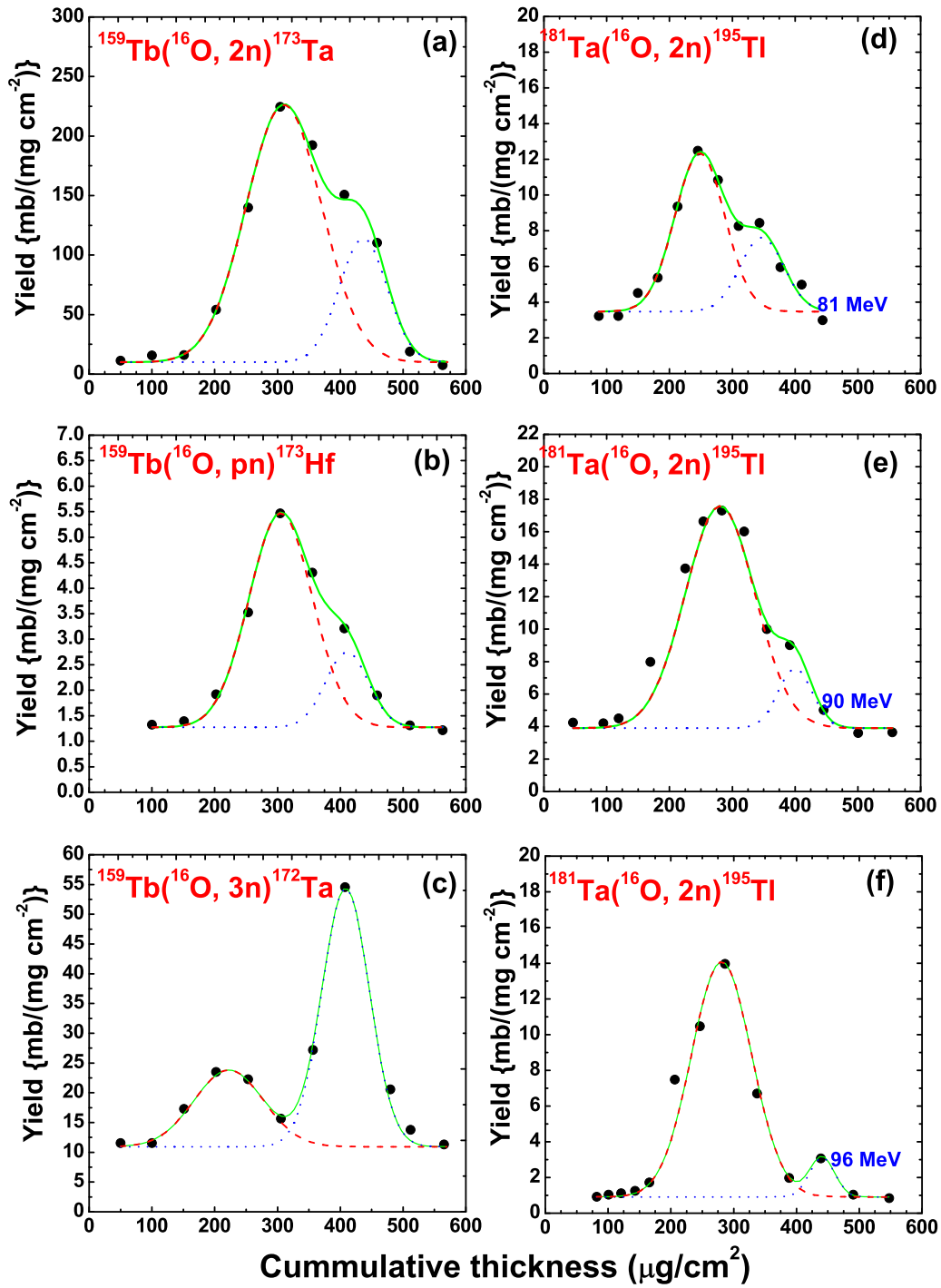


FIG. 2. (a)–(c) The experimental RRDs for the reactions $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, and $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$ at energy ≈ 90 MeV. (d)–(f) The energy-dependent RRD for reaction $^{181}\text{Ta}(^{16}\text{O}, 2n)^{197}\text{Tl}$ at energies ≈ 81 , 90, and 96 MeV, respectively.

assigned to the CN process while the peak at lower range is attributed to the PCN process. (iii) As incident energy increases from 81 to 96 MeV, the maxima of both the PCN and the CN peaks shifts towards the higher range side. This is expected since the increase of incident energy increases the linear momentum associated with both the PCN and the CN processes, which in turn results in higher ranges. (iv) It may also be observed from Figs. 2(d)–2(f) that the area

under the peak corresponding to the PCN process increases with the increase in incident energy while the area under the CN peak decreases. This reflects that the fact that the relative contribution of the PCN process over that of the CN process increases with incident energy.

In order to deduce the relative contributions of the PCN and the CN processes in these reactions, the areas under the Gaussian peaks corresponding to these processes

have been obtained using the ORIGIN program. The relative contributions of the PCN and the CN processes for the reaction $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$ at 88 MeV are found to be $\approx 80\%$ and $\approx 20\%$, while for reactions $^{159}\text{Tb}(^{16}\text{O},2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O},pn)^{173}\text{Hf}$, and $^{159}\text{Tb}(^{16}\text{O},3n)^{172}\text{Ta}$ at 90 MeV they are $\approx 70\%$ and $\approx 30\%$, $\approx 73\%$ and $\approx 27\%$, and $\approx 35\%$ and $\approx 65\%$, respectively, within the experimental uncertainties ($<10\%$) arising due to various factors, viz., uncertainty in the number of target nuclei due to nonuniformity in sample thickness, the fluctuations in the beam current, uncertainty in the determination of detector efficiency, statistical error in counts, and the dead time of the counting system. The energy-dependent contributions of the PCN and the CN processes in reaction $^{181}\text{Ta}(^{16}\text{O},2n)^{195}\text{Tl}$ are found to be $\approx 68\%$ and $\approx 32\%$ at 81 MeV, $\approx 79\%$ and $\approx 21\%$ at 90 MeV, and $\approx 88\%$ and $\approx 12\%$ at 96 MeV respectively. As such, with the increase in energy the PCN contribution is found to increase, as expected.

B. Spin distributions

To further confirm the present findings on the PCN and the CN processes, a second experiment based on particle- γ coincidence technique has been performed at IUAC, New Delhi, India, for measuring the population of spin states during de-excitation of reaction residues. In the present work, the spin distributions of the reaction $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$ at ≈ 88 MeV and $^{159}\text{Tb}(^{16}\text{O},2n)^{173}\text{Ta}$ at ≈ 93 MeV have been measured using the Gamma Detector Array (GDA) along with the Charged Particle Detector Array (CPDA). The experimental conditions are detailed in Ref. [14]. However, a brief account of the experiment is given here for ready reference.

The GDA is an assembly of 12 Compton suppressed, high-resolution HPGe γ spectrometers arranged at 45° , 99° , and 153° angles with respect to the beam axis and there are four detectors at each of these angles. The CPDA is a set of 14-phoswich detectors housed in a 14-cm-diameter scattering chamber, covering nearly 90% of total solid angle. The reaction residues have been identified from their characteristic prompt γ -transition lines. The proton- and α -emitting channels presented in Ref. [14] have been identified from particle ($Z = 1,2$)-gated γ spectra. However, neutron-emission channels have been selected from the singles spectra of observed prompt γ transitions collected at forward 45° and backward 153° angles with respect to the direction. The values of relative production yields of the residues (observed area under the peak of the experimentally measured prompt γ lines) have been plotted as a function of observed spin J_{obs} corresponding to prompt γ transitions [15].

In order to have information about the involved input angular momenta and to investigate the entry state spin population of the reaction $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$, the relative yield has been normalized with minimum observed spin ($J_{\text{obs}}^{\text{min}}$) at highest yield ($Y_{\text{obs}}^{\text{max}}$). The experimentally measured SDs obtained from prompt γ rays recorded in forward and backward directions for the reaction $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$ are shown in Figs. 3(a) and 3(b) at 88 MeV, respectively. As can be seen from these figures, the measured SD and hence its decay pattern for this reaction obtained in the forward and backward directions are distinctly different from each other, indicating

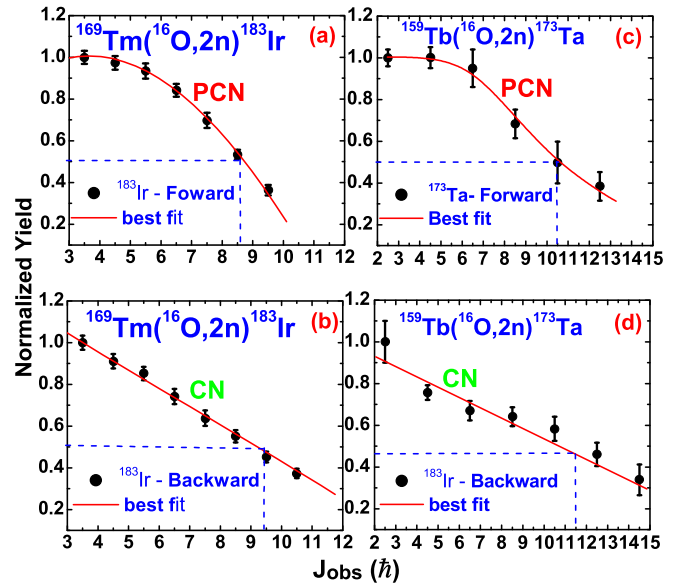


FIG. 3. The experimentally measured spin distributions for reactions $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$ and $^{159}\text{Tb}(^{16}\text{O},2n)^{173}\text{Ta}$ in forward and backward directions. The nomenclature used in the plots indicates the involved reaction mechanisms, i.e., the PCN and the CN. The lines and curves through data points are results of best fits.

widely different reaction mechanisms involved. It may be pointed out the entirely different shapes of SDs in forward and backward directions indicate that the two processes are quite different in nature. Further, in the case of CN it is expected that higher spin states in the residues are populated, giving rise to linearly increasing populations of lower spins due to their feeding from the higher spin states. On the other hand, in the case of the PCN, residues are left with relatively lower excitation energy and hence with lower spin entry state; therefore, the feeding of lower spin states is not linear.

The SDs for the CN and the PCN contributions may be characterized by the parameter, which is termed as the mean input angular momentum associated with the process, specified by the spin where the normalized yield falls to half of its maximum value at lowest observed spin. The mean input angular momenta, i.e., the spin at half yield deduced from experimental SDs of Figs. 3(a) and 3(b) are found to be $\approx 8.5\hbar$ and $\approx 9.5\hbar$ in forward and backward directions for the reaction $^{169}\text{Tm}(^{16}\text{O},2n)^{183}\text{Ir}$ while for reactions $^{159}\text{Tb}(^{16}\text{O},2n)^{173}\text{Ta}$ [shown in Figs. 3(c) and 3(d)], it is found to be $\approx 10.5\hbar$ and $\approx 11.5\hbar$ in forward and backward directions. The observed lower value of the mean input angular momentum in forward direction is due to the fact that emission of LFP (i.e., two PCN neutrons) takes away a significant part of the angular momenta. On the other hand, a relatively higher observed value of the mean input angular momentum in the backward direction is because of the emission of two equilibrated neutrons. As such, it is concluded that distinctly different SDs give direct evidence of the PCN emission process. Thus, the results of the measurements of the SDs further supplement the conclusions drawn from the RRDs measurements. The present work not only strengthens our earlier findings but also provides

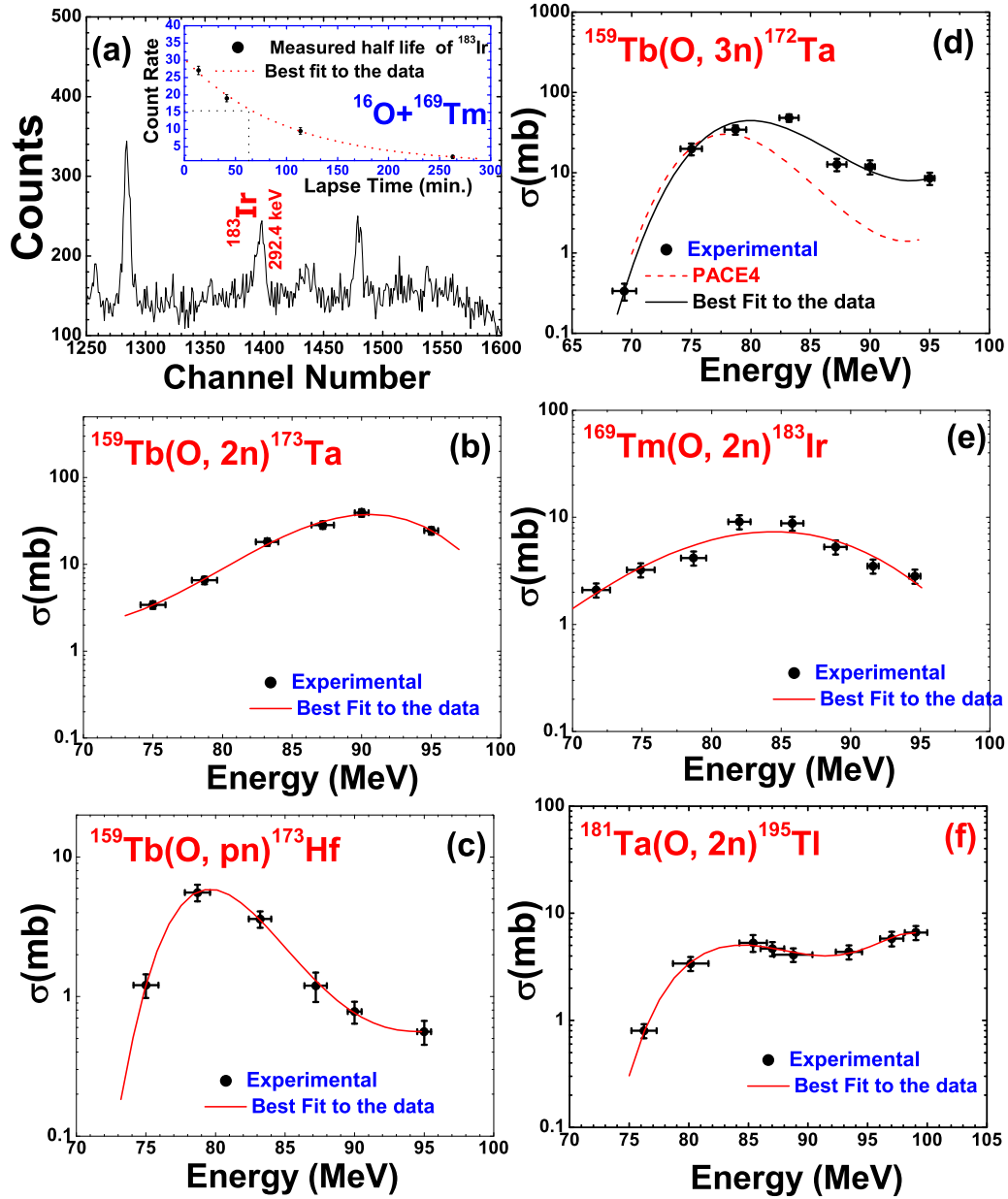


FIG. 4. (a) Observed γ -ray spectrum of ^{169}Tm sample irradiated by ^{16}O beam at ≈ 88 MeV. The measured decay curve of residues ^{183}Ir is shown in the inset. (b)–(f) The experimentally measured EFs for reactions $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$, $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$, and $^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Tl}$, respectively. The theoretically calculated EFs by using code PACE4 gives negligibly small values of cross sections for all these reactions except for $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$. For reaction $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$ experimentally measured EFs and their PACE4 predictions are also shown in (d).

additional information on the mean input angular momenta associated with the PCN and the CN processes.

C. Excitation functions

As already been mentioned, the deviation of the measured EFs from one calculated theoretical CN mechanism may also be used as a measure of PCN emission. In the present work the EFs for the reactions $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$, $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$, and

$^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Tl}$ have been measured in a separate experiments using the stack foil activation technique. In this experiment, the stacks consisting of (^{159}Tb , ^{169}Tm , and ^{181}Ta) target samples followed by Al foils of suitable thickness have been irradiated for ≈ 8 – 10 h in a specially designed General Purpose Scattering Chamber (GPSC) 1.5 m in diameter having an in-vacuum transfer facility. The Al foils serve as energy degrader as well as catcher foil where the recoiling residues are trapped. The pertinent decay data required for cross-sectional measurements of the reaction residue have been taken from Ref. [16]. The activities

produced in each target catcher assembly have been measured using a high-resolution large-volume (100 c.c.) high-purity germanium detector (HPGe) γ -ray spectrometer. A typical observed γ -ray spectrum of a ^{169}Tm sample irradiated by an ^{16}O beam at ≈ 88 MeV is shown in Fig. 4(a). The residues ^{183}Ir have been identified both by its characteristic γ ray and also from the measured half-life. The reaction cross sections for the product residues have been obtained from the measured intensities of the characteristic γ rays using the standard formulation [1].

The analysis of experimental EFs has been performed within the framework of statistical model calculations based on the code PACE4 [17]. The code PACE4 calculates the reaction cross section using the Bass formula [18] through the Monte Carlo procedure. The level density used in this code is calculated from the expression $a = (A/K)$, where A is the mass number of the compound nucleus and K is a free parameter known as the level density parameter constant. In the present work, a value of $K = 8$ is taken in the calculations, which is widely accepted. The detailed discussion of this code along with its parameters are given in Ref. [17].

The experimentally measured EFs for the reactions $^{159}\text{Tb}(^{16}\text{O}, 2n)^{173}\text{Ta}$, $^{159}\text{Tb}(^{16}\text{O}, pn)^{173}\text{Hf}$, $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$, $^{169}\text{Tm}(^{16}\text{O}, 2n)^{183}\text{Ir}$, and $^{181}\text{Ta}(^{16}\text{O}, 2n)^{195}\text{Ti}$ are shown in Figs. 4(b) to 4(f), respectively. The theoretically calculated EFs using code PACE4 give negligibly small values of cross sections for all these reactions except $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$; hence, these values are not shown in these figures. It means that analysis of EFs gives negligible contribution of CN process for the reaction, while the analysis of RRD for the same reactions gives energy-dependent contributions of the CN and the PCN processes. This shows RRD measurements are much sensitive as compared to EFs. Figure 4(d) shows that the absolute measured cross-sectional values for reaction $^{159}\text{Tb}(^{16}\text{O}, 3n)^{172}\text{Ta}$ are higher than predictions of theoretical calculations based on code PACE4 at higher energies. The enhancement of experimentally measured EFs over the

theoretical calculations of code PACE4 may be attributed to the PCN emission of a neutron in the first step of de-excitation before equilibration of composite nucleus takes place. Since the PCN emission is not taken into account in the code PACE4, the deviation of the measured data as compared to the theoretical predictions indicate significant contributions of the PCN process over the entire range of energy studied. The analysis of measured EFs for reaction $^{169}\text{Tm}(^{16}\text{O}, 3n)^{182}\text{Ir}$ gives a contribution of $\approx 35\%$ PCN and $\approx 65\%$ of CN at ≈ 88 MeV as shown in Fig. 4(d). These data are consistent with the corresponding values obtained from the RRD measurements at ≈ 88 MeV. Further, the reasonable agreement between the two sets of experiments speaks favorably on the consistency of these measurements.

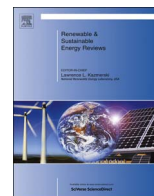
III. CONCLUSIONS

The experimentally measured recoil range distributions for the residues produced via the xn channel show a significant contributions due to the PCN process. The results obtained by recoil range distributions are supported by the spin distribution measurements where the mean input momentum involved is found to be lower for the PCN emission than for the CN emission process. The recoil range distribution measurements are found to be a sensitive tool to decipher the CN and the PCN processes. The auxiliary experiments on the excitation function measurements are found to be consistent with both the recoil range and spin distribution measurements.

ACKNOWLEDGMENTS

The authors are thankful to the director, IUAC, New Delhi, for extending all the facilities for carrying out the experiments. M.K.S. thanks the Council of Scientific and Industrial Research (CSIR), New Delhi, India, Project No. 03(1361)16/EMR-11, for financial support and A. K. Dixit, Principal, S. V. College, Aligarh, for all his support during this work. B.P.S. also thanks DST for providing financial support.

-
- [1] M. K. Sharma, P. P. Singh, D. P. Singh, V. Sharma, A. Yadav, Unnati, I. Bala, R. Kumar, B. P. Singh, and R. Prasad, *Phys. Rev. C* **91**, 044601 (2015).
 - [2] J. Gomez del Campo, D. Shapira, J. McConnell, C. J. Gross, D. W. Stracener, H. Madani, E. Cha'vez, and M. E. Ortiz, *Phys. Rev. C* **60**, 021601(R) (1999).
 - [3] H. C. Britt and A. R. Quinton, *Phys. Rev.* **124**, 877 (1961).
 - [4] T. Otsuka and K. Harada, *Phys. Lett. B* **121**, 106 (1983).
 - [5] P. Vergani, E. Gadioli, E. Vacic, E. Fabrici, E. Gadioli Erba, M. Galmarini, G. Ciavola, and C. Marchetta, *Phys. Rev. C* **48**, 1815 (1993).
 - [6] M. Cavinato, E. Fabrici, E. Gadioli, E. Gadioli Erba, P. Vergani, M. Crippa, G. Colombo, I. Redaelli, and M. Ripamonti, *Phys. Rev. C* **52**, 2577 (1995).
 - [7] C. Birattari *et al.*, *Phys. Rev. C* **54**, 3051 (1996).
 - [8] H. Delagrange, A. Fleury, F. Hubert, and G. N. Simonoff, *Phys. Lett. B* **37**, 355 (1971).
 - [9] H. Morgenstern, W. Bohne, K. Grabisch, D. G. Kover, and H. Lehr, *Phys. Lett. B* **113**, 463 (1982).
 - [10] B. B. Back *et al.*, *Phys. Rev. Lett.* **50**, 818 (1983).
 - [11] M. K. Sharma *et al.*, *Phys. Rev. C* **70**, 044606 (2004).
 - [12] D. P. Singh *et al.*, *Phys. Rev. C* **81**, 054607 (2010).
 - [13] P. Descouvemont, T. Druet, L. F. Canto, and M. S. Hussein, *Phys. Rev. C* **91**, 024606 (2015).
 - [14] P. P. Singh, B. P. Singh, M. K. Sharma *et al.*, *Phys. Lett. B* **671**, 20 (2009).
 - [15] REDWARE level scheme [<http://radware.phy.ornl.gov/agsdir1.html>].
 - [16] E. Browne and R. B. Firestone, *Table of Radioactive Isotopes* (Wiley, New York, 1986).
 - [17] A. Gavron, *Phys. Rev. C* **21**, 230 (1980).
 - [18] R. Bass, *Nucl. Phys. A* **231**, 45 (1974).



Progress in plasmonic solar cell efficiency improvement: A status review



P. Mandal*, S. Sharma

Department of Physics, University of Petroleum and Energy Studies, Bidholi-248007, Dehradun, India

ARTICLE INFO

Article history:

Received 29 June 2015

Received in revised form

28 June 2016

Accepted 7 July 2016

Keywords:

Review

Plasmonic solar cells

Surface patterning

Surface plasmon

Dye-sensitized solar cells

ABSTRACT

Solar cell efficiency improvement has been one of the major concerns to realize ultimately the cost effective efficient solar cells. Among various ways to improve solar cell efficiency, plasmonic light trapping mechanism has been found to be of immense interests recently. The mechanism of strong scattering into the active materials and guiding of light at the excitation of plasmons at the metal-semiconductor interface play significant role for better photon harvesting. The present review concentrates on the recent advances on the application of plasmonics in inorganic semiconductor solar cell efficiency improvements. Various research groups active in this field have employed various metal nanostructures on to the surface of solar cells to achieve higher efficiency. This review partially also concentrates on surface nanopatterning of solar cells with nonmetallic dielectrics. Finally, a brief account on the dye-sensitized solar cell is presented to show the potential of plasmonics in solar cell research.

© 2016 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	537
2. Surface plasmon polariton and localized surface plasmon	540
3. Efficiency improvement by surface texturing with dielectrics	541
4. Application of plasmon	542
4.1. Random distribution of metal nanostructures	542
4.2. Front side metal periodic texturing with metal nanostructures	543
4.3. Back side metal periodic texturing with metal nanostructures	544
4.4. Plasmonic metamaterial perfect absorber for solar cells	546
5. Plasmonic dye-sensitized solar cells	546
6. Future aspects and conclusions	549
Acknowledgments	549
References	549

1. Introduction

Solar cell, a source of cleaned electrical energy has been the most important alternative renewable energy source for many years. The device basically converts sunlight or solar radiation into electrical energy and therefore, it can reasonably be considered as a source of endless energy as long as sunshine persists. This has led to the primary interests in solar cell research by several

research groups worldwide. The most important and technologically developed solar cell module is based on 'silicon' till date. However, a wide variety of solar cell structures also available based on various active materials, such as, thin film CIS (CuInS_2 ($/\text{Se}_2$)), Cu(In, Ga)S_2 ($/\text{Se}_2$) solar cell, single-/multi-junction III–V solar cells etc.

Dye-sensitized/quantum-dot-sensitized solar cells and polymer solar cells also seem to be very interesting and promising since the price of these types of solar cells is expected to be much cheaper. However, in general, many factors such as device efficiency, materials capital and the mature device processing

* Corresponding author.

E-mail address: pmandal@ddn.upes.ac.in (P. Mandal).

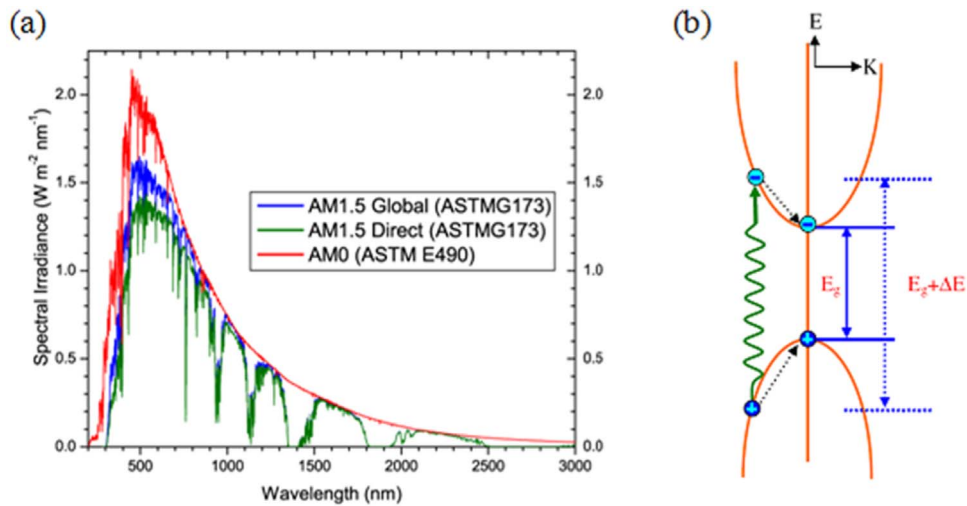


Fig. 1. (a) Represents standard solar spectra and (b) schematic for the electron-hole (e-h) pair generation by the absorption of photon energy by an arbitrary semiconductor having band gap (direct) energy E_g . Photons energy above the band gap (i.e. excess energy ΔE) will be lost due to thermal heating (indicated by black dotted arrows). (a) is adapted from the reference [1].

technology have always been the important issues for making these devices efficient and cheap with large scale productivity. Among the issues, the first and foremost importance is given to the improvement in the solar power conversion efficiency. Effective absorption in the wide solar spectral range covering 350–1250 nm (standard solar spectrum [1]) is shown in Fig. 1(a)) has been an important aspect of solar cell research. The primary interest of increasing the cell efficiency is to study the efficient light trapping mechanism which can result in enhancement in the optical absorption [2–21]. However, efficient absorption does not

guarantee for efficient generation of electron-hole pairs and hence photo-voltage. Photon energy around the band gap of the semiconductor used as the active material will only effectively create charge pairs. The excess photon energy (above the band gap) will be dissipated as heat. A schematic representation (Fig. 1(b)) will help in understanding the process. Therefore, suitable choice of active materials will enable in obtaining better photon absorption and generation of charge pairs. However, no single active semiconductor will be best for the efficient absorption; therefore multi-junction solar cells (or tandem solar cells) have also been

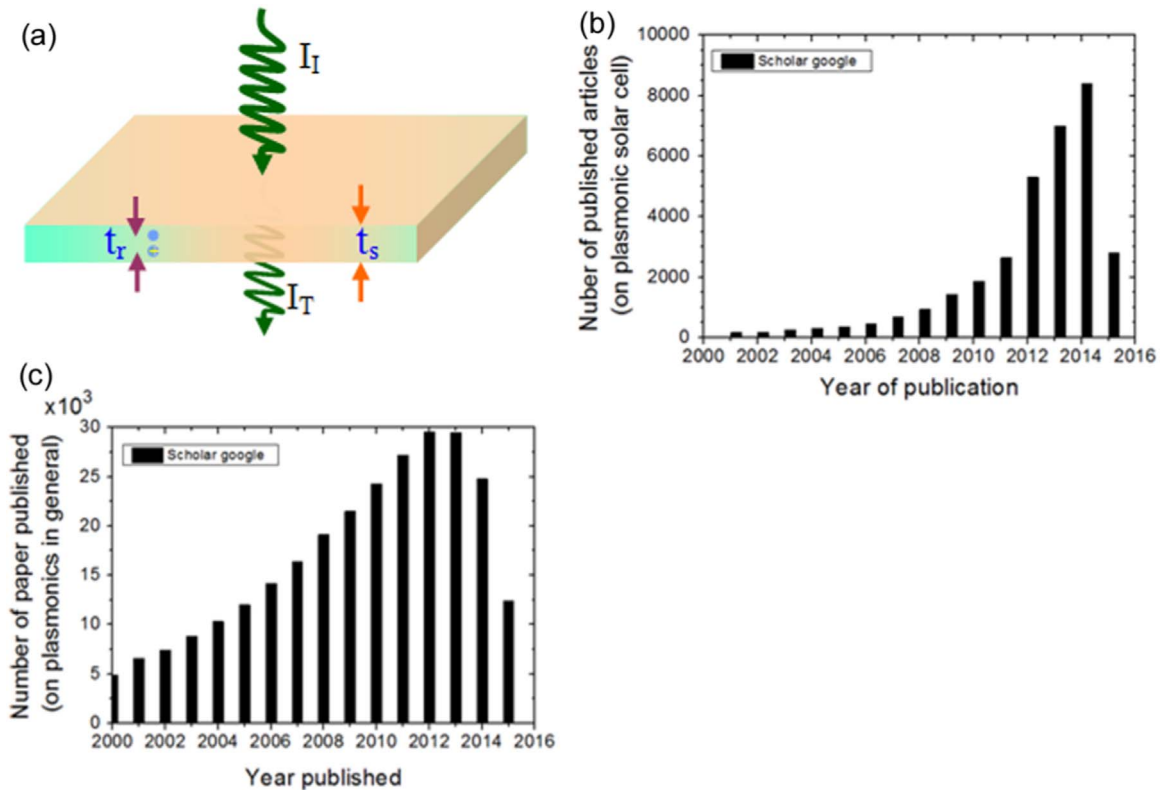


Fig. 2. (a) Represents photon absorption scheme by a semiconductor of thickness t_s . Carrier diffusion length is represented by t_r . I_i and I_T represent incoming and outgoing photon fluxes, respectively. (b) represents the research strength over the years in terms of published papers (to the author's knowledge using search engine 'scholar google') showing interest in plasmonic solar cells. The above search engine has also been used for the general search on plasmon associated with all possible fields of research (corresponding bar graph plot is shown in (c)).

investigated. Even for single semiconductor active material, photons having energy of the order of band gap may not be absorbed fully through a single pass across the semiconductor of thickness ' t_s ' (a photon absorption scheme is depicted in Fig. 2(a)). This simple scheme requires multiple pass through the semiconductor for complete absorption, inferring if the semiconductor is sufficiently thick (or optically thick) then the photons may be fully absorbed by single pass. But this has an adverse effect on the charge collection on the respective electrodes on either side of the semiconductor. Because once a charge pair (electron-hole pair) is generated it will move a certain distance called 'diffusion length (say t_r)' before being recombined together. The recombination process is nothing but a loss of charge carriers and therefore quite unwanted in efficient solar cells. Therefore, optically thick semiconductor (fully absorbed photon for one pass) is not suitable for better charge carrier separation. Also thicker the semiconductor the more will be the amount of materials needed, demanding earth abundant amount of material source which does not sound viable naturally as well as economically. This situation demands the need for thinner semiconductor layer thickness. Thin layer will now satisfy the condition of $t_r > t_s$ but at the same time it, however, also demands multiple pass across the semiconductor layer. This problem can be addressed by applying back side reflector or more effectively a textured back reflector for multiple scattering back into the absorbing material. The other approach is to trap photons inside the semiconductor by some means and increase the photon path length. Efficient scattering of photons into active material can also be achieved by applying textured front surface. Front side or backside surface texturing can be of either random or periodic in nature. After scattering, the some of the scattered light will be converted to guided light waves propagating within the active material between the collecting electrodes. This, in turn, increases the photon path length, leading to efficient photon absorption in the active material. The increase absorption generates further charge carriers resulting in enhanced efficiency. However, if the textured surface is of metal a new resonance phenomenon will arise called surface plasmon resonance. Surface plasmon is the charge density oscillation at the metal surface and exists at the metal-dielectric interface. Surface

plasmon can be excited with the associated electric fields of incoming photon flux. Resonance excitation has tremendous effect on the scattering power of the surface structures or metal nanostructures. Plasmon excitation is doubly beneficial for the solar cells, on one hand a strong scattering of photons can be achieved, while on the other hand, a guiding wave (non radiative) may exist at the interface. Very large local electric field due to surface plasmon resonance may cause strong light scattering and generate huge charge carriers. The use of plasmonics in solar cells research now triggers tremendous interests among the researchers actively involved in this field. Excitation of plasmon and its use in solar cell would be very promising for the enhancement of solar cell efficiency and is expected to be a major thrust in the present decade. Fig. 2 (b) represents the research strength over the years in terms of published papers showing interest in plasmonics in solar cells (Fig. 2(c) represents the bar graph of plasmon related published articles associated with all possible fields of research). Surface plasmon existing at the metal-dielectric interface can be excited by light of suitable wavelength and by the use of metal nanostructures on the front surface, back surface or inside the active materials. The novel metal nanostructures, such as nanoparticles, nanorods or disks have been employed for the excitation of localized surface plasmons, whereas, periodic metal nanostructures (metallic one-dimensional and two-dimensional gratings) on top surface of thin metal film (referred to Fig. 3 for various possibilities) have been employed for the use of propagating surface plasmons as well as localized surface plasmons. It can be noted here that for propagating surface plasmon a thin metal layer is required underneath the periodic dielectric or metallic nanostructures. Owing to the importance we have reviewed here how plasmonic light trapping mechanisms can be applied to improve solar cell efficiency. A general review is rather presented which describe the efforts of the researchers over the years to enhance solar cell efficiency by the application of plasmonics. Although there have been various cell structures, namely, inorganic solar cells (single or multi-junction), organic/polymer solar cell, dye-sensitized solar cells, only solar cells based on inorganic semiconductors are reviewed.

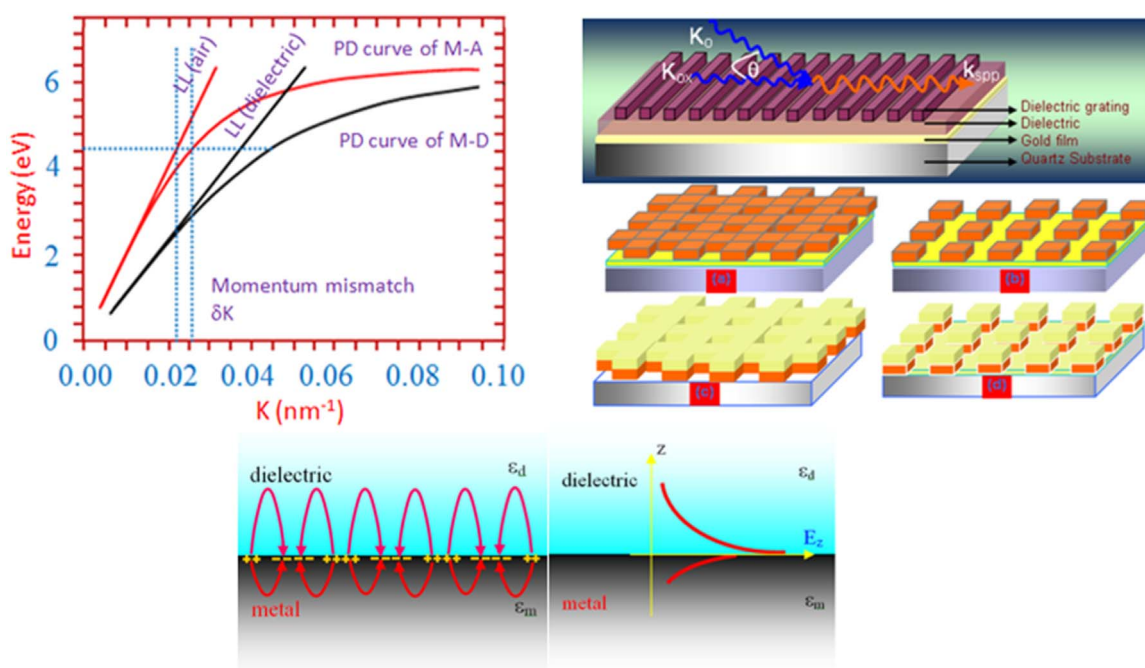


Fig. 3. Clockwise: plasmon dispersion (PD) curves and momentum mismatch (top left), coupling of photon to plasmon in one-dimensional and two-dimensional gratings (top right), and generation of plasmon at metal dielectric interface and decay of field (bottom). The bottom figure is adapted from the reference [25].

2. Surface plasmon polariton and localized surface plasmon

Solar cells efficiency improvement by efficient light trapping through the front/back surface nanopatterning by using dielectric as well as metal nanostructures has been studied for quite some time now. Like dielectric nanopatterning, metal nanostructures have also been employed through either random distribution over the surface or with a definite periodic structure. In this review, we are mainly concerned about the metal nanostructures; therefore, it would be helpful in understanding the plasmonic effect of metal nanostructures on solar cells if we are aware of a few important parameters related to the surface plasmon polariton (SPP) or localized surface plasmon (LSP). Before going into details of the application of plasmon to the solar cell, we therefore would discuss first a few basic important aspects of surface plasmon. Metal nanostructure on the top of an active semiconductor has air-metal (in most cases, or sometime glass-metal interface) and semiconductor-metal interfaces. Surface plasmon an electronic charge density oscillation exists at the these metal-dielectric interface and when light falls on to the surface of solar cell these plasmonic modes get excited provided the momentum matching condition gets satisfied (Fig. 3 shows the momentum mismatch of plasmonic mode). This is essential for one-dimensional (1D) and two-dimensional (2D) definite arrays of metallic nanostructures (relation-1 and relation-2, respectively) as shown in Fig. 3. Relation-1 and relation-2 refer to the momentum matching conditions in the case of propagating surface plasmon polariton excitation in periodically corrugated metal surface. It can be noted here that isolated metal nanoparticle distributed randomly or in a periodic manner would show localized surface plasmon. In relation-1 and relation-2, the term $2\pi/p$ represents grating vector in reciprocal space, whereas, \hat{u} represents the unit vector along the plasmon wave oscillation direction. The unit vectors \hat{i} and \hat{j} are along x direction and y direction, respectively for a square lattice (in the present case). The terms ϵ_d and ϵ_m ($=1-\omega_{mp}^2/\omega^2$, ω_{mp} being metal plasma frequency, neglecting damping term) are dielectric permittivity and metal permittivity, respectively. In the case of metal nanoparticles, such as metal spherical nanostructures, plasmon modes are localized in nature and can be excited without following such kind of momentum matching rules. This is an added advantage over periodic arrangement concerning about propagating surface plasmon. The propagating plasmon propagates in the direction perpendicular to the metal-dielectric surface normal. This waveguiding feature is beneficial for further light trapping into the active material. Obviously, larger propagation length is expected in such type of application as in plasmonic solar cells. Surface plasmon propagation length (L_{spp}) is defined as the length at which the plasmon intensity falls off $1/e$ times due to out-coupling to photons or absorption, and is as written in relation-3. For gold (Au) this characteristic length is in the range of 15 nm to 135 μm , and for silver (Ag) it is in the range of 3.5 μm to 585 μm for the wavelength range of 400 nm to 1000 nm, respectively. The value is calculated using relation-3, and air as dielectric [22–25]. It is, thus, obvious that Ag would be the preferred material as per plasmon concerned. The other issue is enhanced electromagnetic near-field near metal nanostructures which is important for enhancing the scattering of incoming photon flux, and in boosting the process of creation of e-h pairs. As the near-field decays exponentially (Fig. 3), the spatial extension on either side of the metal-dielectric interface can be approximated to the skin depth at low frequency regime which is the distance at which evanescent electric field falls off $1/e$ time, and can be calculated using relation-4. This is an important parameter which signifies the necessity of optimum thickness of the metal and dielectric active layers. Spatial extension is calculated using relation-4 and relation-5 and the

wavelength corresponding to the maximum solar irradiance (~ 500 nm), and found to be nearly 30 nm and 300 nm for gold (Au), and nearly 25 nm and 400 nm for silver (Ag). In the calculation air medium is taken as dielectric for both the cases.

$$|\vec{K}_{SPP}| = \left| K_0 \sin \theta \hat{u} \pm m \frac{2\pi}{p} \hat{i} \right| = K_0 \sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}} \quad (1)$$

$$|\vec{K}_{SPP}| = \left| K_0 \sin \theta \hat{u} \pm m \frac{2\pi}{p} \hat{i} \pm n \frac{2\pi}{p} \hat{j} \right| = K_0 \sqrt{\frac{\epsilon_m \epsilon_d}{\epsilon_m + \epsilon_d}} \quad (2)$$

$$L_{SPP} = \frac{1}{2K_x} = \left[2K_0 \left(\frac{\epsilon_m(\text{real}) \cdot \epsilon_d}{\epsilon_m(\text{real}) + \epsilon_d} \right)^{\frac{3}{2}} \frac{\epsilon_m(\text{imaginary})}{2(\epsilon_m(\text{real}))^2} \right]^{-1} \quad (3)$$

$$Z_m = \left(\frac{1}{K_0} \right) \cdot \left(\frac{\epsilon_m(\text{real}) + \epsilon_d}{\epsilon_m^2(\text{real})} \right)^{1/2} \quad (4)$$

$$Z_d = \left(\frac{1}{K_0} \right) \cdot \left(\frac{\epsilon_m(\text{real}) + \epsilon_d}{\epsilon_d^2} \right)^{1/2} \quad (5)$$

$$\Omega_{\text{scattering}} \propto \frac{d^6}{\lambda^4} \quad (6)$$

$$\Omega_{\text{absorption}} \propto \frac{d^3}{\lambda} \quad (7)$$

Like propagating surface plasmon polariton or propagating plasmon, localized surface plasmon has also been very important in enhancing solar cell efficiency. LSP due to metal nanoparticles distributed randomly or in a regular manner leads to enhanced electromagnetic fields around metal nanostructures, which, in turn, contributes to scattering of light. LSP excitation is possible either polarized light (transverse magnetic (TM) and transverse electric (TE)). Scattering depends on the size of the metal nanoparticles according to the relation-6 (scattering proportional to sixth power of particles size d). In solar cell, absorption (relation-7, absorption proportional to third power of particle size d) by metal particles itself (d - d electronic transition in metal) is not preferred which will basically lead to the loss of photons. Efficient scattering of photons into active materials is expected which may cause in increase of photon path length and hence enhanced absorption. From the relation-8 (combination of relation-6 and relation-7, for details, readers are requested to look at the reference [26]), it can clearly be seen that for an optimum scattering for a range of wavelengths, say for visible to near infrared, particle size must be chosen carefully. For metal particles of size of the order of 100 nm, the ratio of scattering cross-section to absorption cross-section will only be determined by the proportionality factor (because d/λ is always less than 1 in this wavelength range). Proportionality factor can be determined once the dielectric functions of the metal used and its surrounding are known. The particles of optimum size may be introduced inside the active region of the solar cell, or can be deposited on the front side or back side of the solar cells. Plasmonic back patterning would be preferred compared to front side patterning, as in the former case the overall optical scattering back into active layer will be high.

With this brief introduction, we will now focus on the available reports existing in the literature on the plasmonic solar cells. We will, especially, describe the solar cell devices where plasmonic particles or patterns have been introduced in the front side or back side of the cells. Before entering in to the section of 'application of plasmon' to solar cells, we will briefly discuss first the current

activity of application of dielectric nanopatterning or dielectric nanoparticles on solar cells' efficiency improvement.

3. Efficiency improvement by surface texturing with dielectrics

Absorption enhancement by the active layer can be obtained through surface texturing, specially top surface texturing, and has been known for the years [27–48]. The underlying mechanisms involved are total internal reflection and efficient forward light scattering. Scattering to all possible directions gives rise to increased optical path which results in enhanced absorption. The stated mechanisms may be effective in the case of textured surfaces obtained by direct patterning of the top (front) surface of the solar cells. The pattern can be obtained by various lithography processes or by simple self assembling process. Out of various lithography processes, laser interference lithography and nanoimprint lithography seem to be the most interesting cost effective processes with large scale surface patterning capability in a short period of time. There are many reports on the surface dielectric nanopatterning using laser interference lithography and self assembling process. For example, Chang et al. [49] prepared 2D surface texturing by self assembling of one monolayer polystyrene micro spheres of various sizes on the top of a GaAs solar cell (Fig. 4). They have studied systematically the changes occurred in the optical absorption and external quantum efficiency of the solar cell, and found maximum absorption and quantum efficiency when solar cell surface was covered by $1\ \mu\text{m}$ (size) polystyrene spheres. The highest increase in efficiency is found to be $\sim 25\%$ relative to the solar cell efficiency without polystyrene spheres. Because of the higher dielectric constant of the polystyrene spheres, the incoming photons are forward scattered into the active material and the optical path became longer. Increased path length is expected to play major role in enhancing the absorption and efficiency. They have confirmed the experimental results of enhanced efficiency through FDTD (finite difference time domain) simulation of similar solar cell structure. There is another report of solar cell efficiency improvement by using textured front surface. Laser interference lithography was employed to pattern the front surface of InGaP single junction solar cell. The present structure was investigated by Chang et al. [50]. They obtained patterned InGaP solar cell with 2D microholes array with varying array periods but keeping microhole size constant. A general scheme of lithographically patterned top surface is shown in Fig. 4. Their studies shows that an enhancement of $\sim 15\%$ in the solar conversion efficiency can be achieved with a textured microholes with hole size $\sim 5\ \mu\text{m}$. They have also shown experimentally that the

best efficiency can be obtained when array period and microhole size are of the order of $5\ \mu\text{m}$. Increased light trapping and scattering events have been the basis of enhanced efficiency. In a most recent report we [51] have shown that single junction GaAs solar cell efficiency can be enhanced by 23.5% when front surface of solar cell is textured with line pattern or nanoholes pattern. Laser interference lithography is adapted for this study. The patterning is done either in the form of 1D line or 2D nanoholes arrays. It is shown that for maximum light scattering of visible part of the solar spectrum, periodicity of the structures would be order of 500 nm. The experimental results have also been supported by the simulation studies. According to the report, increased photon path length caused by increased light scattering into active material plays a major role in enhancing the device efficiency.

In past years, there have been many reports on solar cell efficiency improvements by surface texturing. For example, Zhao et al. [52] studied multicrystalline Si solar cell with a honeycomb surface texturing on the front surface. They have reported that an energy conversion efficiency of about 19.8% can be achieved which is about 6% more efficient compared to multicrystalline cell without surface texturing (efficiency 18.6%). The same group [53] has also reported a conversion efficiency of 24% in monocrystalline Si solar cell with inverted pyramidal surface texturing in the front surface. Inverted pyramidal surface texturing is effective in efficient surface reflection from the back surface and at the same time the reflection loss from the front surface is minimized. However, in all these solar cells the main mechanism involved for the improvement of the solar cell efficiency is the efficient light scattering back into the active material by periodic surface texturing, which resulted in increased optical path length. This, in turn, increases the overall absorption, causing higher generation rate of charge pairs. It is noteworthy that for the periodically surface textured solar cells, an optimization in terms of period of the periodic structure and textured depth (or height, especially important for the case of surface texturing with metal nanostructures, as discussed in the coming section) is essential for optimum light scattering back into the active materials. In this context, a report by Dewan et al. [54] showed through simulation that the effective light scattering can be obtained if the period of the periodic structure is in the wavelength range of incident light if the wavelength is high (700–1100 nm) and the effective period is below 200 nm if the wavelength range is between 300 nm to 500 nm. The other important parameter of improving the solar cell efficiency is the minimization of the front surface reflection loss. For the purpose, the use of antireflection coating on the front surface has been realized. There are many reports on the reduction of front side reflection loss by surface roughening and by the use of antireflection coating. Interested readers are referred to the

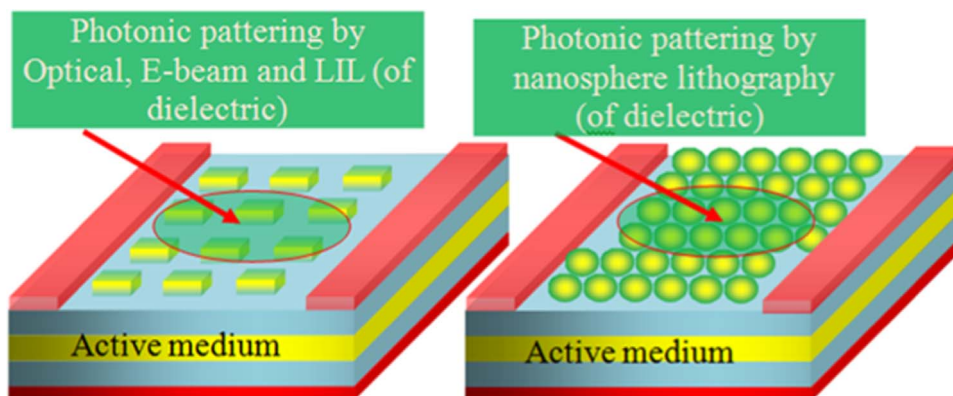


Fig. 4. General scheme of pattern surface on the top of the active solar cell surface is shown. The patterning is done by (left) laser interference lithography, optical lithography, e-beam lithography, and (right) self assembly or nanosphere lithography.

reports by Yerokhov et al. [55], Huang et al. [56], Tao et al. [57], and Dobrzanski et al. [58]. In the present review as we are mainly concerned about the application of plasmonics in the solar cells efficiency enhancement, we will now concentrate on the subject by confining ourselves into the role of metallic nanostructures (periodic or non-periodic) in solar energy conversion. The brief introduction of surface dielectric nanopatterning only outlines the idea which may be further modified in order to achieve best improvement. Most importantly, the application of front surface dielectric nanopatterning on the single junction/multijunction solar cell along with surface plasmonic back reflector is expected to be the most effective configuration. The idea of plasmonic back reflector and/or the front surface metal nanopatterning will now be considered in details in the coming sections. In all these reports, the efficiency improvement of solar cells is realized by the application of random metal nanostructures as well as regular array patterns. We will consider first the case of randomly distributed metal nanostructures and subsequently the case of regular metal nanostructure arrays will be discussed.

4. Application of plasmon

4.1. Random distribution of metal nanostructures

Random distribution of metal nanoparticles onto front surface of solar cell is expected to increase strong light scattering due to plasmonic interaction, which, in turn, increases optical path length. The increased path length would obviously be favorable for enhanced absorption resulting in increased efficiency. For example, Nakayama et al. [59] studied in details the effect of random distribution of Ag nanoparticles of various sizes and shapes on to the single junction GaAs solar cell surface. Their solar cell structure is shown in Fig. 5 (left). Growth of Ag nanoparticles has been performed using anodic aluminum oxide mask. Right panel of Fig. 5 shows SEM (scanning electron microscope) images of Ag nanostructures with varying shape and sizes. Highest efficiency they achieved is of the order of 8%, which is nearly 25% (calculated from the data they have tabulated) more than the reference solar cell. The most effective nanostructure is shown in the right panel (c) of Fig. 5. From their studies they came to the conclusion that, for equal metal particle density, it is the vertical height which matters (compare (b) and (c)) in relation to the improvement of cell efficiency. It is also reported that the decoration of solar cell surface with the metallic nanostructure actually reduced the sheet resistance which is an added advantage. A report by Pillai et al. [60] showed a 16 fold photocurrent enhancement in Si thin film

solar cell by decorating front surface with randomly distributed Ag nanoparticles grown by simple thermal evaporation and post annealing at 200 °C. They have also studied wafer based Si solar cell with similar Ag nanoparticles decoration and found only 7% improvement in the efficiency. Randomly distributed Ag nanoparticles deposited on the front surface act as antireflection coating causing the enhancement in optical absorption. Their studies show that the enhancement of the Si-on-oxide solar cell with smaller Ag nanoparticles is larger compared to wafer based solar cell. Effect of plasmon is significant in the case of thinner cell with smaller Ag particles compared to thicker one with similar configuration. Here the reason is, in thicker cell, the active material alone significantly absorbs, and therefore, further improvement by the effect of plasmon is not very significant. This observation is interesting and it can be noted that the active layer thickness can readily be reduced by employing plasmonic layer, keeping the overall light absorption high. In a similar approach, Akimov et al. [61] have employed silver nanoparticles on top of a Si solar cell and studied plasmon assisted optical absorption. They used a thin oxide layer (SiO_2) between Si active layer and metal nanoparticles, and studied the details of plasmonic absorption enhancement with varying surface coverage by Ag nanoparticle, and the thickness of the oxide layer. They have shown that the coupling between nanoparticles and active Si layer becomes less significant when the surface oxide layer becomes thick. It is also shown that there is an optimum condition existing relative to the surface oxide layer, surface coverage and metal nanoparticles size. In plasmonic solar cells efficient light scattering has been an important issue, as mentioned previously. Scattering into the active layer and its guiding through the active layer is key for strong improvement of absorption and hence efficiency. Mokkaapati et al. [62] addressed this issue. They have shown that part of the scattered component will have photonic mode in the active layer, and the other part will have modes in air. For the device to be more plasmon active the scattered component into the active layer should be strong. The light due to total internal reflection will be trapped before being fully absorbed. It is, however, pointed out by the group that the active scattered components depend on particle size, shape and the effective dielectric constant of the surrounding media. Schaadt et al. [63] also reported the importance of metal nanoparticles' size and shape in relation to the efficiency of solar cell devices. To realize the importance of particles' shape and size they studied a p-n junction Si photodiode with Au nanoparticles of various sizes deposited on the top surface of the photodiode. Surprisingly, they noticed a huge increase, as high as 50–80%, in photocurrent response relative to the device without Au nanoparticles. Apart from the metal nanoparticles' size and shapes,

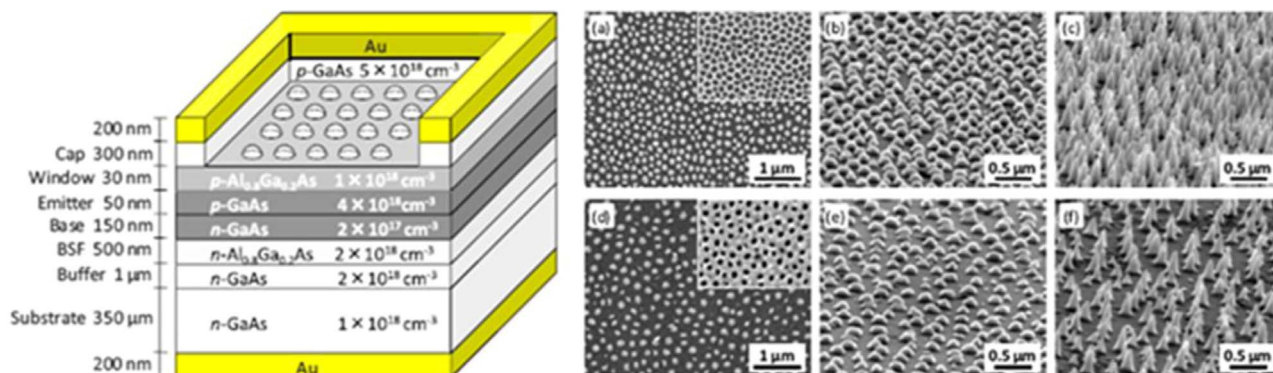


Fig. 5. (left) GaAs solar cell unit and (right) SEM images of the template assisted metal nanoparticles growth with the following specifications: (a)–(c) and (d)–(f) represent the dense ($3.3 \times 10^{-2}/\text{cm}^2$) and sparse ($1.8 \times 10^{-2}/\text{cm}^2$) arrays of 110 nm diameter Ag nanoparticles, respectively. Inset figures represent the corresponding reference templates. Height of the nanoparticles in (b) and (e) is 55 nm, whereas in (c) and (f) it is 220 nm. Figures are adapted from the reference [59].

importance has also given to the density of metal particles i.e. surface coverage. For example, Derkacs et al. [64] fabricated Si solar cell with p-i-n structure and deposited Au nanoparticles on the top surface of the photodiode. They observed a energy conversion efficiency of about 8.3% with Au nanoparticle density of $3.7 \times 10^8/\text{cm}^2$. From finite-element electromagnetic simulation they have confirmed that the particle density of $2.5 \times 10^9/\text{cm}^2$ will be optimum for the studied structure. This result signifies the role of surface coverage (or particle density) by metal nanoparticles.

Till now we have discussed about device efficiency improvement by metal nanoparticles such as Ag and Au. However, there are reports on other metallic structures too, such as of Cu, Al and Ni. These metals (Cu, Al) are cheaper than Au, Ag and Ni and can be used for large production. Using various noble metal nanoparticles (Au, Ag and Cu), Stuart et al. [65] studied forward light scattering. They found better efficiency with larger nanoparticles. They have also found that, out of Au, Ag and Cu, the preferred metal is Ag which gives the maximum light scattering and hence absorption. However, in all these studies the issue related to parasitic absorption by metal nanoparticles itself has not been addressed. Parasitic absorption by metal nanoparticles itself will have adverse effect on efficiency improvement, therefore, needs great attention. To avoid parasitic absorption loss, especially at lower wavelength, metal nanoparticles' shape and size have to be carefully chosen. Koh et al. [66] enlightened this issue. They studied Si solar cells (Fig. 6, left panel) with Ag nanoparticles as well as SiC and TiO₂ dielectric nanoparticles. Dielectric spheres are less absorptive and expected to be strong scattering sources. The optimized parameters used for the experiments for highest absorption are as follows: size and surface coverage are respectively as 25 nm and 11% for Ag, 25 nm and 78.5% for SiC, and, 20 nm and 74.5% for TiO₂. Surprisingly, according to their observation the metal nanoparticles based solar cell offered lowest absorption enhancement of 10% compared to 29% (SiC) and 23% (TiO₂) as in Fig. 6 (right panel).

4.2. Front side metal periodic texturing with metal nanostructures

Periodic metal nanostructures on the top surface of solar cells [67–80] are mainly interesting for localized surface plasmon as well as surface plasmon polariton assisted scattering and guiding of light. The combined effects lead to the longer optical path resulting in enhanced absorption. Based on theoretical prediction, improvement in the solar cell efficiency is possible and can be, in fact, manipulated with periodic structural parameters such as grating period, fill factor and depth. Because of plasmon oscillations at the meta-dielectric interface the number of associated modes will help in getting better photon harvesting and hence absorption [81–95]. We will discuss these issues based on the available literatures. One-dimensional metallic line patterns, two-

dimensional metallic holes or pillars/rods, disks etc have been employed on both the surfaces of solar cells and improvement in the cell efficiency is realized. According to the existing literature it is seen that there are various metals that have been employed to improve efficiency. Silver is preferred due to its less absorptive nature in the visible spectral range, as mentioned earlier. The present section describes the efficiency improvements by using periodic metal nanostructures fabricated onto the front surface of solar cells.

Thin film Si based a model solar cell structure has been simulated in order to show enhanced efficiency by the application of periodic plasmonic layer (Pala et al. [96]). Their simple Si solar cell structure is as shown in Fig. 7. Ag nanostrips of 50 nm are employed onto the front surface of Si solar cell. The structure has a spacer layer of SiO₂ of about 10 nm for higher optical transparency and better electrical surface passivation. Short circuit current enhancement of the order of 43% with the optimum period of 295 nm is observed. Near field light concentration by localized plasmon resonance and increased light trapping due to the coupling with waveguide modes are responsible for the enhanced efficiency. It is also concluded in their report that there could be even more enhancement (> 50%) if the active layer thickness is in the range of 100 nm–1 μm . The increase in Si layer thickness basically governs waveguide modes which may contribute to the overall efficiency improvement. In a slightly different approach, Lee et al. [97] used two-filling factor asymmetric binary grating to study enhancement of light trapping due to guided mode resonance in solar cells. They concluded that about 40% enhancement in the efficiency is achievable relative to a planar Si thin film. Interestingly, enhancement in such a structure is not significantly affected for wide angle of incidence. In support of their statement of light trapping due to guided mode resonance they have presented simulated results performed on grating structure. Their simulated results showed that the incident power periodically spread over the guided mode resonance structure elongating the optical path. Increased light trapping in a plasmonic amorphous Si photovoltaic cell has also been investigated through finite difference time domain simulation by Chao et al. [98]. They have considered both TM and TE polarized light to excite plasmon modes and found that for TM polarized light there is a 28.7% enhancement in absorption (photon absorption increased by a factor of 1.287) relative to the structure without grating. However they have pointed out that the structure demonstrates bad light trapping properties under TE light. This problem can be overcome with the use of mesa type grating and will be best suitable for solar light which is randomly polarized in nature. More recently, simulation based optimal design of 1D and 2D plasmonic solar cells has been studied by Rockstuhl et al. [99]. The simulation has been carried out on the amorphous Si structure having 1D or 2D Ag gratings. Absorption enhancement by a factor of 1.6 relative to the

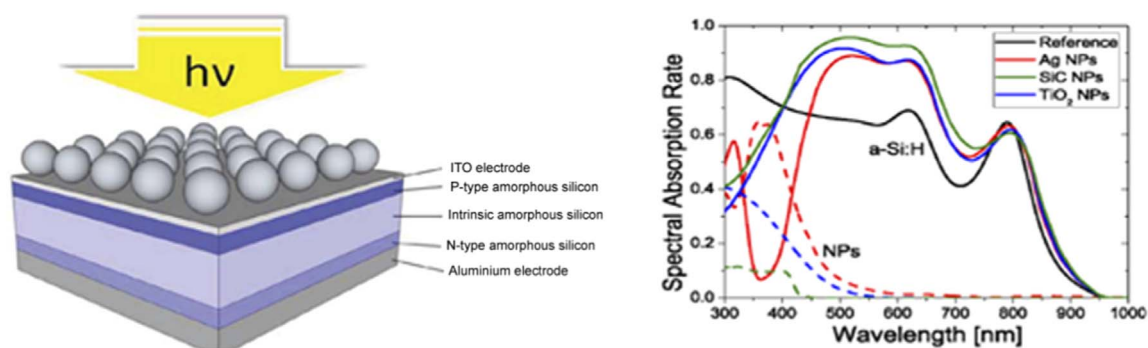


Fig. 6. Si solar cell structure (left) and photoresponse (absorption) (right). Figures are adapted from the reference [66].

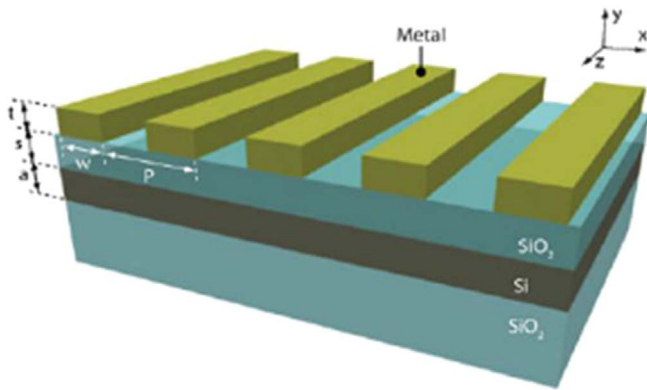


Fig. 7. One-dimensional periodic metal pattern on the top of a Si solar cell. The parameters t , s , a , w and p represent metal height, SiO_2 layers height, absorbing Si layers thickness, metal width and period of the metal pattern, respectively. Figure is adapted from the Reference [96].

absorption in the case of structure without grating has been achieved by considering the optimized silver nanowire width $w_{\text{Ag}} = 140$ nm and height $h_{\text{Ag}} = 69$ nm for entire simulation. Optimized period is found to follow $\Lambda = 1.5 w_{\text{Ag}}$ relation. Large near field enhancement and strong scattering cross section due to localized plasmon excitation are responsible for such enhanced absorption.

It is observed that 1D metallic strips based plasmonic solar cell absorption is very much polarization dependent. Thus 1D structure may be less effective for efficient absorption. This can easily be avoided by applying bi-periodic structures (mesa type) and will be best suit for unpolarized polarized light such as sunlight. Bi-periodic metallic nanostructures in the form of nanodiscs have been employed to enhance solar cell efficiency by the same group [99]. 2D metal nanostructure employed by them is shown in Fig. 8. With the parameters mention in the caption of Fig. 8, the absorption enhancement factor obtained by the group is 1.45, which

is comparable to the 1D case. However, they have mentioned that the absorption enhancement likely to be increased by optimizing the disc diameter and height. In 2009, the group [100] showed through simulation that the disc shaped 2D arrays of Ag on amorphous Si solar cell surface give rise to absorption enhancement of about 50% (with the optimum parameters such as grating period and disc diameter). Efficiency improvement by front surface texturing with other metal nanoparticles such as of Au, and with dielectric nanoparticles such as of SiO_2 onto InP/InGaAsP quantum well solar cells has been investigated for by Derkacs et al. [101]. They obtained an enhancement in the short circuit current of 7.3% and 12.9% in the case of Au and SiO_2 covered solar cell. They concluded that the normal incident light coupling into lateral propagation paths through available slab waveguide modes formed by multiple quantum well layers is responsible for the improvement.

4.3. Back side metal periodic texturing with metal nanostructures

Solar cell efficiency improvement by texturing back surface has also been studied by various researchers [102–122]. Scattering loss evident with front surface texturing may be minimized in this case. However, active semiconductor surface will be affected by back surface nanopatterning which will lead to increased surface recombination; resulting in loss of charge carriers. Nevertheless, the large enhanced field and waveguiding modes will provide increased light trapping and propagation along the interface, thereby, increasing optical path length and hence absorption [103,123–140]. We will now concentrate on the available literature which deals with backside plasmonic surface texturing. Although there are many simulation based reports enlightening the enhancement, experimental evidences are not so less.

Absorption enhancement of a Si solar cell structure has been investigated through simulation by Bai et al. [141]. The structure they used for the simulation is shown in the Fig. 9 (top panel). They used surface texturing in the form of rectangular ridge and

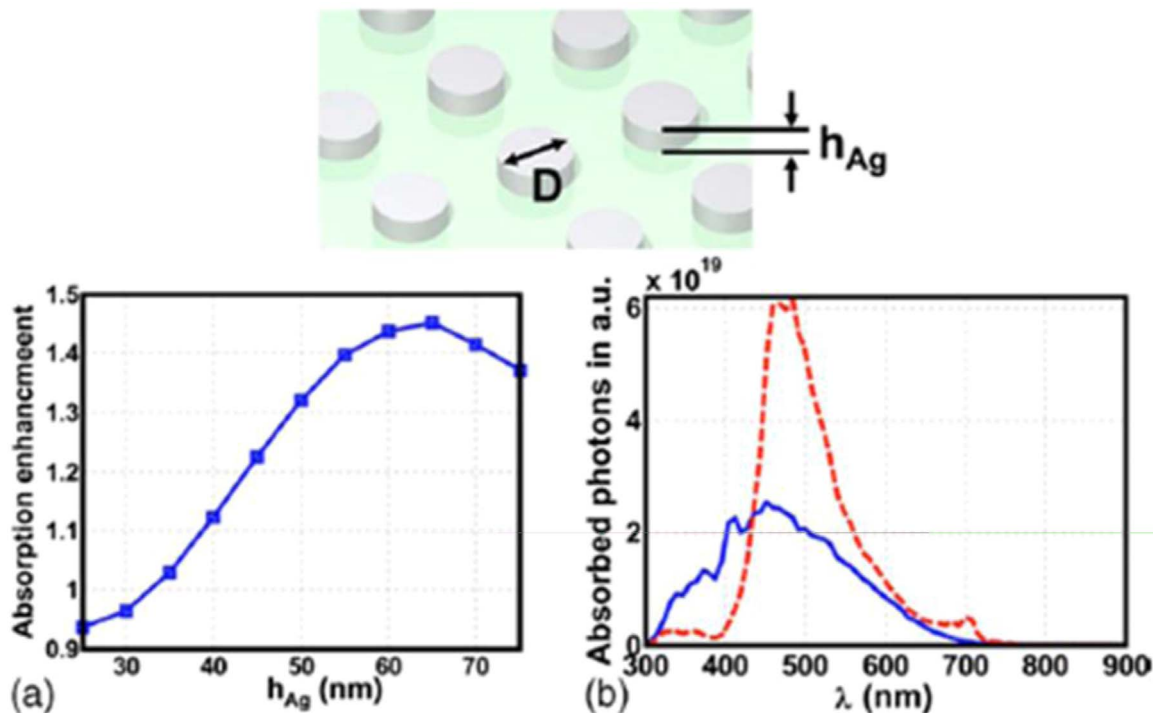


Fig. 8. Nanodiscs array in a square lattice (top row) and absorption enhancement (bottom row) as a function of disks height (a) and wavelength (b). In (b) the disc height is considered to be 65 nm. Disc diameter is 140 nm and period of the arrays is 175 nm. Figures are adapted from the reference [99].

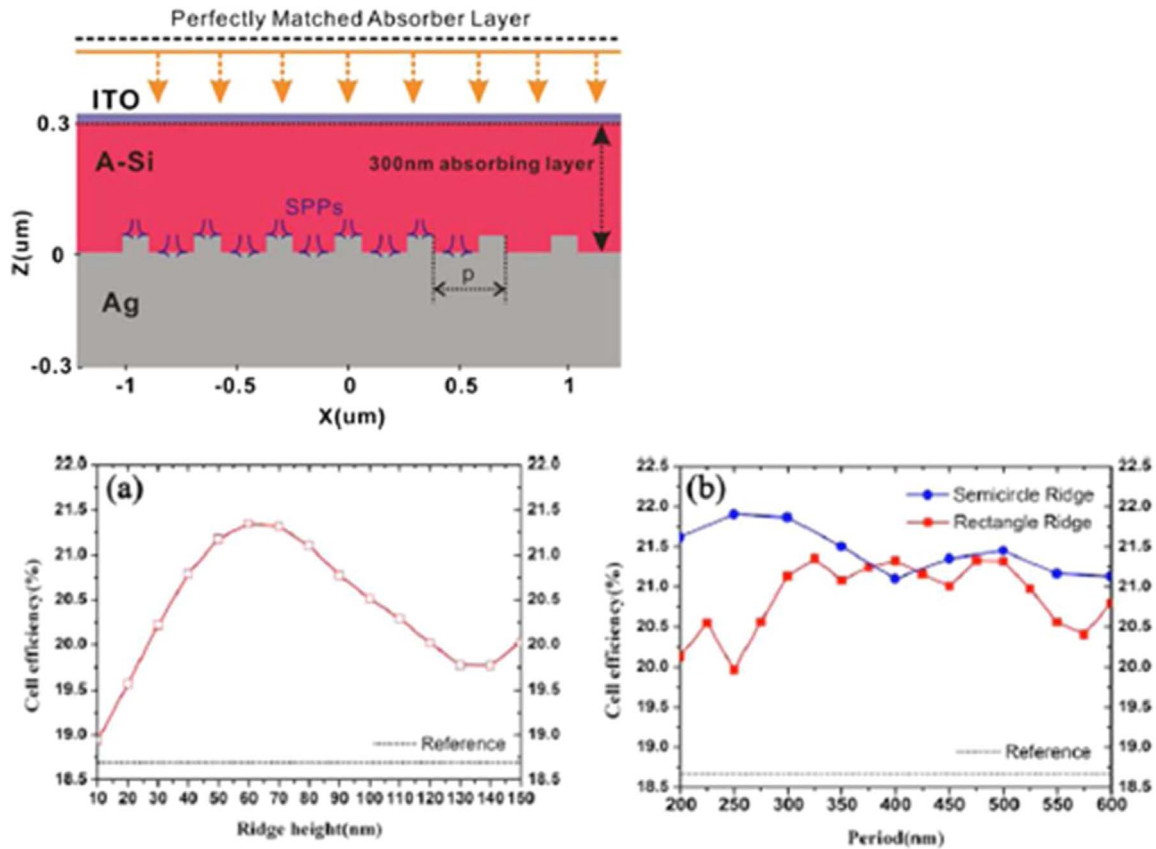


Fig. 9. (top) Amorphous Si solar cell structure used for the simulation with the rectangular ridge and semicircular ridge shaped back surface texturing. The efficiency plot for the rectangular ridge case with varying ridge height is shown in figure (a). With rectangular ridge height of 60 nm and semicircular ridge diameter of 100 nm, the corresponding efficiency plots with varying periods are shown in figure (b). Figures are adapted from the reference [141].

semicircular ridge. With 325 nm period they obtained cell efficiency plots as shown in Fig. 9(a) (bottom panel). By changing ridge height they obtained highest efficiency at a ridge height of 60 nm. However keeping ridge height fixed at 60 nm they varied period of the ridge pattern, and observed that the best set point for the period would be in the range of 300 nm to 525 nm. This optimized parameters offered ~12.5% improvement in the efficiency (Fig. 9(b)). In a similar manner when they simulated the structure with semicircular ridge of diameter 100 nm they obtained a corresponding efficiency plot as shown in Fig. 9(b). In the latter case the efficiency improvement exceeded (17%) that of the former case. Ferry et al. [142] studied nanoimprinted holes pattern over coated with Ag as plasmonic back reflector for amorphous Si solar cell. They have investigated the structure through simulation and

experiment. In their study they have obtained a short circuit current enhancement by 26%, and the overall efficiency increased from 4.5% to 6.2% relative to the flat back surface with metal coating. Back side plasmonic reflector has also been investigated experimentally by Sai et al. [143]. They used honeycomb like Al pattern (anodic oxidation of Al, Fig. 10) onto which Ag (100 nm) film was sputtered and microcrystalline Si solar cell was realized onto the patterned back reflector. The external quantum efficiency was measured (as in Fig. 10) and found to increase from 16.8 to 20.7, which shows an enhancement of about 23%. Better light trapping inside the active semiconductor is thought to be the cause of this enhancement. In a simulation based report, Mokkapatil et al. [144] studied a two-dimensional plasmonic grating onto solar cell (Si) back surface and suggested the possibility of

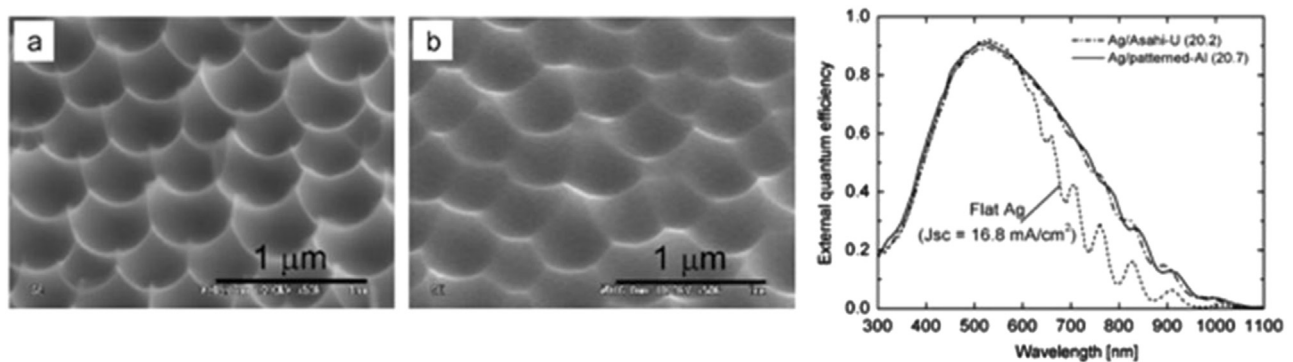


Fig. 10. Honeycomb dimple patterned obtained by anodic oxidation of Al surface at 195 V in a H_3PO_4 solution: after Al_2O_3 etching (a) and after additional H_3PO_4 treatment for 2 h (b). The current-voltage characteristics in the corresponding case of flat back reflector and patterned back reflector are shown in the figure (extreme right). Figures are adapted from the reference [143].

optimum design parameters. They have concentrated onto certain parameters in order to obtain optimum efficiency. They concluded that scattering cross section and diffraction efficiency are the two important factors. It is also pointed out that the grating periods would be such that there would be higher diffraction modes available for higher wavelengths while keeping fill factor at highest possible value. In their study, simulation has been performed on the structure by either varying the metal particle dimension while keeping pitch of the structure constant or by varying the pitch while keeping particle dimension constant. Fig. 11 shows the simulated data which outlines the optimized parameters for Si solar cells. The optimum data corresponds to the particle plasmon resonance wavelength (optimum pitch: 400 nm and Ag square base particle dimension: 200 nm) close to the Si band gap. Plasmon and photon coupling is important to increase the scattering cross-section and hence the solar cell efficiency. Ferry et al. [145] demonstrated that the subwavelength scatterers in the form of back interface coated with corrugated metal film will effectively couple incident light into guided modes in Si or GaAs solar cells. The coupling is insensitive to a great extent of incident angle. They have reported 2.5 times absorption enhancement near Si band gap of Si solar cells. A novel type of plasmonic light trapping back reflector composed of colloidal metal particles synthesized through solution process is studied by Mendes et al. [137]. The self assembled particles uniformly covered the Si surface showing enhanced absorption and short-circuit current density.

4.4. Plasmonic metamaterial perfect absorber for solar cells

It is observed that most of the published results concentrate mainly on the plasmonic effects on the solar cells by introducing plasmonic nanostructures either on the front side or back side. Very limited articles report both sides surface texturing with active plasmonics. With the front side metal nanopatterning, care

should be taken in order to assure overall transmission remains high. In this regard, combination of pure dielectric nanopatterning on the front side and plasmonic nanopatterning on the back side could be very promising for efficient light trapping. This will ensure less reflection loss from front side and at the same time the back side scattering would be the strong resulting in the overall high absorption. Nevertheless the antireflection coating should be effective for broad band light harvesting. It has been observed that even after an effective coating, front side reflection loss can not be reduced to zero. The idea of perfect absorber [146] or perfect impedance matched layer can be employed. Sculptured film and metamaterial perfect absorber although would be promising route for better photon harvesting, however, wide spectral absorbance can only be obtained using perfect absorbers having multiple resonances. This needs a pre-design optimization aiming at the wide spectral absorbance covering UV–vis to NIR [147,148]. Here, we can note that apart from the research on perfect light trapping one has to find plasmonic material that would be cheap. Al and Cu can be considered as good choice in place of Au or Ag [149].

5. Plasmonic dye-sensitized solar cells

Apart from Si solar cells, active research is going on the dye sensitized solar cells (DSSCs). Since the report by Michael Gratzel [150] on the dye sensitized solar cell about 20 years back, DSSCs have attracted tremendous interests [151–159] due to easy materials processability, low production cost, flexibility in designing the structure and high solar to electrical power conversion efficiency of about 10% (low compared to crystalline Si solar cells, a global solar cell efficiency list can be found in reference [151]). For the interest we have put a bar graph on the global research interests on dye-sensitized solar cells as in Fig. 12. The key points in developing the highly efficient dye-sensitized solar cells include, finding various materials which may cover wide solar spectral

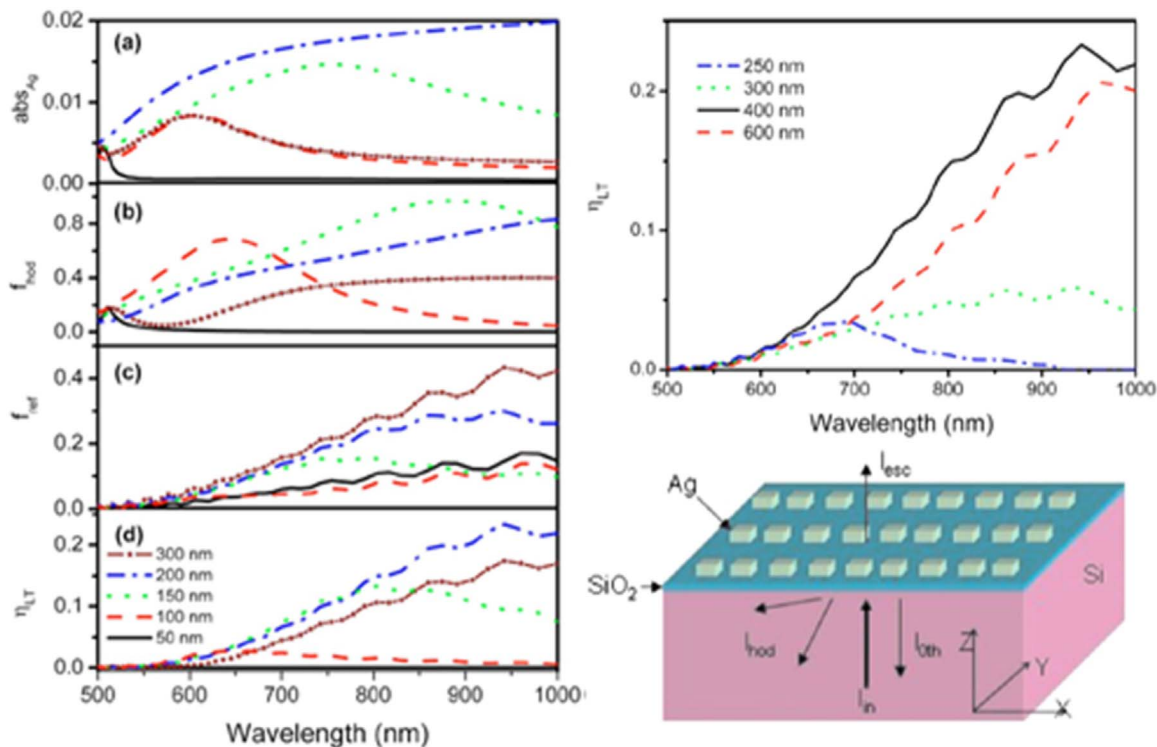


Fig. 11. (Left panel) simulated data (varying particle dimension) for the light coupling into the higher order diffraction modes. (Right panel, top) efficiency plot for the structure with varying periods at constant particle dimension of 200 nm. The modeled structure used for the simulation is shown in the bottom figure (right panel). Figures are adapted from the reference [144].

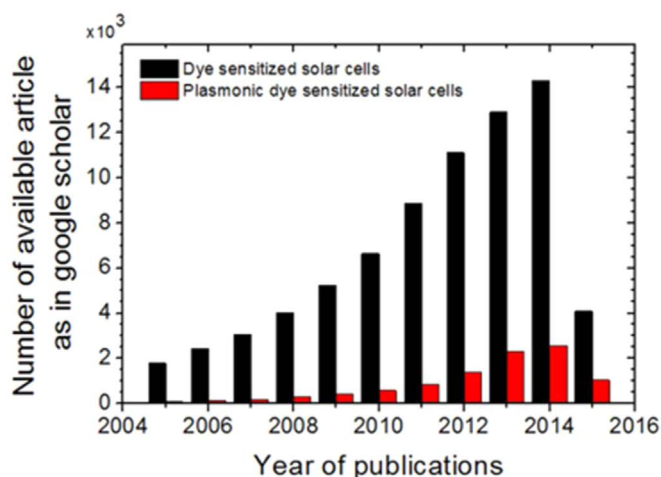


Fig. 12. Bar graphs represent the general interests in term of published articles on the plasmonic dye-sensitized solar cells. Google-scholar search engine is used. The phrases 'plasmonic dye-sensitized solar cells' and 'dye-sensitized solar cells' are used to avail the information on research strength (available articles).

absorption by employing strong absorbing dyes [160,161], efficient generation of electron-hole pairs [162], thereafter, transportation of generated charge carriers [163], and reducing the loss of charge carriers at the defects or interface [164]. Spectral absorption or efficiency improvement can be achieved through light trapping techniques of which plasmon coupling seems to be very effective [162–174]. Interests on plasmon assisted dye-sensitized solar cells are growing up as can be found from the bar graphs of Fig. 12. A plenty of opportunities exist in improving dye-sensitized solar cell efficiency through the use of plasmonic near fields. Strong plasmonic near-fields increase the photon scattering cross-section enormously, thereby increasing the overall dye absorption. A completely new route for the plasmonic near-field engineering is emerging having high impact on PV-technology (photo voltaic=PV), chemical sensing, optical engineering and others. In this communication, we are more concerned about the dye-sensitized PV-technology, therefore, a brief review on the plasmon assisted dye sensitized solar cell work is presented.

In a typical dye-sensitized solar cell structure the following layers are essential. A transparent conductive oxide (TCO) anode coated with TiO_2 layer, dye (Ruthenium complex molecules, for example) which is adsorbed onto the anode, electrolyte such as I^-/I_3^- redox shuttle and Platinum Coated Glass (PCG) or TCO as cathode. Typical thickness of the TiO_2 layer is of about 12–15 μm [152,153,175–179]. The incident solar photons are absorbed by the

photo sensitized dye molecules producing electrons. These excited electrons are then transferred to anode mediated by TiO_2 nanoparticles layer. Dye molecules are then regenerated by collecting electrons from the electrolytes, and electrolytes get neutralized by capturing electrons from the cathode (Fig. 13). It is to be noted that the electrons flow through the external circuit from anode to cathode.

The efficient light scattering occurs when plasmonic nanoparticles are used by mixing with dyes or a thin metal nanoparticles layer is coated on to TiO_2 layer. Strong photon scattering by the enhanced near-fields caused by localized surface plasmon resonance (LSPR) takes place, resulting in better photon absorption by the dye molecules. However, there have been other approaches such as plasmonic back reflector, and front reflector, similar to what is discussed in the case of inorganic solar cells. Photons get coupled to plasmon (surface plasmon polariton: SPP) through 1D or 2D Bragg gratings fabricated using cheap nano-imprinting method/laser interference lithography/nanosphere lithography. The reflectors enhance absorption through excitation of plasmonic modes in the spectral range overlapping with dye absorption band, especially in the long wavelengths ($> 500 \text{ nm}$). Composites of dyes-semiconductor nanoparticles-plasmonic nanoparticles can be of better choice, although, the device performance may get affected adversely due to the probable existence of interface and defects. This problem may be overcome by using core-shell semiconductor nanoparticle or other core-shell structures. With this introduction we now present a brief review on the current status of plasmon assisted dye-sensitized solar cells.

In this section we will mainly concentrate on the efficiency improvement by the use of surface plasmon near-fields in the dye-sensitized solar cells. In recent reports [180–189] it is shown that the surface plasmon resonance in noble metal nanoparticles such as Ag and/or Au nanoparticles greatly enhance light absorption due to strong light scattering, thereby increasing the photocurrent and hence the device efficiency. Li et al. [180] have reported 25% efficiency improvement of DSSC when they used photo-anode of Ag nanoparticle doped TiO_2 nanofibers in comparison to undoped DSSC. The improved efficiency of the cell was collectively attributed to enhanced optical absorption due to surface plasmon effect of Ag nanoparticles as well as the faster electron transport in the photo-anode of Ag doped TiO_2 nanofibers (Fig. 14). Similar reports can be found in reference [181]. It is noteworthy that the crystallinity of plasmonic nanoparticles can play an important role in absorption enhancement [182]. A report by Tanvi et al. [183] has shown that the absorbance peak for composite dye-Ag nanoparticles was significantly enhanced in the case of single-crystalline Ag nanoparticles compared to poly-crystalline nanoparticles.

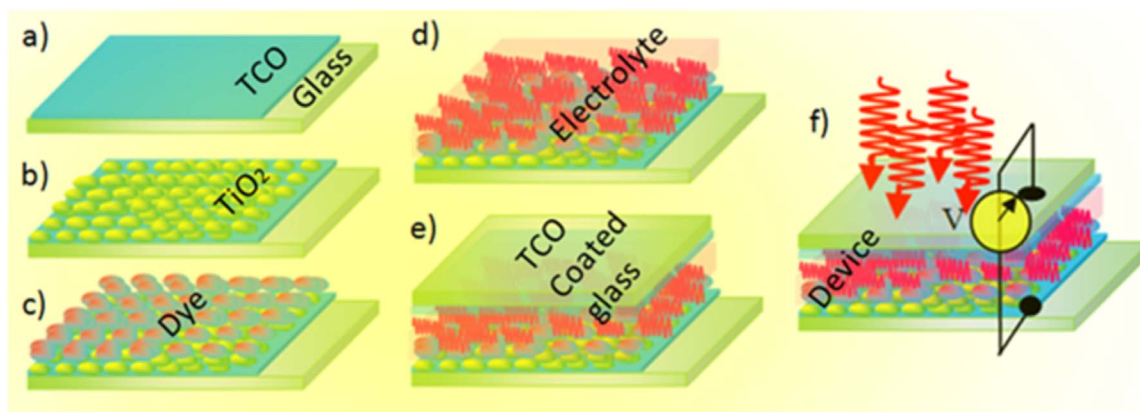


Fig. 13. A schematic representation of the fabrication of dye-sensitized solar cells using various layers placed sequentially (starting with 'a' to 'e'). A complete device structure is shown in figure (f).

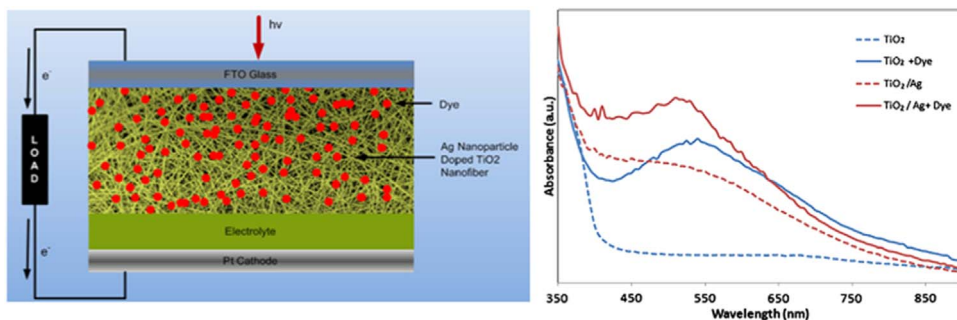


Fig. 14. A schematic representation of Ag doped TiO₂ fiber and dye composite for solar cell device (left). The right side figure represents the enhancement in the absorption due to the incorporation of Ag doing in TiO₂ layer fibers. Figures are adapted from the reference [180].

The enhanced absorption cross section was attributed to the enhanced dephasing time in single-crystalline Ag nanoparticles.

Though, the incorporation of various metal nanoparticles improves the overall dye absorption, the interface between the metal nanoparticles and dye molecules or electrolytes [190–193] may be the cause of corrosion of metal nanoparticles, increased charge recombination, and back reaction of photo-generated carriers leading to loss of charge carriers and poor charge separation at the electrodes. This may lead to lower the overall power conversion efficiency. To overcome these problems, researchers have attempted to put core-shell structures of different shape and size. To minimize the corrosion and recombination of charges, metal nanoparticles are coated with various oxide layers such as with SiO₂ [166,169,170], and TiO₂ [171,194], and its power conversion efficiencies are studied. Recently, Jang et al. [159] have prepared core-shell structures of SiO₂/TiO₂, SiO₂/Au/TiO₂ and SiO₂/TiO₂/Au which were supported on SiO₂ spheres to use as a photo-anode as shown

in the Fig. 15. It shows how the incident photon conversion efficiency (IPCE) spectra get affected by the incorporation of SiO₂/TiO₂, SiO₂/TiO₂/Au, and SiO₂/Au/TiO₂ core-shell nanostructures decorated on TiO₂ anode. One can easily observe the improvement by the incorporation of plasmonic core-shell nanostructures due to enhanced near-fields caused by excitation of surface plasmon resonance of Au nanoparticles. However, oxide shell may reduce the optical absorption as can be found in the references [176,195]. A new type of metal-metal core-shell nanostructures are studied by various research groups in view of improving power conversion efficiency [167,172]. For example, in the report by Dong et al. [167] metal-metal (Au/Ag) core-shell nanorods are studied to show not only the suppression of the charge recombination but also the improvement of the photoresponse due to the surface plasmon resonance effect of the Au/Ag nanorods. The enhanced absorption results in ~40% increase in power conversion efficiencies. Based on the existing literature it can be said that the efficiency

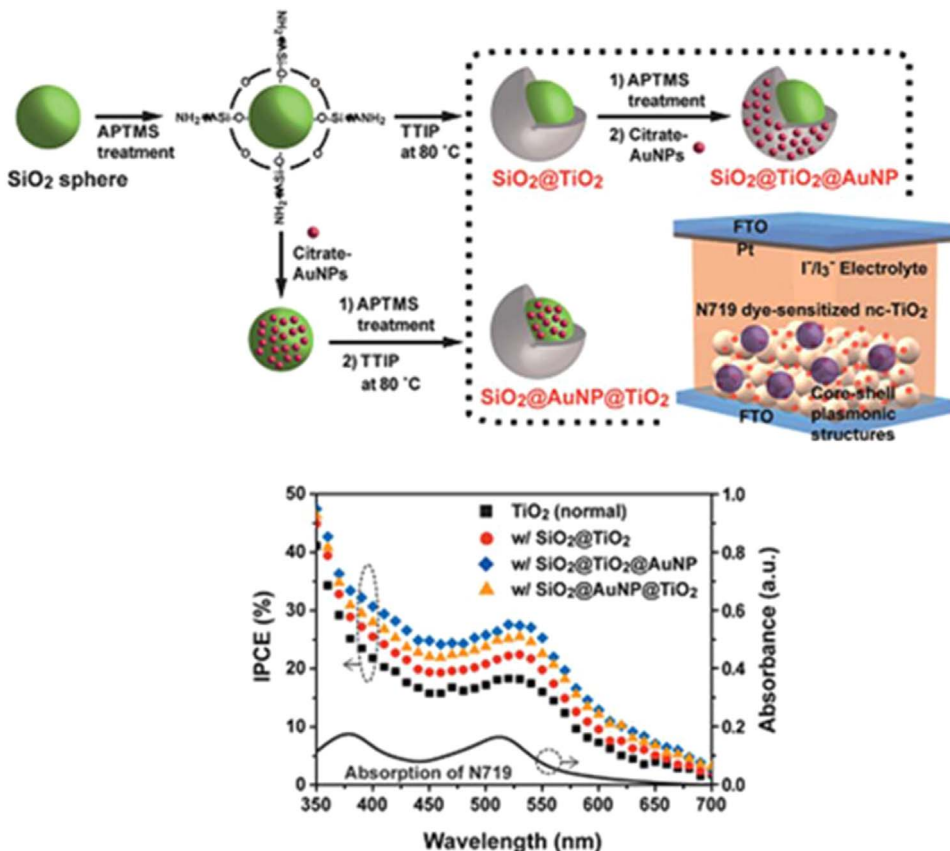


Fig. 15. A schematic representation of core-shell nanoparticles incorporated in dye-sensitized solar cell (top). The effect of incorporation of core-shell structures is observed as a change in IPCE (bottom). Figures are adapted from the reference [159].

improvement of a plasmonic DSSC depends strongly on the intensity of the near-fields caused by localized as well as propagating surface plasmon resonances which crucially depend on the size, shape and structures [196–198]. A great effort is required to achieve the fruitful utilization of the surface plasmon near-fields. On the other hands, discovering new dye-materials and electrolytes may pave the way of finding efficient dye-sensitized solar cells.

6. Future aspects and conclusions

In this review we have shown how plasmonic near-field can be utilized to improve solar photovoltaic device efficiency. The scattering and guiding of light into active materials have been the major concern of plasmonic solar cells. Plasmon near-field arising due to LSP and SPP boosts the light scattering enormously, thereby increasing the chance of light absorption to manifold. Guiding the light inside the active region increases the light path length which helps further light absorption. Front surface patterning, back surface patterning, or both have been utilized to enhance photon harvesting. The composite of metal particles and dyes have been the typical approach for dye-sensitized solar cells. Care should be taken to minimize parasitic absorption by the metal nanoparticles itself.

Plasmon excitation using periodic distribution of particles suffers from polarization and launch angle dependence. 1D pattern requires TM-polarized incident light while for 2D two orthogonal polarizations will excite surface plasmon efficiently. On the other hand, for aperiodic patterns these limitations are not there, although it is to be noted here that designing or fabrication of aperiodic structures is difficult for specific spectral response. Quasi-periodic patterns may be the solutions. Quasi-periodic patterns may offer designable broadband spectral response which is independent of polarization and launch angle [199]. A recent review by Achanta Venu Gopal describes various aspects of quasi-periodic patterns. Interested readers are referred to the reference [200] for further reading on the future prospects of quasiperiodic patterns.

Plasmonic perfect absorber with multiple resonance or broadband resonance can be a potential candidate in this regard. It is also proposed that plasmonic light trapping mechanism along with semiconductor multilayers of different materials having absorption at different wavelengths in solar spectral range can be of good choice for further improvement of cell efficiency. Plasmonic sculptured film can be a good option. Keeping in mind the cost effectiveness, research needs to be carried out on various metals such as aluminum (Al) and copper (Cu).

Acknowledgments

The work was conceptualized and directed by PM. Manuscript sections from 1 to 4, 6 were written by PM. Section-5 was written by SS. PM thanks to K. Pandit for constant encouragement and support.

References

- [1] <http://www.pveducation.org/pvcdrom/appendices/standard-solar-spectra>, as on 15-6-2015.
- [2] Gusak V, Kasemo B, Hägglund C. High aspect ratio plasmonic nanocones for enhanced light absorption in ultrathin amorphous silicon film. *J Phys Chem C* 2014;118:22840–6.
- [3] Theuring M, Wang PH, Vehse M, Steenhoff V, Maydell van K, Agert C, Brolo AG. Comparison of Ag and SiO₂ nanoparticles for light trapping applications in silicon thin film solar cells. *J Phys Chem Lett* 2014;5:3302–6.
- [4] Sharma M, Pudasaini PR, Ruiz-Zepeda F, Vinogradova E, Ayon AA. Plasmonic effects of Au/Ag bimetallic multispiked nanoparticles for photovoltaic applications. *ACS Appl Mater Interfaces* 2014;6:15472–9.
- [5] Li G, Li H, Ho JYL, Wong M, Kwok HS. Nanopyramid structure for ultrathin c-Si tandem solar cells. *Nano Lett* 2014;14:2563–8.
- [6] Wang PH, Millard M, Brolo AG. Optimizing plasmonic silicon photovoltaics with Ag and Au nanoparticle mixtures. *J Phys Chem C* 2014;118:5889–95.
- [7] Zhang A, Guo Z, Tao Y, Wang W, Mao X, Fan G, Zhou K, Qu S. Advanced light-trapping effect of thin-film solar cell with dual photonic crystals. *Nanoscale Res Lett* 2015;10:214.
- [8] Mendes MJ, Morawiec S, Mateus T, Lyubchik A, Águas H, Ferreira I, Fortunato E, Martins R, Priolo F, Crupi I. Broadband light trapping in thin film solar cells with self-organized plasmonic nano-colloids. *Nanotechnology* 2015;26:135202.
- [9] Haug FJ, Ballif C. Light management in thin film silicon solar cells. *Energy Environ Sci* 2015;8:824–37.
- [10] Borra MZ, Güllü SK, Es F, Demircioğlu O, Günöven M, Bek RTA. A feasibility study for controlling self-organized production of plasmonic enhancement interfaces for solar cells. *Appl Surf Sci* 2014;318:43–50.
- [11] Mizuno H, Sai H, Matsubara K, Takato H, Kondo M. Transfer-printed silver nanodisks for plasmonic light trapping in hydrogenated microcrystalline silicon solar cells. *Appl Phys Exp* 2014;7:112302.
- [12] Atwater HA, Polman A. Plasmonics for improved photovoltaic devices. *Nat Mater* 2010;9:205.
- [13] Fahr S, Rockstuhl C, Lederer F. Metallic nanoparticles as intermediate reflectors in tandem solar cells. *Appl Phys Lett* 2009;95:121105.
- [14] Hägglund C, Zäch M, Petersson G, Kasemo B. Electromagnetic coupling of light into a silicon solar cell by nanodisk plasmons. *Appl Phys Lett* 2008;92:053110.
- [15] Beck FJ, Polman A, Catchpole KR. Tunable light trapping for solar cells using localized surface plasmons. *J Appl Phys* 2009;105:114310.
- [16] Catchpole KR, Polman A. Plasmonic solar cells. *Opt Exp* 2008;16:21793–00.
- [17] Cai B, Jia B, Shi Z, Gu M. Near-field light concentration of ultra-small metallic nanoparticles for absorption enhancement in a-Si solar cells. *Appl Phys Lett* 2013;102:093107.
- [18] Shi WB, Fan RH, Zhang K, Xu DH, Xiong X, Peng RW, Wang M. Broadband light trapping and absorption of thin-film silicon sandwiched by trapezoidal surface and silver grating. *J Appl Phys* 2015;117:065104.
- [19] Zhang X, Yu Y, Xi J, Liu T, Sun XH. The plasmonic enhancement in silicon nanocone hole solar cells with back located metal particles. *J Opt* 2015;17:015901.
- [20] Ho WJ, Cheng PY, Hsiao KY. Plasmonic silicon solar cell based on silver nanoparticles using ultra-thin anodic aluminum oxide template. *Appl Surf Sci* 2015. <http://dx.doi.org/10.1016/j.apsusc.2015.05.049>.
- [21] Smeets M, Smirnov V, Meier M, Bittkau K, Carius R, Rau U, Paetzold UW. On the geometry of plasmonic reflection grating back contacts for light trapping in prototype amorphous silicon thin-film solar cells. *J Photon Energy* 2014;5:057004.
- [22] Raether H. Surface plasmons on smooth and rough surfaces and on gratings. Berlin: Springer-Verlag; 1988.
- [23] Maier SA. Plasmonics: fundamentals and applications. New York: Springer; 2007.
- [24] Johnson PB, Christy RW. Optical constants of the noble metals. *Phys Rev B* 1972;6:4073.
- [25] Mandal P, Nandi A, Ramakrishna SA. Propagating surface plasmon resonances in two-dimensional patterned gold-grating templates and surface enhanced Raman scattering. *J Appl Phys* 2012;112:044314.
- [26] Ferry VE, Munday JN, Atwater HA. Design considerations for plasmonic photovoltaics. *Adv Mater* 2010;22:4794–808.
- [27] Yang Z, Li X, Wu S, Gao P, Ye J. High-efficiency photon capturing in ultrathin silicon solar cells with front nanobowl texture and truncated-nanopyramid reflector. *Opt Lett* 2015;40:1077–80.
- [28] Branham MS, Hsu WC, Yerci S, Loomis J, Boriskina SV, Hoard BR, Han SE, Chen G. 15.7% Efficient 10-μm-thick crystalline silicon solar cells using periodic nanostructures. *Adv Mater* 2015;27:2182–8.
- [29] Lin Q, Hua B, Leung SF, Duan X, Fan Z. Efficient light absorption with integrated nanopillar/nanowell arrays for three-dimensional thin-film photovoltaic applications. *ACS Nano* 2013;7:2725–32.
- [30] Battaglia C, Escarre J, Soderstrom K, Erni L, Ding L, Bugnon G, Billet A, Boccard M, Barraud L, De Wolf S, Haug FJ, Despeisse M, Ballif C. Nanoimprint lithography for high-efficiency thin-film silicon solar cells. *Nano Lett* 2011;11:661–5.
- [31] Dong WJ, Yoo CJ, Cho HW, Kim KB, Kim M, Lee JL. Flexible a-Si: H solar cells with spontaneously formed parabolic nanostructures on a hexagonal-pyramid reflector. *Small* 2015;11:1947–53.
- [32] Lin Q, Leung SF, Lu L, Chen X, Chen Z, Tang H, Su W, Li D, Fan Z. Inverted nanocone-based thin film photovoltaics with omnidirectionally enhanced performance. *ACS Nano* 2014;8:6484–90.
- [33] Bermel P, Luo C, Zeng L, Kimerling LC, Ioannopoulos JD. Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals. *Opt Exp* 2007;15:16986–7000.
- [34] Bozzola A, Liscidini M, Andreani LC. Photonic light-trapping versus Lambertian limits in thin film silicon solar cells with 1D and 2D periodic patterns. *Opt Exp* 2012;20:A224–44.
- [35] Zhu J, Hsu CM, Yu Z, Fan S, Cui Y. Nanodome solar cells with efficient light management and self-cleaning. *Nano Lett* 2010;10:1979–84.

- [36] Du QC, Kam CH, Demir HV, Yu HY, Sun XW. Enhanced optical absorption in nanopatterned silicon thin films with a nano-cone-hole structure for photovoltaic applications. *Opt Lett* 2011;36:1713–5.
- [37] Mavrokefalos A, Han SE, Yerci S, Branham MS, Chen G. Efficient light trapping in inverted nanopillar thin crystalline silicon membranes for solar cell applications. *Nano Lett* 2012;12:2792–6.
- [38] Kim SK, Song KD, Park HG. Design of input couplers for efficient silicon thin film solar absorbers. *Opt Exp* 2012;20:A997–1004.
- [39] Meng X, Depauw V, Gomard G, El Daif O, Trompoukis C, Drouard E, Jomois C, Fave A, Dross F, Gordon I, Seassal C. Design, fabrication and optical characterization of photonic crystal assisted thin film monocrystalline-silicon solar cells. *Opt Exp* 2012;20:A465–75.
- [40] Heidarzadeh H, Rostami A, Matloub S, Dolatyari M, Rostami G. Analysis of the light trapping effect on the performance of silicon-based solar cells: absorption enhancement. *Appl Opt* 2015;54:3591–601.
- [41] Yang Y, Green MA, Ho-Bailie A, Kampwerth H, Pillai S, Mehrvarz H. Characterization of 2-D reflection pattern from textured front surfaces of silicon solar cells. *Sol Energy Mater Sol Cells* 2013;115:42–51.
- [42] Gee JM, Schubert WK, Tardy HL, Hund TD, Robison G. The effect of encapsulation on the reflectance of photovoltaic modules using textured multicrystalline-silicon solar cells. In: *Proceedings of the twenty fourth IEEE photovoltaic specialists conference*. 1994; vol. 2: p. 1555–58.
- [43] Baker-Finch SC, McIntosh KR. Reflection of normally incident light from silicon solar cells with pyramidal texture. *Prog Photovolt: Res Appl* 2011;19:406–16.
- [44] Herman A, Trompoukis C, Depauw V, El Daif O, Deparis O. Influence of the pattern shape on the efficiency of front-side periodically patterned ultrathin crystalline silicon solar cells. *J Appl Phys* 2012;112:113107.
- [45] Huang CK, Sun KW, Chang WL. Efficiency enhancement of silicon solar cells using a nano-scale honeycomb broadband anti-reflection structure. *Opt Exp* 2012;20:A85–93.
- [46] Gjessing J, Sudbo AS, Marstein ES. Comparison of periodic light-trapping structures in thin crystalline silicon solar cells. *J Appl Phys* 2011;110:033104.
- [47] Gomard G, Drouard E, Letartre X, Meng X, Fave A, Lemiti M, Garcia-Caurel E, Seassal C. Two-dimensional photonic crystal for absorption enhancement in hydrogenated amorphous silicon thin film solar cells. *J Appl Phys* 2010;108:123102.
- [48] Deparis O, Vigneron J, Agustsson O, Decroupet D. Optimization of photonics for corrugated thin-film solar cells. *J Appl Phys* 2009;106:094505.
- [49] Chang TH, Wu PH, Chen SH, Chan CH, Lee CC, Chen CC, Su YK. Efficiency enhancement in GaAs solar cells using self-assembled microspheres. *Opt Exp* 2009;17:6519–24.
- [50] Chang YA, Li ZY, Kuo HC, Lu TC, Yang SF, Lai LW, Wang SC. Efficiency improvement of single-junction InGaP solar cells fabricated by a novel micro-hole array surface texture process. *Semicond Sci Technol* 2009;24:085007.
- [51] Cho KS, Mandal P, Kim K, Baek IH, Lee S, Lim H, Cho DJ, Kim S, Lee J, Rtermund F. Improved efficiency in GaAs solar cells by 1D and 2D nanopatterns fabricated by laser interference lithography. *Opt Commun* 2011;284:2608–12.
- [52] Zhao J, Wang A, Green MA. 19.8% efficient honeycomb textured multicrystalline and 24.4% monocrystalline silicon solar cells. *Appl Phys Lett* 1998;73:1991.
- [53] Zhao J, Wang A, Altermatt P, Green MA. Twenty-four percent efficient silicon solar cells with double layer antireflection coatings and reduced resistance loss. *Appl Phys Lett* 1995;66:3636.
- [54] Dewan R, Marinkovic M, Noriega R, Phadke S, Salleo A, Knipp D. Light trapping in thin-film silicon solar cells with submicron surface texture. *Opt Exp* 2009;17:23058–65.
- [55] Yerokhov VY, Hezel R, Lipinski M, Ciach R, Nagel H, Mylynych A, Panek P. Cost-effective methods of texturing for silicon solar cells. *Sol Energy Mater Sol Cells* 2002;72:291–8.
- [56] Huang MJ, Yang CR, Chiou YC, Lee RT. Fabrication of nanoporous antireflection surfaces on silicon. *Sol Energy Mater Sol Cells* 2008;92:1352–7.
- [57] Tao M, Zhou W, Yang H, Chen L. Surface texturing by solution deposition for omnidirectional antireflection. *Appl Phys Lett* 2007;91:081118.
- [58] Dobrzański LA, Drygała A. Surface texturing of multicrystalline silicon solar cells. *J Achievements Mater Manuf Engg* 2008;31:77–82.
- [59] Nakayama K, Tanabe K, Atwater HA. Plasmonic nanoparticle enhanced light absorption in GaAs solar cells. *Appl Phys Lett* 2008;93:121904.
- [60] Pillai S, Catchpole KR, Trupke T, Green MA. Surface plasmon enhanced silicon solar cells. *J Appl Phys* 2007;101:093105.
- [61] Akimov YA, Ostrikov K, Li EP. Surface plasmon enhancement of optical absorption in thin-film silicon solar cells. *Plasmonics* 2009;4:107–13.
- [62] Mokkapat S, Beck F, Catchpole K. SPIE newsroom. 2009. <http://dx.doi.org/10.1117/2.1200912.002527>.
- [63] Schaad DM, Feng B, Yu ET. Enhanced semiconductor optical absorption via surface plasmon excitation in metal nanoparticles. *Appl Phys Lett* 2005;86:063106.
- [64] Derkacs D, Lim SH, Matheu P, Mar W, Yu ET. Improved performance of amorphous silicon solar cells via scattering from surface plasmon polaritons in nearby metallic nanoparticles. *Appl Phys Lett* 2006;89:093103.
- [65] Stuart HR, Hall DG. Island size effects in nanoparticle-enhanced photodetectors. *Appl Phys Lett* 1998;73:3815.
- [66] Koh WS, Akimov Y. Enhanced light trapping in thin-film solar cells. SPIE newsroom. 2010. <http://dx.doi.org/10.1117/2.1201003.002654>.
- [67] Luo LB, Xie C, Wang XH, Yu YQ, Wu CY, Hu H, Zhou KY, Zhang XW, Jie JS. Surface plasmon resonance enhanced highly efficient planar silicon solar cell. *Nano Energy* 2014;9:112–20.
- [68] Matheu P, Lim SH, Derkacs D, McPheeters C, Yu ET. Metal and dielectric nanoparticle scattering for improved optical absorption in photovoltaic devices. *Appl Phys Lett* 2008;93:113108.
- [69] Eisele C, Nebel CE, Stutzmann M. Periodic light coupler gratings in amorphous thin film solar cells. *J Appl Phys* 2001;89:7722.
- [70] Alemu N, Chen F. Plasmon-enhanced light absorption of thin-film solar cells using hemispherical nanoparticles. *Phys Status Solidi A* 2014;211:213–8.
- [71] Yun JH, Lee E, Park HH, Kim DW, Anderson WA, Kim J, Litchinitser NM, Zeng J, Yi J, Kumar MMD, Sun J. Incident light adjustable solar cell by periodic nanolens architecture. *Sci Rep* 2014;4:6879.
- [72] Munday JN, Atwater HA. Large integrated absorption enhancement in plasmonic solar cells by combining metallic gratings and antireflection coatings. *Nano Lett* 2011;11:2195–201.
- [73] Kim J, Lee E, Ju M, Kim H, Yi J, Moon SJ, Hyun MS, Kim DW. Surface-concentrated light and efficient carrier collection in microhole-patterned Si solar cells. *Opt Exp* 2013;21:A607–15.
- [74] Zhou Z, Shi L. Optimized design of silicon thin film solar cells with silicon nanogratings. *Optik* 2015;126:614–7.
- [75] Burford N, El-Shenawee M. Optimization of silver nanotoroid arrays for the absorption enhancement of silicon thin-film solar cells. *Plasmonics* 2015;10:225–32.
- [76] Sun C, Wang X. Efficient light trapping structures of thin film silicon solar cells based on silver nanoparticle arrays. *Plasmonics* 2015;10:1307–14. <http://dx.doi.org/10.1007/s11468-015-9934-1>.
- [77] Sun C, Su J, Wang X. A design of thin film silicon solar cells based on silver nanoparticle arrays. *Plasmonics* 2015;10:633–41.
- [78] Diukman I, Orenstein M. How front side plasmonic nanostructures enhance solar cell efficiency. *Sol Energy Mater Sol Cells* 2011;95:2628–31.
- [79] Gou F, Li X, Chen J, Su G, Liu C, Zhang Z. Broadband absorption enhancement in ultrathin-film solar cells by combining dielectric nanogratings and metallic nanoribbons. *J. Nanophoton* 2015;9:093596.
- [80] Schuster CS, Morawiec S, Mendes MJ, Patrini M, Martins ER, Lewis L, Crupi I, Krauss TF. Plasmonic and diffractive nanostructures for light trapping—an experimental comparison. *Optica* 2015;2:194–200.
- [81] Lesina AC, Paternoster G, Mattedi F, Ferrario L, Berini P, Ramunno L, Paris A, Vaccari A, Calliari L. Modeling and characterization of antireflection coatings with embedded silver nanoparticles for silicon solar cells. *Plasmonics* 2015. <http://dx.doi.org/10.1007/s11468-015-9957-7>.
- [82] Winans JD, Hungerford C, Shome K, Rothberg LJ, Fauchet PM. Plasmonic effects in ultrathin amorphous silicon solar cells: performance improvements with Ag nanoparticles on the front, the back, and both. *Opt Exp* 2015;23:A92–105.
- [83] Spinelli P, Hebbink M, de Waele R, Black L, Lenzenmann F, Polman A. Optical impedance matching using coupled plasmonic nanoparticle arrays. *Nano Lett* 2011;11:1760–5.
- [84] Catchpole KR, Polman A. Design principles for particle plasmon enhanced solar cells. *Appl Phys Lett* 2008;93:191113.
- [85] Lim SH, Mar W, Matheu P, Derkacs D, Yu ET. Photocurrent spectroscopy of optical absorption enhancement in silicon photodiodes via scattering from surface plasmon polaritons in gold nanoparticles. *J Appl Phys* 2007;101:104309.
- [86] Yan W, Tao Z, Ong TMB, Gu M. Highly efficient ultrathin-film amorphous silicon solar cells on top of imprinted periodic nano dot arrays. *Appl Phys Lett* 2015;106:093902.
- [87] Kanade P, Yadav P, Kumar M, Tripathi B. Plasmon induced photon manipulation by ag nanoparticle-coupled graphene thin film: light trapping for photovoltaics. *Plasmonics* 2015;10:157–64.
- [88] Tsao YC, Fisker C, Pedersen TG. Nanoimprinted backside reflectors for a-Si: H thin-film solar cells: critical role of absorber front textures. *Opt Exp* 2014;22:A651–62.
- [89] Chen X, Jia B, Zhang Y, Gu M. Exceeding the limit of plasmonic light trapping in textured screen-printed solar cells using Al nanoparticles and wrinkle-like graphene sheets. *Light: Sci Appl* 2013;2:e92.
- [90] Dai H, Li M, Li Y, Yu H, Bai F, Ren X. Effective light trapping enhancement by plasmonic Ag nanoparticles on silicon pyramid surface. *Opt Exp* 2012;20:A502–9.
- [91] Lenzenmann FO, van Lare MC, Salpakari J, Spinelli P, Notta JB, Dörenkämper M, Bakker NJ, Weeber AW, Polman A. Plasmonic light-trapping in a-Si: H solar cells by front-side Ag nanoparticle arrays: a benchmarking study. *Physica Status Solidi A* 2013;210:1571–4.
- [92] Villesen TF, Uhrenfeldt C, Johansen B, Hansen JL, Ulriksen HU, Larsen AN. Aluminum nanoparticles for plasmon-improved coupling of light into silicon. *Nanotechnology* 2012;23:085202.
- [93] Akimov YA, Koh WS, Ostrikov K. Enhancement of optical absorption in thin-film solar cells through the excitation of higher-order nanoparticle plasmon modes. *Opt Exp* 2009;17:10195.
- [94] Dmitruk NL, Borkovskaya OY, Mamontova IB, Mamykin SV, Malynych SZ, Romanyuk VR. Metal nanoparticle-enhanced photocurrent in GaAs photovoltaic structures with microtextured interfaces. *Nanoscale Res Lett* 2015;10:72.
- [95] Han SE, Chen G. Optical absorption enhancement in silicon nanohole arrays for solar photovoltaics. *Nano Lett* 2010;10:1012–5.

- [96] Pala RA, White J, Barnard E, Liu J, Brongersma ML. Design of plasmonic thin-film solar cells with broadband absorption enhancements. *Adv Mater* 2009;21:3504–9.
- [97] Lee YC, Huang CF, Chang JY, Wu ML. Enhanced light trapping based on guided-mode resonance effect for thin-film silicon solar cells with two filling-factor gratings. *Opt Exp* 2008;16:7969–75.
- [98] Chao CC, Wang CM, Chang YC, Chang JY. Plasmonic multilayer structure for ultrathin amorphous silicon film photovoltaic cell. *Opt Rev* 2009;16:343–6.
- [99] Rockstuhl C, Fahr S, Lederer F. Absorption enhancement in solar cells by localized plasmon polaritons. *J Appl Phys* 2008;104:123102.
- [100] Rockstuhl C, Lederer F. Photon management by metallic nanodiscs in thin film solar cells. *Appl Phys Lett* 2009;94:213102.
- [101] Derkacs D, Chen WV, Matheu PM, Lim SH, Yu PKL, Yu ET. Nanoparticle-induced light scattering for improved performance of quantum-well solar cells. *Appl Phys Lett* 2008;93:091107.
- [102] Palanchoke U, Kurz H, Noriega R, Arabi S, Jovanov V, Magnus P, Aftab H, Salleo A, Stiebig H, Knipp D. Tuning the plasmonic absorption of metal reflectors by zinc oxide nano particles: Application in thin film solar cells. *Nano Energy* 2014;6:167–72.
- [103] Pattnaik S, Chakravarty N, Biswas R, Dalal V, Slafer D. Nano-photonic and nano-plasmonic enhancements in thin film silicon solar cells. *Sol Energy Mater Sol Cells* 2014;129:115–23.
- [104] Lee YY, Ho WJ, Chen YT. Performance of plasmonic silicon solar cells using indium nanoparticles deposited on a patterned TiO₂ matrix. *Thin Solid Films* 2014;570:194–9.
- [105] Shi L, Zhou Z, Tang B. Optimization of Si solar cells with full band optical absorption increased in all polarizations using plasmonic backcontact grating. *Optik* 2014;125:789–94.
- [106] de la Morena SS, Recio-Sánchez G, Torres-Costa V, Martín-Palma RJ. Hybrid gold/porous silicon thin films for plasmonic solar cells. *Scr Mater* 2014;74:33–7.
- [107] Wang DH, Kyaw AKK, Du QG, Choi DG, Park JH. Tailoring of the plasmonic and waveguide effect in bulk-heterojunction photovoltaic devices with ordered, nanopatterned structures. *Org Electron* 2014;15:3120–6.
- [108] Islam K, Alnuaimi A, Battal E, Okyay AK, Nayfeh A. Effect of gold nanoparticles size on light scattering for thin film amorphous-silicon solar cells. *Sol Energy* 2014;103:263–8.
- [109] Ferry VE, Verschuuren MA, van Lare MC, Schropp REI, Atwater HA, Polman A. Optimized spatial correlations for broadband light trapping nanopatterns in high efficiency ultrathin film a-si: h solar cells. *Nano Lett* 2011;11:4239.
- [110] Palanchoke U, Jovanov V, Kurz H, Obermeyer P, Stiebig H, Knipp D. Plasmonic effects in amorphous silicon thin film solar cells with metal back contacts. *Opt Exp* 2012;20:6340–7.
- [111] Springer J, Poruba A, Müllerova L, Vanecek M, Kluth O, Rech B. Absorption loss at nanorough silver back reflector of thin-film silicon solar cells. *J Appl Phys* 2004;95:1427.
- [112] Zeman M, van Swaaij RACMM, Metselaar JW, R. Schropp REI. Optical modeling of a-Si: H solar cells with rough interfaces: Effect of back contact and interface roughness. *J Appl Phys* 2000;88:6436–00.
- [113] Sai H, Fujiwara H, Kondo M, Kanamori Y. Enhancement of light trapping in thin-film hydrogenated microcrystalline Si solar cells using back reflectors with self-ordered dimple pattern. *Appl Phys Lett* 2008;93:143501.
- [114] Zeng L, Bermel P, Alamariu BA, Yi Y, Broderick KA, Liu J, Hong C, Duan X, Joannopoulos J, Kimerling LC. Demonstration of enhanced absorption in thin film Si solar cells with textured photonic crystal back reflector. *Appl Phys Lett* 2008;93:221105.
- [115] Tan HR, Santbergen R, Smets AHM, Zeman M. Plasmonic light trapping in thin-film silicon solar cells with improved self-assembled silver nanoparticles. *Nano Lett* 2012;12:4070–6.
- [116] Ferry VE, Verschuuren MA, Li HB, Verhagen E, Walters RJ, Schropp RE, Atwater HA, Polman A. Light trapping in ultrathin plasmonic solar cells. *Opt Exp* 2010;18:A237–45.
- [117] Atalla MRM, Suliman SA. On the propensity of guiding surface-plasmon-polariton waves by the back-contact of an amorphous silicon p-i-n solar cell. *J Opt* 2015;17:045002.
- [118] Ouyang Z, Pillai S, Beck FJ, Kunz O, Varlamov S, Catchpole KR, Campbell P, Green MA. Effective light trapping in polycrystalline silicon thin-film solar cells by means of rear localized surface plasmons. *Appl Phys Lett* 2010;96:261109.
- [119] Beck FJ, Mookapati S, Polman A, Catchpole KR. Asymmetry in photocurrent enhancement by plasmonic nanoparticle arrays located on the front or on the rear of solar cells. *Appl Phys Lett* 2010;96:033113.
- [120] Moulin E, Sukmanowski J, Schulte M, Gordijn A, Royer FX, Stiebig H. Thin-film silicon solar cells with integrated silver nanoparticles. *Thin Solid Films* 2008;516:6813–7.
- [121] Iftiqar SM, Jung J, Park H, Cho J, Shin C, Park J, Jung J, Bong S, Kim S, Yi J. Effect of light trapping in an amorphous silicon solar cell. *Thin Solid Films* 2015;587:117–25.
- [122] Sidharthan R, Murukeshan VM. Improved light absorption in thin film solar cell using combination of gap modes and grating back reflector. *Thin Solid Films* 2013;548:581–4.
- [123] Yin Y, Yu Z, Liu Y, Ye H, Zhang W, Cui Q, Yu X, Wang P, Zhang Y. Design of plasmonic solar cells combining dual interface nanostructure for broadband absorption enhancement. *Opt Commun* 2014;333:213–8.
- [124] Sai H, Kondo M. Light trapping effect of patterned back surface reflectors in substrate-type single and tandem junction thin-film silicon solar cell. *Sol Energy Mater Sol Cells* 2011;95:131–3.
- [125] Zheng GG, Zhang W, Xu LH, Chen YY, Liu YZ. Absorbance enhancement of thin film solar cells with front double dielectric and back metallic grating. *Infra Phys Technol* 2014;67:52–7.
- [126] Lin HY, Kuo Y, Liao CY, Yang CC, Kiang YW. Surface plasmon effects in the absorption enhancements of amorphous silicon solar cells with periodical metal nanowall and nanopillar structures. *Opt Exp* 2011;20:A104–18.
- [127] Paetzold UW, Moulin E, Michaelis D, Boettler W, Waechter C, Hagemann V, Meier M, Carius R, Rau U. Plasmonic reflection grating back contacts for microcrystalline silicon solar cells. *Appl Phys Lett* 2011;99:181105.
- [128] Paetzold UW, Moulin E, Pieters BE, Carius R, Rau U. Design of nanostructured plasmonic back contacts for thin-film silicon solar cells. *Opt Exp* 2011;19:A1219–30.
- [129] Paetzold UW, Moulin E, Pieters BE, Rau U, Carius R. Optical simulations of microcrystalline silicon solar cells applying plasmonic reflection grating back contacts. *J Photon Energy* 2012;2:027002.
- [130] Santbergen R, Temple TL, Liang R, Smets AHM, van Swaaij RACMM, Zeman M. Application of plasmonic silver island films in thin-film silicon solar cells. *J Opt* 2012;14:024010.
- [131] Paetzold UW, Meier M, Moulin E, Smirnov V, Pieters BE, Rau U, Carius R. Plasmonic back contacts with non-ordered Ag nanostructures for light trapping in thin-film silicon solar cells. *Mater Sci Engg B* 2013;178:630–4.
- [132] Moulin E, Sukmanowski J, Luo P, Carius R, Royer F, Stiebig H. Improved light absorption in thin-film silicon solar cells by integration of silver nanoparticles. *J Non-Cryst Solids* 2008;354:2488–91.
- [133] Haug FJ, Söderström T, Cubero O, Terrazzoni-Daudrix V, Ballif C. Plasmonic absorption in textured silver back reflectors of thin film solar cells. *J Appl Phys* 2008;104:064509.
- [134] Zhao H, Öztürk B, Schiff EA, Sivec L, Yan B, Yang J, Guha S. Back reflector morphology effects and thermodynamic light-trapping in thin-film silicon solar cells. *Sol Energy Mater Sol Cells* 2014;129:104–14.
- [135] Morawiec S, Mendes MJ, Filonovich SA, Mateus T, Mirabella S, Águas H, Ferreira I, Simone F, Fortunato E, Martins R, Priolo F, Crupi I. Broadband photocurrent enhancement in a-Si: H solar cells with plasmonic back reflectors. *Opt Exp* 2014;22:A1059–70.
- [136] Jang J, Kim M, Kim Y, Kim K, Baik SJ, Lee H, Lee JC. Three dimensional a-Si: H thin-film solar cells with silver nano-rod back electrodes. *Curr Appl Phys* 2014;14:637–40.
- [137] Mendes MJ, Morawiec S, Simone F, Priolo F, Crupi I. Colloidal plasmonic back reflectors for light trapping in solar cells. *Nanoscale* 2014;6:4796–05.
- [138] Pahud C, Isabella O, Naqavi A, Haug FJ, Zeman M, Herzog HP, Ballif C. Plasmonic silicon solar cells: impact of material quality and geometry. *Opt Exp* 2013;21:A786–97.
- [139] Morawiec S, Mendes MJ, Mirabella S, Simone F, Priolo F, Crupi I. Self-assembled silver nanoparticles for plasmon-enhanced solar cell back reflectors: correlation between structural and optical properties. *Nanotechnol* 2013;24:265601.
- [140] Tan H, Sivec L, Yan B, Santbergen R, Zeman M, Smets AHM. Improved light trapping in microcrystalline silicon solar cells by plasmonic back reflector with broad angular scattering and low parasitic absorption. *Appl Phys Lett* 2013;102:153902.
- [141] Bai W, Gan Q, Bartoli F, Zhang J, Cai L, Huang Y, Song G. Design of plasmonic back structures for efficiency enhancement of thin-film amorphous Si solar cells. *Opt Lett* 2009;34:3725–7.
- [142] Ferry VE, Verschuuren MA, Li HBT, Schropp REI, Atwater HA, Polman A. Improved red-response in thin film a-Si: H solar cells with soft-imprinted plasmonic back reflectors. *Appl Phys Lett* 2009;95:183503.
- [143] Sai H, Fujiwara H, Kondo M. Back surface reflectors with periodic textures fabricated by self-ordering process for light trapping in thin-film microcrystalline silicon solar cells. *Sol Energy Mater Sol Cells* 2009;93:1087–90.
- [144] Mookapati S, Beck FJ, Polman A, Catchpole KR. Designing periodic arrays of metal nanoparticles for light-trapping applications in solar cells. *Appl Phys Lett* 2009;95:053115.
- [145] Ferry VE, Sweatlock LA, Pacifici D, Atwater HA. Plasmonic nanostructure design for efficient light coupling into solar cells. *Nano Lett* 2008;8:4391–7.
- [146] Hao J, Wang J, Liu X, Padilla WJ, Zhou L, Qiu M. High performance optical absorber based on a plasmonic metamaterial. *Appl Phys Lett* 2010;96:251104.
- [147] Aslam MI, Ali SM. A wideband metamaterial absorber for solar cell applications. In: *Proceedings of International Conference on Energy and Sustainability*. 2013, p. 113–116.
- [148] Wang H, Wang L. Perfect selective metamaterial solar absorbers. *Opt Exp* 2013;21:A1078–93.
- [149] Pors A, Uskov AV, Willatzen M, Protsenko IE. Control of the input efficiency of photons into solar cells with plasmonic nanoparticles. *Opt Commun* 2011;284:2226–9.
- [150] O'Regan B, Gratzel M. Low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO₂ films. *Nature* 1991;353:737–40.
- [151] Green MA, Emery K, Hishikawa Y, Warta W, Dunlop ED. Solar cell efficiency tables (version 44). *Prog Photovolt: Res Appl* 2014;22:701–10.
- [152] Cong J, Liang J, Sumathy K. Review on dye-sensitized solar cells (DSSCs): Fundamental concepts and novel materials. *Renew Sustain Energy Rev* 2012;16:5848–60.
- [153] Hamann TW, Jensen RA, Martinson ABF, Ryswyk HV, Hupp JT. Advancing beyond current generation dye-sensitized solar cells. *Energy Environ Sci* 2008;1:66–78.
- [154] Lim SP, Pandikumar A, Huang NM, Lim HN. Enhanced photovoltaic performance of silver/titanium plasmonic photo anode in dye-sensitized solar cells. *RSC Adv* 2014;4:38111–8.

- [155] Schaadt DM, Feng B, Yu ET. Enhanced semiconductor optical absorption via surface plasmon excitation in metal nanoparticles. *Appl Phys Lett* 2005;86:063106.
- [156] Wen C, Ishikawa K, Kishima M, Yamada K. Effects of silver particles on the photovoltaic properties of dye-sensitized TiO_2 thin films. *Sol Energy Mater Sol Cells* 2000;61:339–51.
- [157] Zhu K, Neale NR, Miedaner A, Frank AJ. Enhanced charge-collection efficiencies and light scattering in dye-sensitized solar cells using oriented TiO_2 nanotubes arrays. *Nano Lett*. 2007;7:69–74.
- [158] Kuang D, Brilliet J, Chen P, Takata M, Uchida S, Miura H, Sumioka K, Zakeeruddin SM, Grätzel M. Application of highly ordered TiO_2 nanotube arrays in flexible dye-sensitized solar cells. *ACS Nano* 2008;2:1113.
- [159] Jang YH, Jang YJ, Kochuveedu ST, Byun M, Lin Z, Kim DH. Plasmonic dye-sensitized solar cells incorporated with Au– TiO_2 nanostructures with tailored configurations. *Nanoscale* 2014;6:1823–32.
- [160] Wang P, Zakeeruddin SM, Moser JE, Nazeeruddin MK, Sekiguchi T, Agrätzel M. Stable quasi-solid-state dye-sensitized solar cell with an amphiphilic ruthenium sensitizer and polymer gel electrolyte. *Nat Mater* 2003;2:402–7.
- [161] Grätzel M. Recent advances in sensitized mesoscopic solar cells. *Acc Chem Res* 2009;42:1788–98.
- [162] Hagglund C, Zach M, Kasemo B. Enhanced charge carrier generation in dye sensitized solar cells by nanoparticle plasmons. *Appl Phys Lett* 2008;92:013113.
- [163] Park JT, Chi WS, Jeon H, Kim J. Improved electron transfer and plasmonic effect in dye-sensitized solar cells with bi-functional Nb doped TiO_2/Ag ternary nanostructures. *Nanoscale* 2014;6:2718–29.
- [164] Lou Y, Yuan S, Zhao Y, Hu P, Wang Z, Zhang M, Shi L, Li D. Molecular-scale interface engineering of metal nanoparticles for plasmon-enhanced dye sensitized solar cells. *Dalton Trans* 2013;42:5330–7.
- [165] Gangishetty MK, Scott RWJ, Kelly TL. Panchromatic enhancement of light-harvesting efficiency in dye-sensitized solar cells using thermally annealed Au@ SiO_2 triangular nanoprisms. *Langmuir* 2014;30:14352–9.
- [166] Brown MD, Suteewong T, Kumar RSS, D'Innocenzo V, Petrozza A, Lee MM, Wiesner U, Snaith HJ. Plasmonic dye-sensitized solar cells using core-shell metal-insulator nanoparticles. *Nano Lett* 2011;11:4384–9.
- [167] Dong H, Wu Z, El-Shafei A, Xia H, Xi J, Ning S, Jiao B, Hou X. Ag-encapsulated Au plasmonic nanorods for enhanced dye-sensitized solar cell performance. *J Mater Chem A* 2015;3:4659–68.
- [168] Chang S, Li Q, Xiao X, Wong KY, Chen T. Enhancement of low energy sunlight harvesting in dye-sensitized solar cells using plasmonic gold nanorods. *Energy Environ Sci* 2012;5:9444–8.
- [169] Choi H, Chen WT, Kamat PV. Know thy nano neighbor. Plasmonic versus electron charging effects of metal nanoparticles in dye-sensitized solar cells. *ACS Nano* 2012;6:4418–27.
- [170] Guo K, Li M, Fang X, Liu X, Zhu Y, Hu Z, Zhao X. Enhancement of properties of dye-sensitized solar cells by surface plasmon resonance of Ag nanowire core-shell structure in TiO_2 films. *J Mater Chem A* 2013;1:7229–34.
- [171] Qi J, Dang X, Hammond PT, Belcher AM. Highly efficient plasmon-enhanced dye-sensitized solar cells through metal@oxide core-shell nanostructure. *ACS Nano* 2011;5:7108–16.
- [172] Yun J, Hwang SH, Jang J. Fabrication of Au@Ag core/shell nanoparticles decorated TiO_2 hollow structure for efficient light-harvesting in dye-sensitized solar cells. *ACS Appl Mater Interfaces* 2015;7:2055–63.
- [173] Kawawaki T, Takahashi Y, Tatsuma T. Enhancement of dye-sensitized photocurrents by gold nanoparticles: effects of plasmon coupling. *J Phys Chem C* 2013;117:5901–7.
- [174] Kawawaki T, Takahashi Y, Tatsuma T. Enhancement of dye-sensitized photocurrents by gold nanoparticles: effect of dye-particle spacing. *Nanoscale* 2011;3:2865–7.
- [175] Reddy KG, Deepak TG, Anjuresee GS, Thomas S, Vadukunpully S, Subramanian KRV, Nair SV, Nair AS. On global energy scenario, dye-sensitized solar cells and the promise of nanotechnology. *Phys Chem Chem Phys* 2014;16:6838–58.
- [176] Standridge SD, Schatz GC, Hupp JT. Distance dependence of plasmon-enhanced photocurrent in dye-sensitized solar cells. *J Am Chem Soc* 2009;131:8407–9.
- [177] Sheehan SW, Noh H, Brudvig GW, Cao H, Schmittenmaier CA. Plasmonic enhancement of dye-sensitized solar cells using core-shell nanostructures. *J Phys Chem C* 2013;117:927–34.
- [178] Ding IK, Zhu J, Cai W, Moon SJ, Cai N, Wang P, Zakeeruddin SM, Grätzel M, Brongersma ML, Cui Y, McGehee MD. Plasmonic dye-sensitized solar cells. *Adv Energy Mater* 2011;1:52–7.
- [179] Jung H, Koo B, Kim JY, Kim T, Son HJ, Kim BS, Kim JY, Lee DK, Kim H, Cho J, Ko MJ. Enhanced photovoltaic properties and long-term stability in plasmonic dye-sensitized solar cells via noncorrosive redox mediator. *ACS Appl Mater Interfaces* 2014;6:19191–200.
- [180] Li J, Chen X, Ai N, Hao J, Chen Q, Strauß S, Shi Y. Silver nanoparticle doped TiO_2 nanofiber dye sensitized solar cells. *Chem Phys Lett* 2011;514:141–5.
- [181] Muduli S, Game O, Dhas V, Vijayamohan K, Bogle KA, Valanoor N, Ogale SB. TiO_2 -Au plasmonic nanocomposite for enhanced dye-sensitized solar cell (DSSC) performance. *Sol Energy* 2012;86:1428–34.
- [182] Goubet N, Yan C, Polli D, Portals H, Arfaoui I, Cerullo G, Pileni MP. Modulating physical properties of isolated and self-assembled nanocrystals through change in nanocrystallinity. *Nano Lett* 2013;13:504–8.
- [183] Tanvi, Mahajan A, Bedi RK, Kumar S, Saxena V S, Aswal DK. Effect of the crystallinity of silver nanoparticles on surface plasmon resonance induced enhancement of effective absorption cross-section of dyes. *J Appl Phys* 2015;117:083111.
- [184] Amiri O, Salavati-Niasari M, Farangi M. Enhancement of dye-sensitized solar cells performance by core shell Ag@organic (organic = 2-nitroaniline, PVA, 4-chloroaniline and PVP): effects of shell type on photocurrent. *Electrochim Acta* 2015;153:90–6.
- [185] Son S, Hwang SH, Kim C, Yun YJ, Jang J. Designed synthesis of $\text{SiO}_2/\text{TiO}_2$ core/shell structure as light scattering material for highly efficient dye-sensitized solar cells. *ACS Appl Mater Interfaces* 2013;5:4815–20.
- [186] Atalla MRM. Plasmonic absorption enhancement in a dye-sensitized solar cell using a Fourier harmonics grating. *Plasmonics* 2015;10:151–6.
- [187] Tripathi SK, Rani M, Singh N. ZnO:Ag and TZO:Ag plasmonic nanocomposite for enhanced dye sensitized solar cell performance. *Electrochim Acta* 2015;167:179–86.
- [188] Kim HY, Rho WY, Lee HY, Park YS, Suh JS. Aggregation effect of silver nanoparticles on the energy conversion efficiency of the surface plasmon-enhanced dye-sensitized solar cells. *Sol Energy* 2014;109:61–9.
- [189] Du J, Qi J, Wang D, Tang Z. Facile synthesis of Au@ TiO_2 core-shell hollow spheres for dye-sensitized solar cells with remarkably improved efficiency. *Energy Environ Sci* 2012;5:6914–8.
- [190] Wen C, Ishikawa K, Kishima M, Yamada K. Effects of silver particles on the photovoltaic properties of dye-sensitized TiO_2 thin films. *Sol Energy Mater Sol Cells* 2000;61:339–51.
- [191] Ihara M, Tanaka K, Sakaki K, Honma I, Yamada K. Enhancement of the Absorption Coefficient of cis-(NCS) $_2$ Bis(2,20-bipyridyl-4,40-dicarboxylate)ruthenium(II) Dye in Dye-Sensitized Solar Cells by a Silver Island Film. *J Phys Chem B* 1997;101:5153–7.
- [192] Ishikawa K, Wen CJ, Yamada K, Okubo T. The photocurrent of dye-sensitized solar cells enhanced by the surface plasmon resonance. *J Chem Eng Jpn* 2004;37:645–9.
- [193] Törngren B, Akitsu K, Ylinen A, Sandén S, Jiang H, Ruokolainen J, Komatsu M, Hamamura T, Nakazaki J, Kubo T, Segawa H, Österbacka R, Småtth JH. Investigation of plasmonic gold-silica core-shell nanoparticle stability in dye-sensitized solar cell applications. *J Colloid Interface Sci* 2014;427:54–61.
- [194] Jeong NC, Prsittichai C, Hupp JT. Photocurrent enhancement by surface plasmon resonance of silver nanoparticles in highly porous dye-sensitized solar cells. *Langmuir* 2011;27:14609–14.
- [195] Chen H, Blaber MG, Standridge SD, DeMarco EJ, Hupp JT, Ratner MA, Schatz GC. Computational modeling of plasmon-enhanced light absorption in a multicomponent dye sensitized solar cell. *J Phys Chem C* 2012;116:10215–21.
- [196] Deepa KG, Lekha P, Sindhu S. Efficiency enhancement in DSSC using metal nanoparticles: a size dependent study. *Sol Energy* 2012;86:326–30.
- [197] Chander N, Khan AF, Thouti E, Sardana SK, Chandrasekhar PS, Dutta V, Komarala VK. Size and concentration effects of gold nanoparticles on optical and electrical properties of plasmonic dye sensitized solar cells. *Sol Energy* 2014;109:11–23.
- [198] Zhang D, Wang M, Brolo AG, Shen J, Li X, Huang S. Enhanced performance of dye-sensitized solar cells using gold nanoparticles modified fluorine tin oxide electrodes. *J Phys D: Appl Phys* 2013;46:024005.
- [199] Kasture S, Ajith PR, Yallapragada VJ, Patil R, Nikesh VV, Muley G, Gopal AV. Plasmonic quasicrystals with broadband transmission enhancement. *Sci. Rep* 2014;4:5257.
- [200] Gopal AV. Plasmonic quasicrystals. *Prog Quantum Electron* 2015;39:1–23.

Proliferating Migration of Uttarakhand's Youth – Reasons, Remedies and Recommendations

University Of Petroleum and Energy Studies

Author Guides:

1.) Dr Vickram Sahai

Associate Professor

2.) Dr Rati Oberoi

Assistant Professor (S.S.)

Student Authors:

1. Satvik Shukla (B.Tech – Electronics Engineering: 2nd Year)

2. Tanisha Gupta (B.Tech – Electronics Engineering: 2nd Year)

**3. Ritika Raj (B.Tech – Computer Science with Specialization in Oil &
Gas Informatics: 3rd Year)**

**4. Aatreyee Dhar (B. Tech – Computer Science with specialization in IT
Infrastructure: 3rd Year)**

Introduction

The soul of India lives in its villages, this was the observation made by father of the nation, Mahatma Gandhi, many years ago. India is a country where nearly two-thirds of its people reside in villages. But this rural and urban population distribution pattern of India is witnessing a change where the urban population and land use is consistently increasing. As is with most of the Indian states, a considerable proportion of Uttarakhand's population also resides in the rural areas. It is primarily a rural state, having 16,826 villages, with a considerable proportion of its population residing in the hilly rural areas. The hill areas offer tough working and living conditions for its inhabitants. These areas do not provide conducive terrains for the establishment of small-scale industries (vis-à-vis handicraft industries or cottage industries) to generate

employment and other sources of income, due to which the people here opt for agriculture and the other activities associated with it. The state is currently witnessing a brew, which is socio-economic, i.e. “Intra-State Migration” of the youth from the hilly areas to the cities. Of the total number of villages, 12,699 or 81% have a population of less than 500. Of the 13 districts in Uttarakhand, most of the districts, more than 75-85 % of rural settlements have a population of less than 500. Only 17 % of the rural settlements have a population ranging between 500-1999 and the villages with population of 2000 or more are very rare (2.7 %). The continuously increasing population pressure on land and the poor agriculture status has led to widespread migration. However, migration is not new to the state of Uttarakhand but the current picture, which is very unfavorable, has made it a frontline issue. The Census of 2011, show that out of Uttarakhand’s 16,793 villages, 1053 have no inhabitants turning them into “Ghost Villages” and another 405 have a population of less than 10 (Venkatesh S, 2016). The migration is so substantial that the economy of the state is termed as money order economy. Lack of basic infrastructure, limited communication, income disparities, religious persecution, desire for better life drive people out of hill areas to the cities. This ensues barren landholdings, reduced labor force, underdevelopment, ghost villages, food insecurity, economic loss etc. The workers engaged in agriculture face many ordeals due to the adverse environmental and geographical configuration of these areas. This has led to poor agricultural conditions marked by problems of poor technology, lack of irrigation facilities and poor land/landholdings etc. Due to the migration of men to the cities which further increases the adversities on women, which are not only confined to the boundaries of domestic chores but also their roles have expanded to such an extent that they have become the backbone of the economy. This is further reflected in the growing literacy rate, awareness level and political indulgence among them. The problem of migration is crucial from the strategic point, too, as out of the 625km. long international border, 350 km. is shared with China and another 293km. is shared with Nepal, i.e, of the 13 districts of Uttarakhand, five are border districts constituting 47% of the area. A well-articulated strategy and effective implementation of already existing policies is fundamental to elevate the levels of development in rural sectors, thus plummeting the levels of migration. Setting up SEZs (Special Economic Zones) in rural areas with the availability of the raw material and labor, employment through MGNREGS, skill training through NSDCs, encouraging agricultural diversification like

horticulture, organic farming, medical tourism, infrastructure development etc. can accelerate development in countryside areas and prove crucial in minimizing migration.

Aim

To study the factors leading to the intra-State migration of youth from the hilly, rural areas to the cities in Uttarakhand and the migration across the international border, lining Uttarakhand, and propose feasible solutions to check the same.

Statement of Purpose

A qualitative approach towards understanding the causes behind migration among youth of Uttarakhand, role of the new policy by the state to tackle this frontline issue and proposing remedies and recommendations to the state government that can help to implement suitable measures in future.

Objectives

Keeping in perspective the widespread migration of men to the cities with the root cause including obscure agricultural condition, the study aims to do an in-depth analysis of the following aspects related to the issue:

- Agricultural profile of Uttarakhand.
- Causes responsible for the current scenario and their simultaneous effects.
- Solutions and recommendations to effectively deal with the situation.

Research Question

- On what basis, can one state that the agricultural condition of a village or a hilly area is a sound one?

Literature Review

The literature review revolves mainly on the economy of the hilly areas and the primary dependence of the population on agriculture. The significance of this occupation does not resonate in terms of income generation but the potential to engage a major portion of the

population in the same. Agriculture's contribution to the state domestic product is 22.4% and about 75-85% of the population is engaged in this occupation. Due to the industrialization constraints in the hilly areas, the service sector, too, is not developed and, thus, the bulk of focus is on the agricultural sector and its prosperity. A large part of the state is characterized by a difficult terrain, undulating topography, sparse population, agro-based economy and degraded forest area. Around 30% of the forest area in the hills is degraded. There are plans to promote cultivate Jatropha and bamboo on this land, which will help in the effective dealing of the problem. Due to physical, geographical and environmental reasons, the scope of input-intensive agricultural policies is severely constrained. This compels its inhabitants to sustain on agriculture or migrate to the cities for employment. The state is currently facing a major challenge to retain the local people through employment opportunities and, consequently, income generation to enhance their quality of life. The land pattern in such regions is very complex. More than half of the land is occupied by the terraced slopes (80%), the remaining portion (20%) lies in the valleys and the plains, which is fairly irrigated.

On the basis of various studies and researches, the literature review indicates that despite playing a key role in the economy, the agriculture of the state is not at all developed and is accentuated due its tough and, at times, adverse geographical conditions.

Causes leading to the intra-state migration in Uttarakhand

- 1. Agricultural problems:** Agriculture is one of the main sources of income and subsistence in Uttarakhand. A two-day open discussion on extensive migration from the hills of Uttarakhand, chaired by Assembly Speaker Govind Singh Kunjwal and former Almora MP Pradeep Tamta, concluded at Pauri. Tamta said agriculture had been the main source of livelihood in the villages for ages but the gradual disenchantment with it triggered the migration (Moudgil A. 2015). He feels revival of agriculture would play a key role in curtailing the exodus and bringing the people back to their native land. The government is in a process of bringing an act on Chakbandi (land consolidation), which would be effective in bringing the greenery back to our villages, he added. But the interest of people in farming is continuously decreasing due to various factors:

- (i) **Scattered landholdings:** In the hilly areas, the agricultural land is scattered, which means it becomes very difficult for the peasant/farmer to execute and manage any agriculture or any other process over it. Another factor observed was the shift in the joint family structure which has caused the landholdings to grow smaller in size through consecutive generations making it difficult to carry out any profitable agricultural activity. Uneconomic land holdings also make it difficult for growing families to sustain themselves, as according to government data as high as 72% are under the category of marginal holdings and 47% are below the size of 0.5 hectares. (Joshi S. 2012) Change in the dynamics of the household, as a result of absent family members who had migrated has led to this. Traditionally, joint families are better able to manage collective resources, such as cattle, due to the availability of labor.

- (ii) **Scarcity of water:** One factor is the lack of water. What usually happens is that an uphill family diverts most of the water from the local guhl (small stream) in order to begin rice cultivation for transplantation, and downhill families suffer the consequence of significantly less water. Thus, the uphill families involved in farming of crops, like rice, which need plenty of water find such alternatives to water their fields which leads to water problems for the downhill people. Thus, availability of water to irrigate fields is a problem which further gives rise to problems of basic needs compelling people to move to other suitable areas. The state faces a shortage of clean water and a way needs to be found to get clean water from the rivers and streams of the state.

- (iii) **Antediluvian Technological Methods:** Use of primitive methods affects the yield to a considerable extent thereby making the problems even worse for the farmers. The women of the state work throughout the year and are able to get food for only 2-3 months. There is a need to develop and use methods for a better produce from the same land.

- (iv) **Lack of investment in agriculture due to limited finances:** No major efforts have been made to improve agriculture and its productivity through its diversification in favor of high value crops. In fact, farmers are willing to diversify their farm production to improve their income, but the required support of agricultural extension services like development of irrigation channels and use of better seeds, improved inputs, technology and marketing remain a major concern.
- (v) **Barren Land or Landholdings:** Landholdings in Uttarakhand are typically small and segmented. The villages that have witnessed migration in the recent past now have to deal with several plots of untended land interspersed with active farmland. Untended land turns barren or is covered with resilient weeds and shrubs (such as lantana and parthenium) that are very difficult to clear. Moreover, such land is being increasingly managed by immigrants from Nepal.

2. Lack of basic amenities in villages (Kasniyal B D, 2015): People have always moved in search of work, in response to environmental shocks and stresses, to escape religious persecution and political conflict. However, improved communications, transport networks, conflict over natural resources and new economic opportunities have created unprecedented levels of mobility.

- (i) **Lack of proper healthcare facilities:** The healthcare centers that have opened are blighted by severe lack of medical professionals and serve, more often than not, as referral centers to hospitals in cities, such as Dehradun and Nainital. With much of the young and able-bodied youth having migrated, it is the elderly who have to live with shoddy healthcare facilities. And they, too, want to move out.
- (ii) **Lack of proper educational institutions:** Access to primary education has also improved significantly, with all hill districts having at least one primary school for every two villages. But, similar growth is not visible in the number of high schools in hilly areas. This means that most villages have a de facto urban dependency if they want good education. Villages do not have a good school system; schools are too far away and it is very difficult to study there. Also, the

quality of education in such schools is pathetic. The students do not pass their tenth grade examinations and ultimately move to other places either out of shame or in search of livelihood.

- (iii) **Lack of basic infrastructure:** The attraction to cities arising due to hardships of village life in hills such as poor transport connectivity, lack of water, networking and inaccessible markets have further accelerated the process of migration of youth. Many areas are remote leading to reduced access to quality and modern life. Government data shows that around 2000 villages in hill districts are situated at the distance of 5 km or more from road head. Another cause because of which there is a problem of providing basic infrastructure is the focus on prevention and conservation of forests (Forest Conservation Act) which has caused delay or led to stopping of approximately 200 development projects. (Joshi S. 2012)
- (iv) **Lack of employment and skills required for satisfactory employment:** Theyouth who migrate for education have no motivation to come back to the hills due to no satisfactory employment option. There are hardly any industries where these people can be employed. There is lack of self-employment due to the paucity of government initiatives to promote tourism, agriculture and other small scale opportunities in the hills for the youth as due to migration the art or handicraft (for eg: Aipan, a Kumaoni region handicraft) that could have been passed on to the next generation and become a developed small scale industry and, thus, leading to diminished income making people disconnected from their origin and culture. The medical officers posted in rural areas are either not regular or do not find suitable infrastructure to work. The rest of the people who are unskilled do not have adequate finances to get skilled, ultimately finding no other way but to move to cities to acquire skill training and jobs. The villages have very few small scale industries and cottage industries which could have served as easy employment centers for village women and other people. Income disparities in rural and urban areas further add to the woes of rural people. They feel it useless to work in villages because they do not get equal rights as do the people in cities.
- (v) **Wildlife threat:** The land pattern of the state shows that the forests occupy dominant proportion of the land and cover more than half of the reported area in

the state. The risk of human-animal conflict creates a sense of insecurity among the inhabitants giving them another reason to migrate to the cities where they are safe. A lot many children die every year due to leopard prowling. The small amount of agriculture that is carried on also faces danger from the monkey menace that constantly destroy the harvest and the lack of young labor to effectively guard this is also a problem.

- (vi) **Poverty:** Inhabitants of the hilly areas are deprived of the resources required for cultivation. They don't have land which they can use to mortgage for agriculture. The per capita income is very less and lack of knowledge of the developed world outside the hills inhibits the growth of these people who still follow the old school. Also, deficiency of development policies in the hills fails to increase their standard of living.
- (vii) **Natural calamities:** The mass erosion of the land where the villages are located and the fact that they could be wiped out by the landslides which are caused by incessant rains and other natural calamities like flash floods have also become a reason due to which the youth are migrating to greener pastures, abandoning their villages which should otherwise be relocated. Around 435 villages have been identified in past that are in need to be relocated. In the recent June 16 floods, 4000 villages have been badly hit and another 60 have been devastated. Also, the earthquake which hit the Uttarkashi region in 1991 caused a lot of migration and even after 13 years the relocation of the people affected has not been completed. (Joshi S. 2012)
- (viii) **Lopsided economic development:** One of the crucial reasons for the unbridled outmigration is the inequality in the economic growth between the hill districts and the plains propelled by underdevelopment in the hills owing to lack of infrastructure like roads, irrigation and electricity, aggravated by geographical disparity. Confirming to the report as per the State's Directorate of Economics and Statistics (Annual Plan 2013-14), only one of the hill districts sees an average per capita income more than the State and the three major plain districts seem to hog most of economy. (Venkatesh, S. 2016) Lack of healthcare facilities in the hill districts propels migration of villagers to the plains leaving the elderly who

are forced to follow the younger generation as they cannot be abandoned with run-down and discreditable medical care which is evident from what one of the 11 residents of Bandul, a village near Pauri who hopelessly prepares for the road only travelled has to say,” “My son has shifted to Gurgaon where he earns about Rs. 7,000 a month. I have begun having health troubles but there are no medical centres nearby. Given the chance, I would move to a town.(Venkatesh, S. 2016)

Solutions to Effectively Tackle Intra-State Migration in Uttarakhand

- 1. Consolidation of land and land holdings:** Efforts and steps as undertaken by Project Chakbandi should be implemented for better and more profitable utilization of the land. Ganesh Gareeb was the pioneer for this project who realized the importance of land consolidation and through his continuous efforts, when Uttarakhand became a state, a Land Consolidation Advisory Committee was formed by the initiative of Minister Harish Singh Rawat in 2003.(www.ghughuti.org)

Prepared by the Parvatiya Chakbandi Salhakar Samiti, the draft has spelt out a well laid out plan for undertaking consolidation of scattered land holdings and revitalize barren agriculture in the hill districts of the state. After the Bill becomes an Act, it will be implemented in the hill districts of the state, excluding Haridwar, Udham Singh Nagar, plain areas of Dehradun, Tehri, Pauri, Nainital, and Champawat. (www.ghughuti.org)

“The basic premise of the chakbandi exercise is to amalgamate and redistribute fragmented land holdings scattered in different corners in the hilly districts. Unfortunately, we were not able to undertake in the hilly districts, though it is already implemented in the plain districts since the UP days,” said Kedar Singh Rawat, chairman of the samiti. (Sharma N. 2015)

- 2. Cost-effective water schemes:**

Implementation of such schemes for the betterment of the occupation which will further help in considering other developmental aspects of agriculture. Judicious use of water also involves growing crops which require limited supply of water. Crops like rice should be grown in the rainy season as they require less water. For this, awareness is needed. According to a survey,

among State's 11 mountain districts, a mere 18% of land remains irrigated compared to over 95% in the plain districts of Haridwar and Udham Singh Nagar.

3. Reviving the Ancient Art Forms of Hilly Regions:

A civilization's traditions, culture and customs are its founding stones. Music, dance and various other forms of art and craft exude the most creative and awe inspiring characteristics of a society. One such art form among the many from the hills is Aipan, which hails from Kumaon. So, efforts should be taken to revive these art forms and train the youth and the women in these, almost lost arts, which have a huge demand in the international market and thus these would provide the base for setting up small scale and cottage industries providing employment as well as a self-sustainable system to these people.

4. Awareness about better agricultural practices:

Latest technology and easy accessibility of raw materials should be provided which will result in an increase in soil fertility. Awareness on various farming techniques and focus on reviving agricultural practices and employing better techniques, such as the Baranaaja or Twelve Grains, (www.ghughuti.org). It is a unique version of poly cropping in the Garhwal region which is growing a number of crops, mixed randomly on the same field that optimally tapped the soil and solar energy along with working as a defense against pests.

The Beej Bachao Andolan (www.ghughuti.org) here which raised awareness on how the new practices of using high yield varieties of seeds was fickle and even with higher investments the greater yield does not last long, thus, became a stepping stone to the Twelve Grains Movement. Practices such as these should be better campaigned and promoted throughout the state.

5. Better railway network:

A wide network of railways to all districts should be developed as a strategic defense initiative also generating employment.

6. Improvement in government education:

Better education should be provided at the government schools till the secondary level and the foremost focus should be on reducing the drop-out rates at the school. Also, efforts should be taken to provide the students with vocational courses at the school level itself. It should be ensured that each and every one is benefitted by the National Skill Development Corporation (www.nsdindia.org)

7. Disaster management initiative:

www.ijellh.com

The fear caused due to the recent natural calamities is also one of the major causes of migration. Also, it has been predicted that the villages that had been cut away due to the 2013 floods and have yet not been fully reconnected. Once done would ensure more migration. To prevent this better Disaster Management Initiatives should be undertaken with the help of the people from the villages itself by organizing them into groups and raising awareness and turning such initiatives into a mass movement. Participation of the people itself in these ventures would reduce the migration to a considerable level. The relocation and rehabilitation of the affected people would also garner predominant faith in the government plans and reducing migration.

8. Co-existence:

65% of the total forest area is covered in thick vegetation due to which the government finds it difficult to carry out development activities. Also, due to the Forest Conservation Act, almost 200 projects are awaiting clearances. The Hydro power potential is also unable to be exploited. So, methods should be found so that the environment and the ecology should peacefully co-exist with the humans. (Joshi S. 2012)

9. Green bonus:

Due to the huge amount of forests present in the state, a lot of green bonus is generated and this has become a method of revenue for the state. But, the fact should be realized that having plantations or conserving forests and afforestation is not the correct solution to counter the emissions produced by the burning of fossil fuels and, thus, this should not be made a hindrance in carrying out development activities.

10. Other Solutions:

Following the example of Himachal Pradesh, the ghost villages should be opened to the tourists and the public and promoted as adventure tourism.

Medical tourism and Ayurveda should be promoted.

By generating better living conditions and employment opportunities, the focus should be on reverse migration.

Conclusion

The root cause is the agricultural problem, which further includes farming done at an individual level and not at a cooperative one. The methodological approach and implementation, too, is obsolete with the respect to the present technology, with the water paucity and the finance
www.ijellh.com

constraints making the current situation even worse. Dearth of basic facilities, vis-à-vis health care, education, employment etc. compels the village and hilly population to migrate to the cities. People of the rural regions will reach new heights when the aforementioned problems will be eradicated from their roots. Although, the relevant and necessary measures have been taken by the government such as the National Skill Development Corporation (NSDC) which aims to provide skill-based training via industry specific courses, thereby taking a step towards the improvement of the condition. The need of the hour is ensure that each and every individual is benefitted from such government plans which will further help in the development and prosperity of both the individual as well as the nation. Awareness will be an innovative step to tackle with the problem such as growing crops which require less water which will be helpful in water management. Techniques like the Baranaaja or Twelve Grains and movements like the Beej Bachao Andolan added to the future enhancement of the occupation and there is a need to come up with more such plans.

References:

1. Venkatesh, S. (2016). How Abandoned Villages in Uttarakhand are posing a Major Threat to the State. Down to Earth. Retrieved from <http://www.youthkiawaaz.com/2016/01/uttarakhand-migration-threat-to-economy/>
2. Joshi S. (2012) A State grappling with the problem of migration. Retrieved from <http://www.thehindu.com/news/national/other-states/a-state-grappling-with-the-problem-of-migration/article2843029.ece>
3. Sharma N. (2015). Chakbandi to become Reality in State Soon. Retrieved from <http://www.tribuneindia.com/news/uttarakhand/chakbandi-to-become-reality-in-state-soon/142559.html>
4. Moudgil A. (2015). U'khand: Experts call for measures to curb extensive migration from hills
5. Kasniyal, B.D. (2015). Uttarakhand Villagers Abandon Hills for Btter Facilities in Cities. Retrieved from <http://www.tribuneindia.com/news/nation/uttarakhand-villagers-abandon-hills-for-better-facilities-in-cities/147816.html>
6. Bhatt J. (2013). Migration from Uttarakhand Villages to Increase when Road Connectivity Restored. Retrieved from <http://hillpost.in/2013/07/migration-from-uttarakhand-villages-to-increase-when-road-connectivity-restored/93783/>
7. Venkatesh, S. (2016). How Abandoned Villages in Uttarakhand are Posing a Major Threat to the State. Down to Earth. Retrieved from <http://www.youthkiawaaz.com/2016/01/uttarakhand-migration-threat-to-economy/>
8. Land Consolidation – Chakbandi Campaign in Uttarakhand, Retrieved from <http://www.ghughuti.org/land-consiladation-chakbandi-campaign-in-uttarakhand/>
9. Sharma N. (2015). Chakbandi to become Reality in State Soon. Retrieved from <http://www.tribuneindia.com/news/uttarakhand/chakbandi-to-become-reality-in-state-soon/142559.html>
10. Land Consolidation – Chakbandi Campaign in Uttarakhand, Retrieved from <http://www.ghughuti.org/land-consiladation-chakbandi-campaign-in-uttarakhand/>

11. Gusain, R. Unique Movement to Conserve Traditional Seeds. Retrieved from
<http://www.uttaranchal.org.uk/bba.php>
12. www.nsdindia.org

International Conference on Computational Modeling and Security (CMS 2016)

Security Algorithms for Cloud Computing

Akashdeep Bhardwaj^{a*}, GVB Subrahmanyam^b, Vinay Avasthi^c, Hanumat Sastry^d

^aUniversity of Petroleum and Energy Studies, Bidholi Dehradun 248001, India

^bTech Mahindra, Infocity, Hyderabad 500081 India

^cUniversity of Petroleum and Energy Studies, Bidholi Dehradun 248001, India

^dUniversity of Petroleum and Energy Studies, Bidholi Dehradun 248001, India

Abstract

With growing awareness and concerns regards to Cloud Computing and Information Security, there is growing awareness and usage of Security Algorithms into data systems and processes. This paper presents a brief overview and comparison of Cryptographic algorithms, with an emphasis on Symmetric algorithms which should be used for Cloud based applications and services that require data and link encryption. In this paper we review Symmetric and Asymmetric algorithms with emphasis on Symmetric Algorithms for security consideration on which one should be used for Cloud based applications and services that require data and link encryption.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the Organizing Committee of CMS 2016

Keywords: Cryptography, Security Algorithm, Symmetric, Asymmetric, RSA, RC6, AES, 3DES, MD5

1. Introduction

Imagine two people who share critical secret information have to split up. This requires them to share and communicate their data and information from a distance, even as there lays a threat of an eavesdropper having the ability to stop, interfere or intercept their communications and seeks that same information. They decide to lock their information in a box using a lock that only the other knows the combination to and has the key to open it. The box is locked and sent over to the other user who uses the combination key to unlock the box and read its contents. In simple terms, Cryptography [1] can be seen as a method of storing and disguising confidential data in a cryptic form so that only those for whom it is intended can read it and are able to communicate information in the presence

* Corresponding author. Tel.: +91-987-327-6660.

E-mail address: Bhrdwh@yahoo.com

of an adversary and the security algorithms mitigate security issues by use of cryptography, authentication and distributing keys securely. Cryptography is thus the science of making data and messages secure by converting the end user data to be sent into cryptic non-readable form and encrypting or scrambling the plaintext by taking user data or that referred to as clear text and converting it into cipher text [2] and then performing decryption which is reverting back to the original plain text. With this ability, Cryptography is used for providing the following security:

- Data Integrity: information has value only if it is correct, this refers to maintaining and assuring the accuracy and consistency of data, its implementation for computer systems that store use data, processes, or retrieve that data.
- Authentication for determining whether someone or something is, in fact, who or what it is declared to be.
- Non Repudiation: is the assurance that a party, contract or someone cannot deny the authenticity of their signature and sending a message that they originated.
- Confidentiality: relates to loss of privacy, unauthorized access to information and identity theft.



Fig 1: Encryption and Decryption process

In pure science terms [3], Cryptography is the science of using mathematics for making plain text information (P) into an unreadable cipher text (C) format called encryption and reconverting that cipher text back to plain text called as decryption with the set of Cryptographic Algorithms (E) using encryption keys (k1 and k2) and the decryption algorithm (D) that reverses and produces the original plain text back from the cipher text. This can be interpreted as Cipher text $C = E \{P, \text{Key}\}$ and Plain text $P = D \{C, \text{Key}\}$

Defining some terms used in Cryptography:

- Plaintext is the original intelligible source information or data that is input to algorithms
- Cipher text is the scrambled message output as random stream of unintelligible data
- Encryption Algorithm substitutes and performs permutations on plain text to cipher text
- Decryption Algorithm is encryption run in reverse by taking the secret key and transforming the cipher text to produce the original plain text
- Keys are used as input for encryption or decryption and determines the transformation
- Sender and Recipients are persons who are communication and sharing the plaintext

With respect to Cloud computing, the security concerns [4] are end user data security, network traffic, file systems, and host machine security which cryptography can resolve to some extent and thus helps organizations in their reluctant acceptance of Cloud Computing. There are various security issues that arise in the Cloud:

- Ensuring Secure Data Transfer: In a Cloud environment, the physical location and reach are not under end user control of where the resources are hosted.
- Ensuring Secure Interface: integrity of information during transfer, storage and retrieval needs to be ensured over the unsecure internet.
- Have Separation of data: privacy issues arise when personal data is accessed by Cloud providers or boundaries between personal and corporate data do not have clearly defined policies.
- Secure Stored Data: question mark on controlling the encryption and decryption by either the end user or the Cloud Service provider.
- User Access Control: for web based transactions (PCI DSS), web data logs need to be provided to compliance auditors and security managers.

Security Algorithms are classified broadly as:

- Private Key / Symmetric Algorithms: Use single secret key are used for encrypting large amount of data and are have fast processing speed. These algorithms use a single secret key that is known to the sender and receiver. RC6, 3DES, Blowfish, 3DES are some prime examples of this algorithms.

- **Public Key / Asymmetric Algorithms:** Use a key pair for cryptographic process, with public key for encryption and private for decryption. These algorithms have a high computational cost and thus slow speed if compared to the single key symmetric algorithms. RSA and Diffie Hellman are some types of public key algorithms.
- **Signature Algorithms:** Used to sign and authenticate use data are single key based. Examples include: RSA, DH
- **Hash Algorithms:** Compress data for signing to standard fixed size. Examples include: MD5, SHA
- Some other ways of classifying Algorithms based on their processing features as below

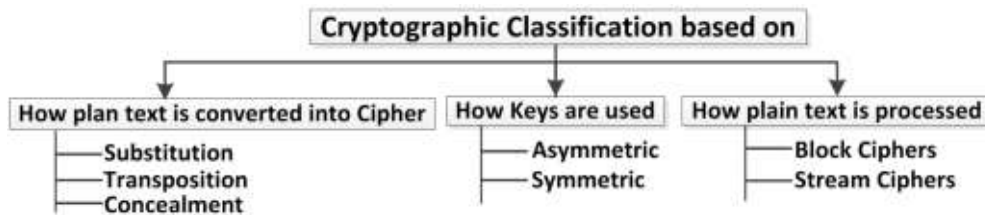


Fig 2: Classification of Algorithms

With several Cloud services, Servers and hosted applications under IT management, most Cloud providers have no defined process to ensure security of data from threats and attacks [5]. Cyberattacks target the end user data for which the Cloud Service providers seek to try and secure by using Cryptographic algorithms whose primary goal is to make it as difficult as possible to ensure decrypting the generated cipher text from the plain text. When the key length is long, that makes it harder to decrypt the cipher texts, which in turn make the algorithms efficient and effective.

2. Asymmetric Algorithms

Asymmetric Algorithms [6] a pair of related key, one key for encryption called the Public key and a different but inter related key for Decryption called the Private keys when performing transformation of plain text into cipher text. The main asymmetric algorithms are ECC, Diffie-Hellman and RSA.

2.1 RSA:

RSA Algorithm named after its inventors (Rivest, Shamir, and Adelman) is best suited for data traveling to/from Web and Cloud based environments. In working with Cloud Computing, the end user data is first encrypted and then stored on the Cloud. When the data is required, the end user simply needs to place a request to the Cloud Service provider for accessing the data. For this the Cloud service provider first authenticates the user to be the authentic owner and then delivers the data to the requester using RSA Asymmetric Algorithm. This algorithm has support from .NET Security Framework as well.

Here two keys involved – first the Public Key [7] which known to all and the other Private Key which is known only to the end user. Data conversion from plain text to cipher text is done using Public Key by the Cloud service provider and the cipher text to plain text decryption is done by the end user using Private Key as the Cloud service consumer. Once the user data is encrypted with the Public Key, that cipher data can only be decrypted with the corresponding Private Key only. In this Algorithm, prime numbers are used to generate the public and private keys based on mathematical formulas and by multiplying the numbers together. This uses the block size data in which plain text or the cipher texts are integers between 0 and 1 for some n values. Here the processed plaintext is also encrypted in blocks and the binary value of each block needs to be less than the number (n). RSA being multiplicative homomorphic which essentially means that to find the product of the plain text, multiply the cipher texts so that the outcome of the result is the cipher text of the product.

2.2 Diffie-Hellman Key Exchange (D-H):

This is a method for exchanging cryptographic keys [8] by first establishing a shared secret key to use for the inter communication and not for encryption or decryption. This key exchange process ensures the two parties that have no prior knowledge of each other to jointly establish a shared secret key over unsecure internet.

Transformations of keys are interchanged and both end up with the same session key that looks like a secret key. Then each can then calculate a third session key that cannot easily be derived by an attacker who knows both exchanged values. This key encrypts the subsequent communications using a symmetric key cipher but is vulnerable to the Man-in-the Middle (MITM) attack. This key exchange is not used for exchanging real large data unlike RSA.

3. Symmetric Algorithms

Symmetric algorithms involve a single shared secret key [9] to encrypt as well as decrypt data and are capable of processing large amount of data and from computing standpoint are not very power intensive, so has lower overhead on the systems and have high speed for performing encryption and decryption. Symmetric algorithms encrypt plaintexts as Stream ciphers bit by bit at a time [10] or as Block ciphers on fixed number of 64-bit units.

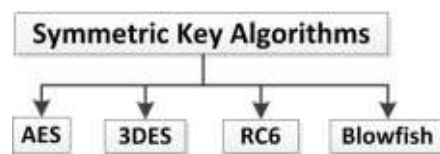


Fig 3: Symmetric Algorithms

There are however few problems with Symmetric Algorithms:

- **Exchanging Shared Secret Key [11] over unsecure internet.**
Symmetric-key algorithms share secret keys required by the sender and receiver during encryption or decryption process. In case a third person gains access to the secure secret key, cipher text messages can easily be decrypted. The fact of having one single secret key algorithm is the most critical issue faced by Cloud service providers when dealing with end users who communicate over unsecure internet. The only option is to have that secret key be changed often or kept as secure as possible during the distribution phase.
- **Problem confirming if the content is altered or actually sent by the claimed sender.**
If a hacker has the secret key, decrypting the cipher text, modifying the information being sent with that key and send to the receiver. Since a single key is involved during the crypto process, either side of the transactions can get compromised. Such data integrity and non-repudiation issues however need to involve the use of Digital signatures or Hashing functions like MD5.
- **Tools for cracking Symmetric encryption**
By use of Brute force [12] by running hacking tools that have the ability crack the combinations and keys to determine the plaintext message and perform Cryptanalysis where the attacks are focused on the characteristics of the algorithm to deduce a specific plaintext or the secret key. Then hackers are able to figure out the plaintext for messages that would use this compromised setup.

4. Related Work Performed

With DDoS and Malware attacks on the rise, Cloud Providers are giving more focus on having end user data as secure as possible and having low priority for cloud performance due to inconsistent selection of algorithms for encryption and encoding. By selecting the right cryptographic scheme end user data security can be achieved without losing out on cloud performance. Since Algorithm analysis is an essential in gathering the knowledge against any accidental or unintentional use algorithm that may prove to be inefficient or significantly impact application system performance due to encryption or decryption. For those cloud based web applications or portals needing real time or time sensitive data, an algorithm that might be taking a long time to long to run would prove a hindrance for the real time application as it may render the results to be useless. Such in efficient algorithm might end up needing lots of computing power or storage to execute over the cloud, making the algorithm useless in that

environment.

4.1. Comparison Parameters

Authors compared Symmetric encryption algorithms and encoding algorithms using size and time to decide on selection of the right algorithms based on the parameters as

- File Size: indicates file of different size to be taken
- Encryption Computation Time: time an algorithm takes to produces a cipher text from a plain text
- Encoding Computation Time: time taken by encoding algorithm to produce a hash code

Performance metrics were collected based on the following:

- Encryption & Decryption Time: This is calculated as the time required for encryption which involves converting the plain text payload file into cipher text. The authors used the encryption time to find the through put which indicated the computation cost i.e. the encryption speed. The decryption time is calculated for the amount of time required for converting the cipher text back into the plain text.
- CPU Processing Time: This is determined as the time CPU is committed for the process and reflects the CPU load during the encryption process. The CPU Clock Cycle and Battery power are the energy consumed during encryption and decryption process.
- Size of payload to be tested: This is actual size of the test file that is being used for the experimental work.

The authors then used the below infrastructure for our data gathering research work:

- Connectivity: 1Mbps WAN circuit link connected to a public Cloud server provider
- Cloud Simulation: Hosted Web application server on the IaaS systems for cloud environment
- Working environment:
 - Programming language environment - Java
 - Setup one 64 bit Windows Server 2008 Operating system
 - Running on VMware based a Virtual machine
 - Over hardware as Intel Core i5-3230M CPU @ 2.66GHz, 8GB memory.

4.2 Performance Evaluation

The below mentioned actions were performed as input using different algorithms to encrypt the data (text file) to determine the time required for reading the file, encrypting it, creating the encrypted data, then sending the data to a cloud location and receiving a confirmation.



Fig 4: Encrypting text file to send to the cloud

The input variables are

- File upload: D:\SACC\Data\Encrypt.txt (input)
- Choosing algorithm:
- Encoding Hash
- Key size:
- Mode: Encrypt or Decrypt

5. Performance Results

The data from experimental work on Symmetric algorithms is depicted below by using varied file sizes as input and recording the computation cost for those algorithms. Encoding algorithms checks for data integrity for end user data on the cloud and computation cost data obtained for different algorithms by varying the size of payload.

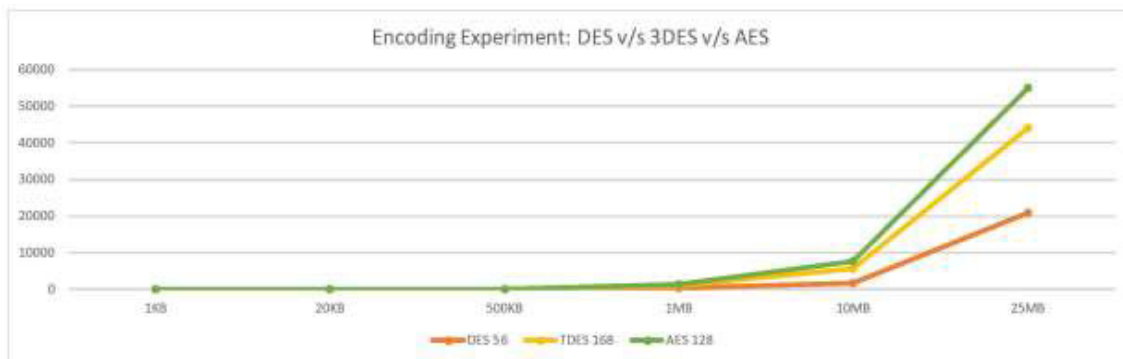


Fig 5: Computational Cost for Encryption

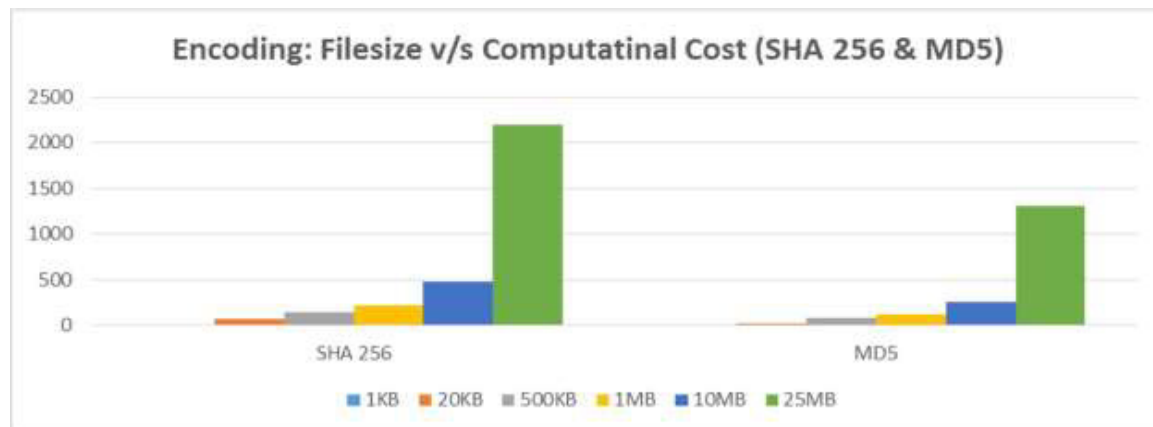


Fig 6: Computational Cost for Decryption

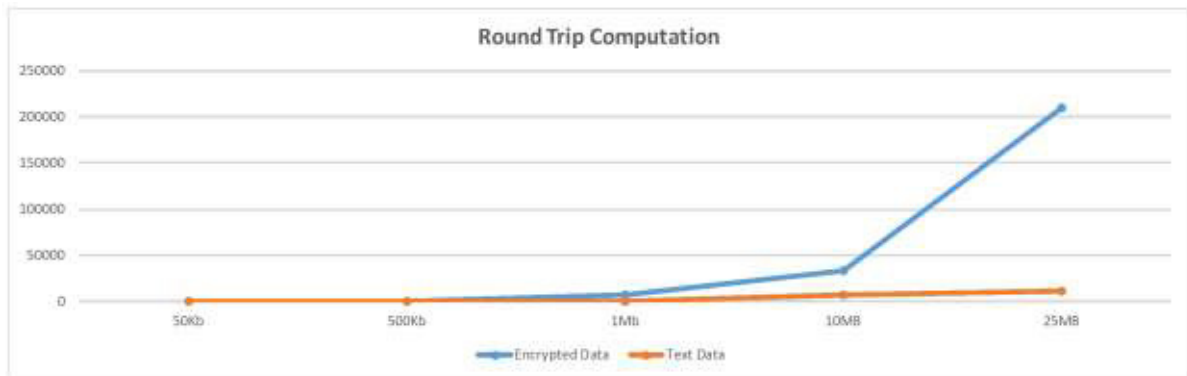


Fig 7: Execution Time (Text file to encrypted file and back)

Further observations from the work performed:

- Data Security for Cloud based applications can be increased by using RSA and AES Encryption algorithms
- When using keys as 1024 bit RSA and 128 bit AES, determining the private key is not possible even if the attacker has the public keys generated
- After the end user logs in to the Cloud web portal, accesses the applications but does not log out and in fact just leaves the session idle, then in this case if an attacker breaks in to the user system attempting to download and access the data from the user system, then the attacker would be required to enter the private key.
- In case the attacker in his attempt to break in to the user system is successful, even able to somehow guess the private key and then go on to download the encrypted data.
- The attacker might be successful in getting the encrypted data but still accessing the original data might still not be possible.

6. Conclusions

With Cloud computing emerging as a new in thing in technology industry, public and private enterprise and corporate organizations are either using the Cloud services or in process of moving there but face security, privacy and data theft issues. This makes Cloud security a must to break the acceptance hindrance of the cloud environment. Use of security algorithms and ensuring these are implemented for cloud and needs to be properly utilized in order to ensure end user security. The authors analyzed Symmetric algorithms for different encryption and encoding techniques, found AES to be a good candidate for key encryption and MD5 being faster when encoding.

References

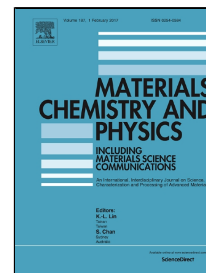
1. Leena Khanna, Anant Jaiswal, "Cloud Computing: Security Issues and Description of Encryption Based Algorithms to Overcome Them", IJARCSSE 2013
2. G Devi, Pramod Kumar "Cloud Computing: A CRM Service Based on a Separate Encryption and Decryption using Blowfish algorithm" IJCTT 2012
3. Simarjeet Kaur "Cryptography and Encryption in Cloud Computing", VSRD International Journal of CS and IT, 2012
4. Nelson Gonzalez, Charles Miers, Fernando Redigolo, Marcos Simplicio, Tereza Carvalho, Mats Naslund, Makan Pourzandi "A quantitative analysis of current security concerns and solutions for cloud computing", Springer 2012.
5. Ronald Krutz, Russell Vines, "Cloud Security: A Comprehensive Guide to Secure Cloud Computing" Wiley Publishing 2010
6. Behrouz Forouzan, "Cryptography and Network Security", McGraw-Hill Special Indian Edition 2007
7. Wayne Jansen, Timothy Grance "Guidelines on Security and Privacy in Public Cloud Computing", National Institute of Standards and Technology 2011
8. Akhil Behl "Emerging Security Challenges in Cloud Computing", IEEE 2011
9. Maha Tebba, Saïd Haji Abdellatif Ghazi, "Homomorphic Encryption Applied to the Cloud Computing Security", World Congress on Engineering 2012
10. Cloud Security Alliance (CSA), "Security Guidance for critical Areas of Focus in cloud computing V3.0" CSA 2015
11. Ayan Mahalanobis, "Diffie-Hellman Key Exchange Protocol, Its Generalization and Nilpotent Groups." 2005
12. Neha Jain, Gurpreet Kaur, 'Implementing DES Algorithm in Cloud for Data Security', VSRD International Journal of CS and IT, 2012

13. Mandeep Kaur, Manish Mahajan, "Implementing Various Encryption Algorithms to Enhance The Data Security Of Cloud In Cloud Computing" VSRD International Journal of Computer Science & Information Technology 2012
14. Jeeva, Dr. Palanisamy, Kanagaram, "Comparative Analysis of Performance Efficiency and Security Measures of some Encryption Algorithms", IJERA ISSN: 2248-9622 Vol. 2, Issue 3, 2012
15. Dr. Sarbari Gupta, "Securely management cryptographic keys used within a cloud environment", NIST Cryptographic Key management workshop, 2012
16. Dr. R. Chandramouli "Key Management Issues in the Cloud Infrastructure", Workshop on Cloud Computing, 2013
17. Sandro Rafaeli, "Survey of key management for secure communication", ACM Computing Surveys, 2013
18. S. Anahita Mortazavi, Alireza Nemaney Pour, Toshihiko Kato, "An Efficient Distributed Group Key Management using Hierarchical Approach with Diffie-Hellman and Symmetric Algorithm: DHSA", CNDS Feb 2011
19. ENISA, "Algorithms, Key Sizes and Parameters Report, 2013", recommendations version 1.0 – October 2013
20. Y. Fan, L. Xiao-ping, D. Qing-kuan and L. Yan-ming, "A Dynamic Layering Scheme of Multicast Key Management," IEEE 5th International Conference on Information Assurance and Security, Xian 2009
21. Rajesh Ingle, G. Sivakumar, "EGSI: TGKA based Security Architecture for Group Communication in Grid", 10th IEEE/ACM International Conference on Cluster, Cloud and Grid Computing, pp. 34-42, 17-20 May, 2010.
22. NIST, "Cloud Computing Synopsis and Recommendations", Special publication 800-146, May 2012

Accepted Manuscript

Superior Mechanical Properties of Poly Vinyl Alcohol-Assisted ZnO Nanoparticle Reinforced Epoxy Composites

Sudipta Halder, Tankeshwar Prasad, Nazrul Islam Khan, M.S. Goyat, Sri Ram Chauhan



PII: S0254-0584(16)30968-3
DOI: 10.1016/j.matchemphys.2016.12.055
Reference: MAC 19386
To appear in: *Materials Chemistry and Physics*
Received Date: 03 August 2016
Revised Date: 11 December 2016
Accepted Date: 27 December 2016

Please cite this article as: Sudipta Halder, Tankeshwar Prasad, Nazrul Islam Khan, M.S. Goyat, Sri Ram Chauhan, Superior Mechanical Properties of Poly Vinyl Alcohol-Assisted ZnO Nanoparticle Reinforced Epoxy Composites, *Materials Chemistry and Physics* (2016), doi: 10.1016/j.matchemphys.2016.12.055

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Highlights

- PVA-assisted ZnO modification was done by sol-gel method to control their particle size
- Average particle size of ~62 nm with uniform nanorods like structures was achieved
- Enhancement in T_g value of epoxy system suggests strong interfacial interactions with fillers
- Significant improvement in mechanical properties was obtained at 2 wt.% filler content

Superior Mechanical Properties of Poly Vinyl Alcohol-Assisted ZnO Nanoparticle Reinforced Epoxy Composites

Sudipta Halder^{a*}, Tankeshwar Prasad^a, Nazrul Islam Khan^a, M.S. Goyat^b, Sri Ram Chauhan^a

^a Department of Mechanical Engineering, National Institute of Technology Silchar, Silchar-788010, Assam, India

^b Department of Physics, College of Engineering Studies, University of Petroleum & Energy Studies, Dehradun, Uttarakhand, 248007, India

*Author to whom correspondence should be addressed. E-mail: sudiptomec@gmail.com, Fax: +91-3842-224797

Abstract

This work demonstrates the effect of poly vinyl alcohol modified ZnO nanoparticles on thermal and mechanical properties of epoxy nanocomposites. Modified ZnO nanoparticles with controlled particle size were successfully synthesized by sol-gel method. The surface modification of nanoparticles was confirmed by FT-IR, EDX and TGA analysis. Epoxy nanocomposites were prepared using different wt.% (0-3 wt.%) of unmodified and modified ZnO nanoparticles. A significant improvement in T_g was achieved for synthesized epoxy nanocomposites. The dispersion and interfacial interactions between unmodified (ZN) and PVA modified ZnO (PZN) with the epoxy matrix were investigated by morphological study of fractured surface of epoxy nanocomposites. The improvement in mechanical performance of PVA-assisted ZnO nanoparticle reinforced epoxy composites was found superior at 2 wt.% of particle concentration. The maximum enhancement in mechanical properties such as tensile strength, tensile modulus, compressive strength, flexural strength and flexural modulus was ~24, 47, 48, 44 and 77% respectively with respect to that of neat epoxy system. We have exposed the potential of the PVA-assisted ZnO nanoparticle as good choice of reinforcement for composite industries.

Key Words: ZnO nanoparticles; Fracture toughness; Thermal stability; Mechanical testing; Sol-gel method

1. Introduction

Over the past few years, the introduction of nanoparticles (NPs) such as carbon, metallic or inorganic particles into polymer matrices led to significant improvement in thermal, mechanical and electrical properties of the

nanocomposite [1-5]. However, another option is the incorporation of low-cost micron sized particles into polymers which can stiffen and increase the strength of the material under certain loading conditions. But, the use of micron sized particles in these days is not so common due to generation of some other undesirable properties [6]. Therefore, worldwide research is going on to modify polymer matrices using different nano-fillers. In these days, low-cost and abundant ZnO nanoparticles are in great demand because of its morphological diversity, controlled particle size and crystalline structure. These characteristics of ZnO nanoparticles have a significant effect on the property improvement not only for semiconductors, solar cells, gas sensors, paints, cosmetics, catalysts and varistors [7-9], but also for other applications [10]. This enriches the motivation to scientists with regard to the special characteristics and morphology of ZnO nanoparticles to scrutinize not only its properties but also approaches to develop technically less sensitive and cost effective processing routes [11]. Over the past few years, the nanosized ZnO as ceramic filler has been widely used in the field of high strength polymer composites [12], electronic glass (ceramics), high-density polyethylene include safety helmets and stadium seating, as it can improve mechanical performance and thermal resistance properties of the polymers [13-15]. But, all types of pristine nanoparticles are very prone to aggregate in viscous polymers due to their high surface area and energy, which results in poor thermal and mechanical properties of the nanocomposites. Thus, to yield a better compatibility between the nano-fillers and the polymer matrix, the use of particle surface modification by various techniques is recommended to enhance the desired properties of nanocomposites [16-19]. It is already well-established that the surface modification of nano-fillers has been a convenient method to reduce the surface energy of nanoparticles while improving its dispersion to achieve required performance characteristics of nanocomposites [20]. The modified surface of nanoparticles is more reactive for the miscibility of the nanofillers into the viscous polymer matrices [21-23]. Recently, the preparations of ZnO/epoxy composites with homogeneous dispersion were reported by Ding et al. [22]. The investigation showed that the composite containing 5 wt% of ZnO particles (about 100–200 nm) had the optimal mechanical properties. The addition of ZnO nanoparticles into the epoxy matrix resulted in a significant increment in the thermal stability and activation energy of thermal degradation. The epoxy nanocomposite exhibited an increase in storage modulus and glass transition temperature compared to the neat epoxy. Rashmi R. Devi et al. [23] reported the vinyl trichlorosilane (VTCS) modified ZnO nanoparticles based Wood polymer nanocomposite. The mechanical properties including modulus of rupture (MOR), tensile strength, hardness, and thermal stability were studied. They found considerable amount of properties enhancement with uniformly dispersed nano-ZnO particles in the polymer

matrices. Mallakpour et al. [24] reported the modification of ZnO nanoparticles using poly vinyl alcohol (PVA) to increase the compatibility and dispersibility in the PVC matrix. They reported, some improvement in tensile strength, strain, and elongation at break about 27%, 19%, and 19% respectively at high filler content of 8 and 12 wt.%. In another work, Xie et al. [25] reported the surface modification of nano-Sb₂O₃ by polymethyl methacrylate (PMMA) and its effect on mechanical properties of Sb₂O₃/PVC nanocomposites. They showed the improvement in dispersibility of particles in the matrix as well as improvement in stiffness and strength of the matrix at 2.5 wt.% of filler content. Chatterjee et al. [26] investigated the effect of nano-calcium carbonate (CaCO₃) containing PMMA nanoparticles on thermal and mechanical properties of polypropylene (PP) composites. Their results showed the homogeneous dispersion of CaCO₃ nanoparticles in PP matrix due to grafting with PMMA and also significant enhancement of thermal and mechanical properties were observed at 1 wt.% filler content in the matrix.

In the above context, **study on the effect of surface** modified ZnO nanoparticles on thermal and mechanical properties of epoxy resin system has been reported for the first time. In this study, we focused on the homogeneous dispersion of ZnO nanoparticles and their impact on the thermal and mechanical properties of the resulting composites. However, earlier **reported studies based on ZnO** nanofillers reinforced polymers reveal only nanofiller dispersion and interfacial interactions as the two main issues governing the properties of nanocomposites [27, 28]. **Most of the literatures reported the functionalization of commercially available ZnO nanoparticles by attaching non-covalent functional groups on their surfaces,** but such kind of functionalization produces poor interfacial adhesion between ZnO particles and polymer matrix leading to insignificant enhancement in mechanical properties. **To achieve a desired set of properties in polymer nanocomposites, adopting a suitable method capable of controlling the interfacial interaction is of quite important.** Therefore, in the present study, ZnO nanofillers were functionalized with PVA through sol-gel method and then reinforced them into polymer matrix to enhance the interfacial interaction between the filler and matrix. The effects of unmodified and PVA-assisted ZnO nanoparticles on the thermal and mechanical properties of epoxy nanocomposites with respect to varied filler concentration were investigated. The possible toughening mechanisms were identified by microscopic investigation of the fractured surfaces of nanocomposites.

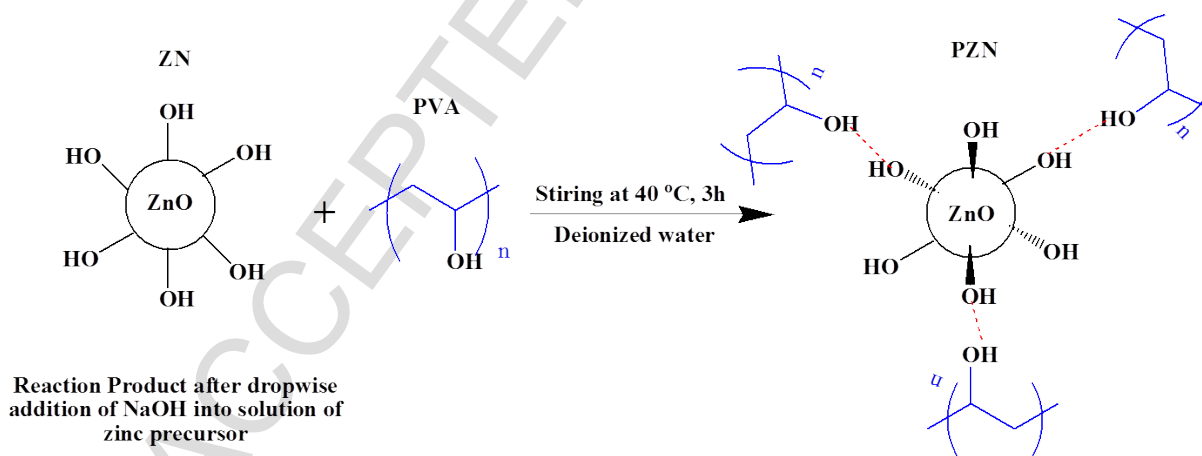
2. Experimental

2.1 Materials

Zinc Sulfate heptahydrate ($\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$) used as a precursor and sodium hydroxide (NaOH) as a reagent were supplied by Merck specialties private ltd, Mumbai, India. Poly vinyl alcohol (PVA), ER grade (MW: 22,000 g/mol) was used as a capping agent and purchased from Loba chemie private ltd., Chennai, India. Diglycidylether of bisphenol-A (DGEBA) based epoxy resin (Lapox L12) with density 1.20 g cm^{-3} and Triethylenetetramine, (TETA, K6) hardener with density 0.95 g cm^{-3} were supplied by Atul ltd., Gujarat, India.

2.2. Surface modification of ZnO nanoparticles

Surface modification of ZnO nanoparticles was carried out by using sol-gel method. Initially, $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ and NaOH were mixed in a molar ratio of 1:2 and dissolved in deionized water to obtain an aqueous solution. Further, 0.28 g of PVA (equivalent to 10 wt.% PVA in ZnSO_4) was added to 10 ml of deionized water. Then, continuous stirring was employed at 60°C for 25 min to achieve a homogenous solution. The resultant solution was mixed with zinc solution under continuous stirring followed by drop wise addition of NaOH solution. The reaction was carried out for approximately 3 h till the formation of white precipitate. The white precipitate was centrifuged, washed and dried in vacuum at 80°C for 12 h. The resultant dried material was grounded to fine powder using agate mortar. Finally, the powder was calcined for 2 h at 500°C . The ZnO nanoparticles modified with 10 wt.% PVA in ZnSO_4 are designated as PZN. Using the same method, unmodified ZnO nanoparticles were synthesized but without using PVA and designated as ZN. The possible reaction mechanism between PVA monomer and hydroxyl groups of ZnO nanoparticle is illustrated in Scheme-1.



Scheme 1: Schematic representation of reaction mechanism between PVA monomer and hydroxyl groups of ZnO nanoparticles.

2.3. Preparation of ZnO/epoxy nanocomposites

Varying amount of unmodified and PVA modified ZnO nanoparticles (1, 2 and 3 wt.%) was chosen for preparation of ZnO/epoxy nanocomposites. High shear mixing was used for blending of the nanoparticles into the epoxy network. Initially, the nanoparticles suspended epoxy resin was continuously stirred at 2000 rpm for 15 min. Later, in a stoichiometric ratio, hardener was added to the ZnO/epoxy resin mixture under continuous stirring for 5 min at 400 rpm. Thereafter, the resulting mixture was sonicated in an ultrasonic bath for 25 min to obtain homogenous mixture. After the ultrasonication, the mixture was degassed at room temperature to remove the entrapped air bubbles. Finally, the mixture was poured into rubber molds followed by pre-curing at room temperature for 24 h and post-curing at 80 °C for 2 h to obtain different sets of cured samples. For ease in writing, the neat epoxy is designated as NE, unmodified and PVA modified ZnO/epoxy nanocomposites consisting 1, 2 and 3 wt.% of ZnO are designated as (ZEC-1, ZEC-2 and ZEC-3) and (PZEC-1, PZEC-2 and PZEC-3) respectively.

2.4. Characterization

Morphology, particle size and particle size distribution of nanoparticles were investigated using FESEM, (Zeiss, model: Supra 55VP, Germany). Chemical characterization was done by FTIR spectrometer (Bruker, model: Hyperion Microscope with Vertex 80, Germany) to identify surface functional groups of modified ZnO nanoparticles. The structural properties of the synthesized nanoparticles were investigated by powder X-ray diffractometer (XRD), (Phillips, model: X'Pert PRO diffractometer, USA) with Cu K α -radiation and energy dispersive X-ray (EDX) analysis (Zeiss, model: Supra 55VP, Germany). Glass transition temperature and thermal decomposition behavior of samples were measured using a simultaneous thermal analyzer i.e. simultaneous application of Thermogravimetry (TGA) and Differential Scanning Calorimetry (DSC) (NETZSCH, model: STA 449F3, Germany). The analysis was performed on 10 mg of samples in a temperature range of 50 to 800 °C with a heating rate of 20 °C/min under nitrogen gas purging. The mechanical properties such as tensile, compression and flexural strength were determined using a computerized universal tensile machine (INSTRON, Model 8801, USA). The tensile properties of samples were investigated according to ASTM D 638 (type-V) standard. The compression tests were conducted in accordance to ASTM D695-96 with specimen size (cube) of 12.5 x 12.5 x 12.5 mm³ and the flexural properties from single edge notched bending (SENB) specimen were investigated according to ASTM D5045 with specimen size of 53 x 12 x 6 mm³. All SENB specimens were tapped by a sharp razor blade to create sharp cracks of ~ 6 mm length. At least five specimens were tested for each composition and property. The tensile

and SENB test fracture specimens were examined under FESEM to investigate the toughening mechanisms of ZnO/epoxy nanocomposites.

3. Results and discussion

3.1. PVA-assisted growth mechanism of ZnO nanoparticles

During sol-gel process, the growth of pristine ZnO nanoparticles generally follows the diffusion-limited Ostwald ripening mechanism. But, in presence of capping agent (PVA), the growth of nanoparticles follows a combination of diffusion and surface reaction mechanism [29]. This can be explained on the basis of Scheme-1. During the hydrolysis process, the increase in zinc precursor concentration significantly increases the nucleation rate of ZnO, which results in formation of large amount of ZnO nuclei due to the decomposition of intermediate parts of $\text{Zn}(\text{OH})_2$ precipitates. On the other hand, at a constant pH (10), some of $\text{Zn}(\text{OH})_2$ precipitates may have transformed into the growth units of $[\text{Zn}(\text{OH})_4]^{2-}$ due to alkaline condition. It is well-known that the polar growth of ZnO nano-crystal along (0001) direction proceeds through the adsorption of growth units of $[\text{Zn}(\text{OH})_4]^{2-}$ onto (0001) plane. These grown nuclei may aggregate together due to excess saturation [30]. There is a possibility of ionic–dipolar interaction due to electrostatic force between PVA and the growth units along the crystal plane (0001), which may lead to the formation of nucleus of rod-like twinning crystal. The growth of the individual crystallite in a twin crystal occurs along the polar c-axis due to incorporation of growth units on the growth interface plane (0001), which possess highest surface energy in nature and thus leads to the formation of rod-like ZnO nano-crystals [31].

3.2. Morphology of ZnO nanoparticles

Fig. 1(a-b) shows the FESEM images of unmodified and modified ZnO nanoparticles, which clearly reveal varying particle size, their size distribution and change in morphologies. The average particle size of unmodified and modified ZnO nanostructures was determined from FESEM image in terms of aspect ratio using Image J software. The average particle size of unmodified ZnO nanoparticles was ~78 nm (aspect ratio: 5.38) with mixed morphology of nanorod and nanoflakes like structures (Fig. 1a). But, the modified ZnO reveals the average particle size ~62 nm (aspect ratio: 4.53) with uniform nanorod like structures (Fig. 1b and Fig. 2a). Fig. 2(b) shows the histograms of particle size distribution of unmodified ZnO nanoparticle (ZN) and PVA modified ZnO nanoparticle (PZN). Generally, the use of capping agent during the synthesis of nanoparticles plays a vital role in controlling the particle size and its morphology [32].

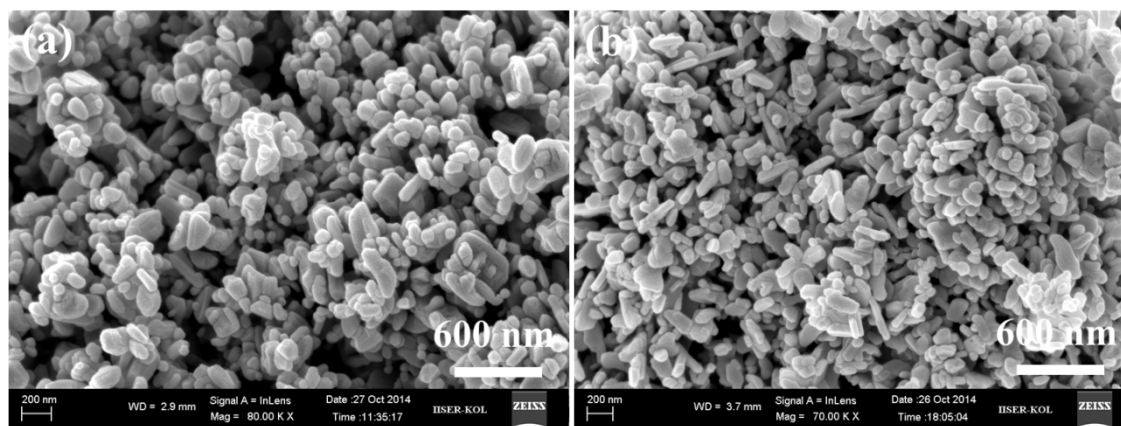


Fig. 1 Morphology of (a) unmodified ZnO nanoparticle (ZN) and (b) PVA modified ZnO nanoparticles (PZN).

However, in absence of any capping agent, there is no control over the nucleation and growth of nanocrystals and which may lead to the rapid growth of the nanoparticles without allowing any phase separation. This may lead to anisotropic growth of the nanostructures along any direction. Also, the surface of nanostructure is quite unstable which leads to the formation of irregular shape and size (Fig. 1a), in order to attain its normalized surface energy [33]. But, when the capping agent is introduced during synthesis, then the synthesized ZnO nanostructures exhibit well-developed facet planes. The variations in the morphology may affect the crystallinity of the material because of controlled growth rate of nanostructures in presence of the capping agent is possible [34-37].

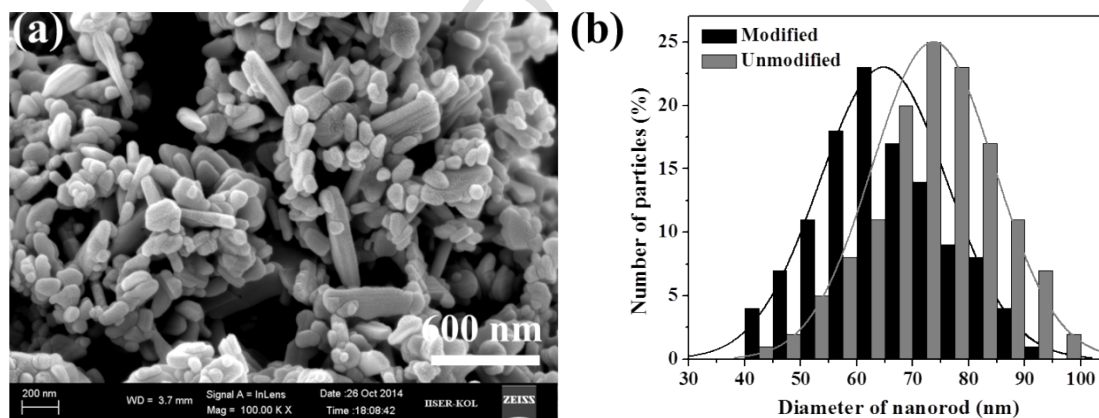


Fig. 2 (a) Morphology of modified ZnO nanoparticle at high magnification (b) histogram of particle size distribution of Unmodified ZnO nanoparticle (ZN) and PVA modified ZnO nanoparticle (PZN).

3.3. Structural property of ZnO nanoparticles

Fig. 3 shows the FT-IR spectra of unmodified ZnO nanoparticles (ZN) and modified ZnO nanoparticles (PZN). The spectra of ZN reveal a broad peak at 472 cm^{-1} , which is a characteristic absorbance peak of ZnO. Another band at

3500 cm^{-1} is attributed to the presence of hydroxyl groups ($-\text{OH}$) on the surface of ZnO nanoparticles. However, from the spectra of PZNs, peaks at 1430 cm^{-1} and 1326 cm^{-1} represent C-C stretching and C-H deformation vibrations, respectively. The C-O stretching and O-H bending vibrations of the PVA chains were detected by the absorption peaks at 1100 cm^{-1} [38]. Thus, the spectra of PZNs reveal characteristic peaks of both ZnO and hydroxyl group from PVA, which confirms the successful surface modification of ZnO.

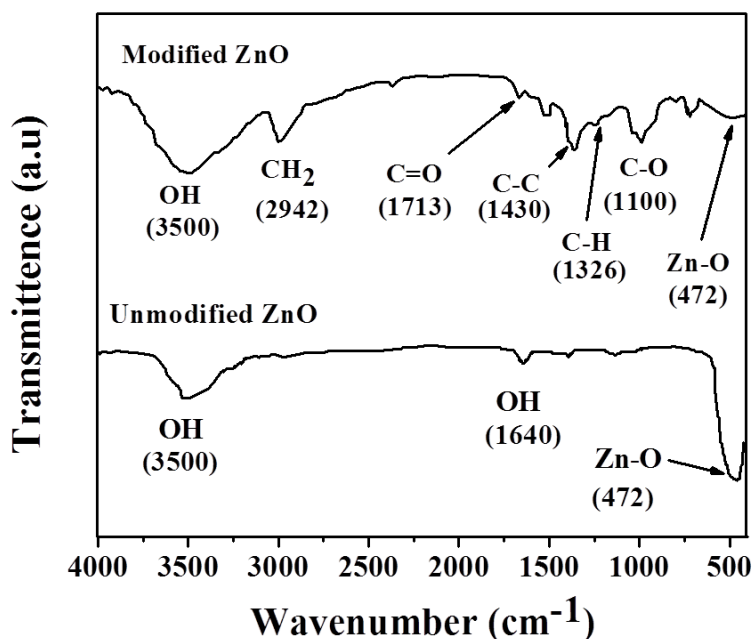


Fig. 3 FTIR spectra of Unmodified ZnO nanoparticle (ZN) and PVA Modified ZnO nanoparticle (PZN).

Fig. 4 shows the TGA curves of unmodified (ZN) and unmodified ZnO (PZN). As reflected by the TGA curves, weight loss occurs in two steps for PZN while no such behavior is observed for ZN. Initial weight loss at 50 $^{\circ}\text{C}$ is attributed to the removal of the physically absorbed volatile solvents and moisture absorbed on the surface of hydrophilic PVA chains in PZN.

Table. 1 Elemental analysis of unmodified and PVA modified ZnO nanoparticles from EDX spectra.

Unmodified ZnO			Modified ZnO		
Element	Weight %	Atomic %	Element	Weight %	Atomic %
C K	-	-	C K	9.16	18.35
O K	29.96	63.71	O K	41.22	61.96
Zn K	70.04	36.29	Zn K	49.62	19.69
Total	100		Total	100	

In second step, rapid weight loss between 200 and 400 °C may be attributed to dehydration process of hydroxyl groups from PVA chains and formation of oxides. A noticeable weight loss about 5.30% is observed in case of PZNs. This may be due to presence of PVA on the surface of nanoparticles, which again confirms the successful functionalization of the surface of ZnO nanoparticles.

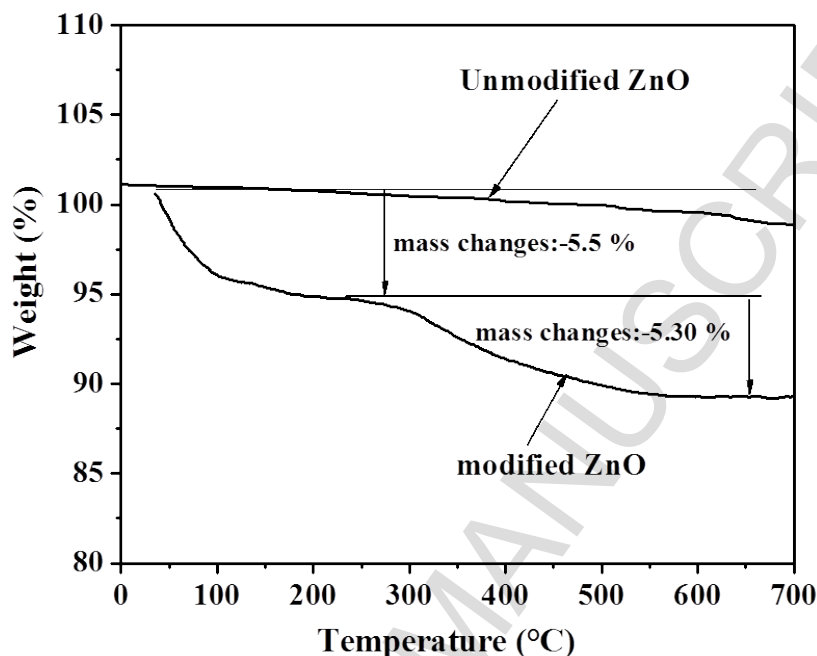


Fig. 4 TGA thermogram of unmodified ZnO nanoparticles (ZN) and PVA modified ZnO nanoparticles (PZN).

Typical XRD patterns of ZN and PZN are shown in Fig. 5 (a). Almost all the observed diffraction peaks of ZN and PZN can be indexed to wurtzite structure of ZnO (hexagonal phase, space group P6₃mc, and JCPDS no. 36-1451). No new diffraction peak is observed in case of the PZN compared to that of the ZN. The crystallite size of unmodified and modified ZnO nanoparticles have been determined using Scherer's formula [39] and found as 43.5 and 32.7 nm. The XRD patterns show the capped polymer does not affect the crystallite structure of ZnO nanoparticles, which is found in close agreement with the results reported by others [40]. The average crystallite size of PZN is about 32.7 nm, which indicates that the capping agent plays a vital role in controlling the growth of the crystals during the synthesis process. Fig. 5(b) shows the EDX spectra of ZN and PZN, which indicates high level purity of nanoparticles. The presence of C atom in case of PZN shows an evidence of PVA bonding onto the surface of ZnO nanoparticles. The elemental analysis in atomic wt.% is depicted in Table 1. Thus, again it is verified that the surface of ZnO nanoparticles is successfully modified by the capping agent PVA.

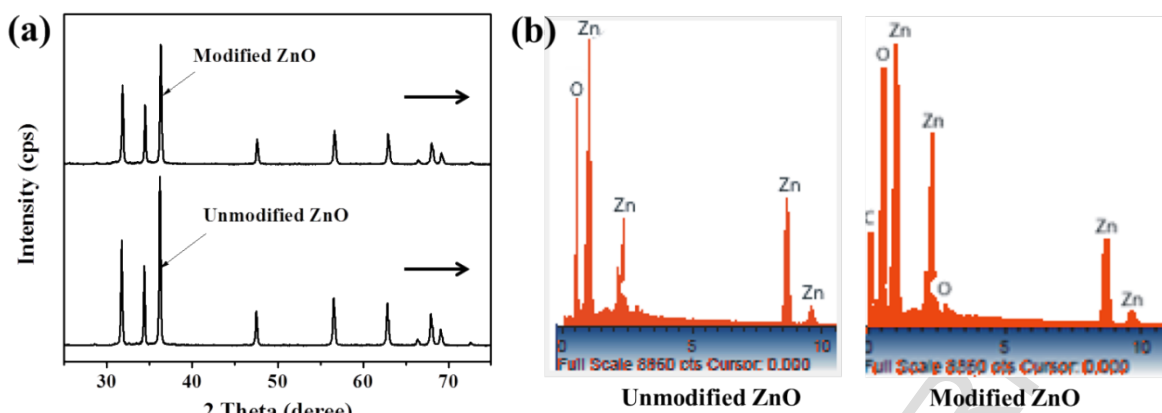


Fig. 5 (a) XRD pattern (b) EDX spectra of Unmodified ZnO nanoparticles (ZN) and PVA modified ZnO nanoparticles (PZN).

3.4. Dispersion behavior of nanocomposites

The FESEM micrographs of ZEC-1 and PZEC-1 are shown in Fig. 6. The image (Fig. 6 (a)) clearly indicates the particles in the agglomeration form, whereas the homogeneous distribution and dispersion of modified ZnO nanoparticles can be easily observed (Fig. 6 (b)) compared to that of the unmodified nanoparticles. The surface energy of the nanoparticles decreases due to the strong interaction between CH–OH groups of PVA and hydroxyl groups present on the surface of nanoparticles. The decrease in surface energy also reduces the clustering tendency of nanoparticles [41]. **The curing behavior and the possible reaction mechanism** of PVA modified ZnO nanoparticles with epoxy network are illustrated in Scheme-2. The possible interactions between the modified ZnO nanoparticles and epoxide groups have been shown through strong hydrogen bonding.

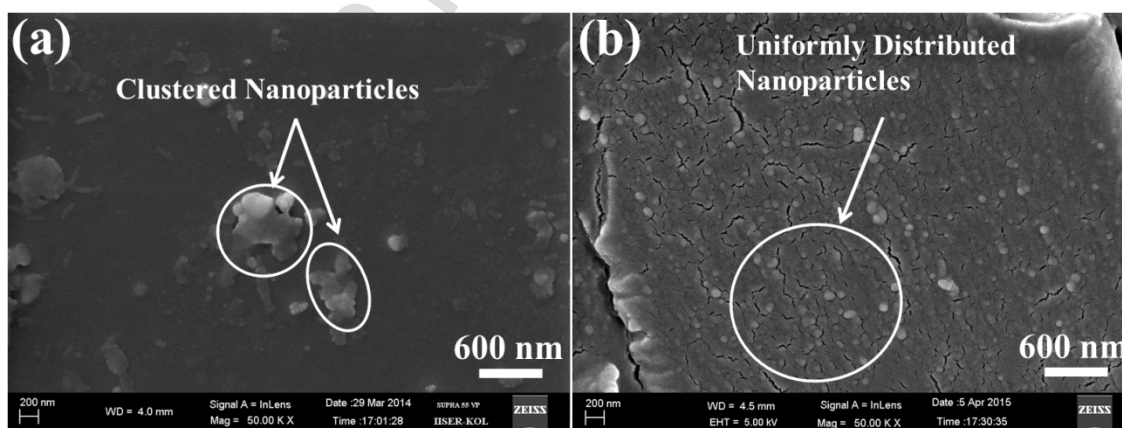
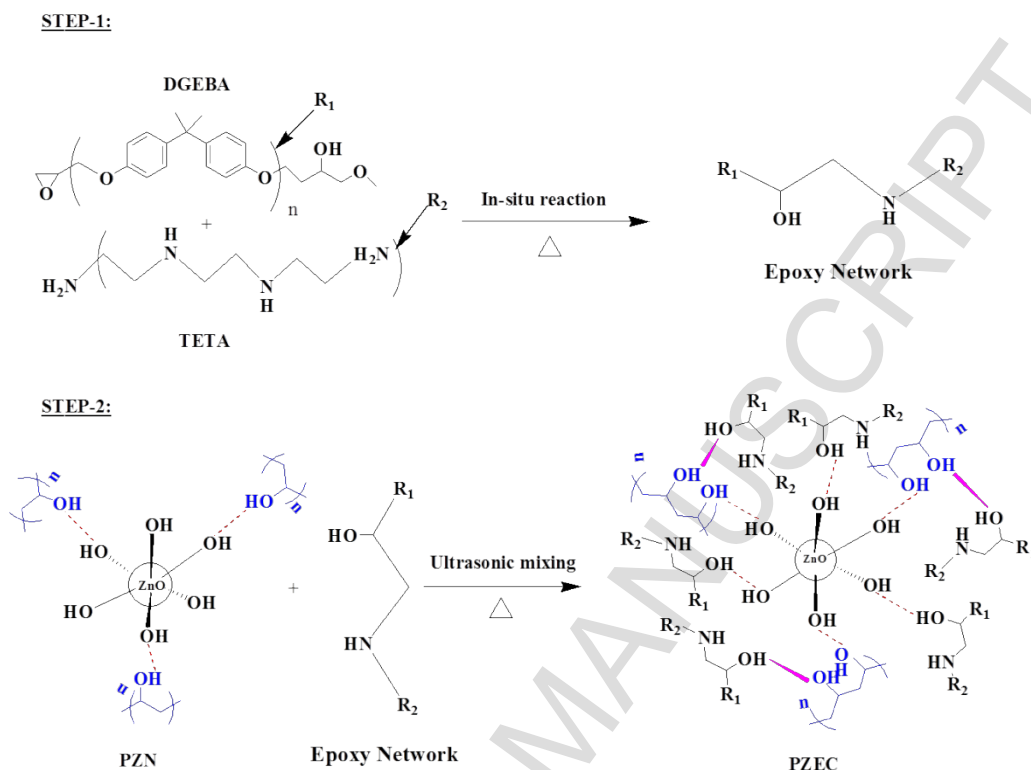


Fig. 6 FESEM micrographs of epoxy nanocomposites containing (a) unmodified ZnO nanoparticles (1 wt. %), (b) PVA modified ZnO nanoparticles (1 wt. %).

The polymer stabilizing agent, PVA also creates steric hindrance between inorganic nanoparticles to reduce their agglomeration [42].



Scheme 2: Schematic representation of curing reaction of epoxy network in presence of PZN.

3.5. Glass transition temperature and thermal stability of nanocomposites

The variation in glass transition temperature (T_g) versus the concentration of ZN and PZN in epoxy matrix is shown in Fig. 7. The increase in concentration of unmodified and modified nanoparticles upto 2 wt.% increases the T_g value of the composites. But, further increase in concentration from 2 to 3 wt.% results in decline in T_g value in both the cases. In case of modified nanoparticles reinforced epoxy matrix, the T_g value for each variation of nanoparticle content from 0 to 3 wt.%, is found higher than the unmodified particles reinforced epoxy. The maximum enhancement in T_g value of PZEC-2 is ~24% higher than the neat epoxy. This indicates the restriction offered by nanoparticles to the mobility of polymer chains of epoxy network [43-44]. Therefore, this enhancement in T_g value suggests strong interfacial interactions between the polymer matrix and surface modified ZnO nanoparticles through strong hydrogen bonding. The difference in T_g values may be due to the difference in the degree of cross-linking density of epoxy network. The higher value of T_g may be attributed to the higher degree of cross-linking density,

whereas low value of T_g with increase in particle content may be attributed to the non-homogeneous dispersion of nanoparticles in the epoxy network [45-46].

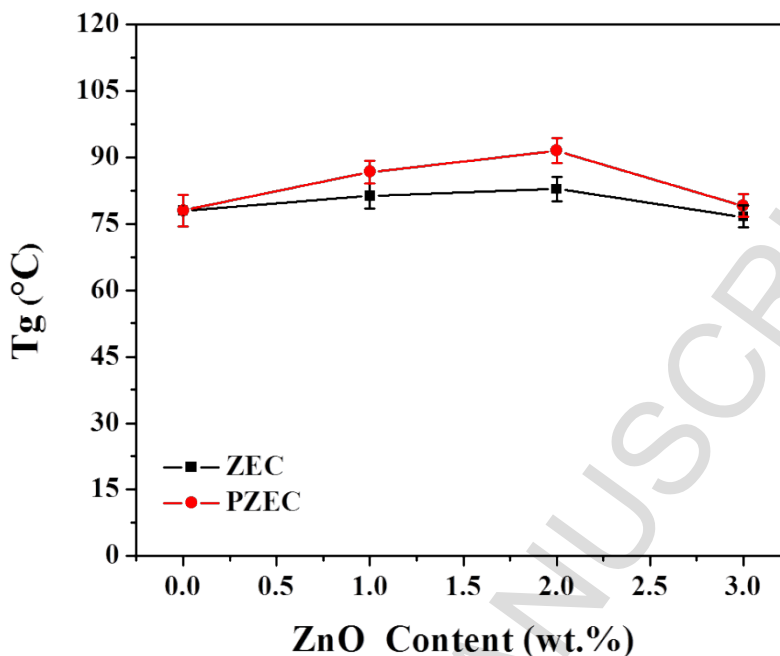


Fig. 7 Variation in glass transition temperature (T_g) with respect to concentration of ZN and PZN in epoxy matrix.

The TGA curves of unmodified and modified epoxy nanocomposites containing 1-3 wt.% of ZnO are shown in Fig. 8. The TGA plots clearly reveal single step degradation behavior for each concentration of unmodified and modified nanocomposites. The value of T_{IDT} (initial decomposition temperature, 5% wt. loss), T_{MDT} (maximum decomposition temperature, 50% wt. loss) and char yield percentage (CYP) of the composites measured from TGA plots are illustrated in Table. 2.

Table. 2 Thermal properties of NE, PVA modified and unmodified ZnO/epoxy nanocomposites.

Samples	ZnO content (wt. %)	T_{IDT} (°C)	T_{MDT} (°C)	Char yield (%) at (700°C)
NE	0	310.5	375	10.91
Modified ZnO/epoxy nanocomposites				
PZEC-1	1	319.6	376.5	15.04
PZEC-2	2	326.3	381.5	17.52
PZEC-3	3	322.9	377.8	14.37
Unmodified ZnO/epoxy nanocomposites				
ZEC-1	1	324.7	375.5	11.47
ZEC-2	2	319.8	376.5	12.68
ZEC-3	3	300	373.7	10.69

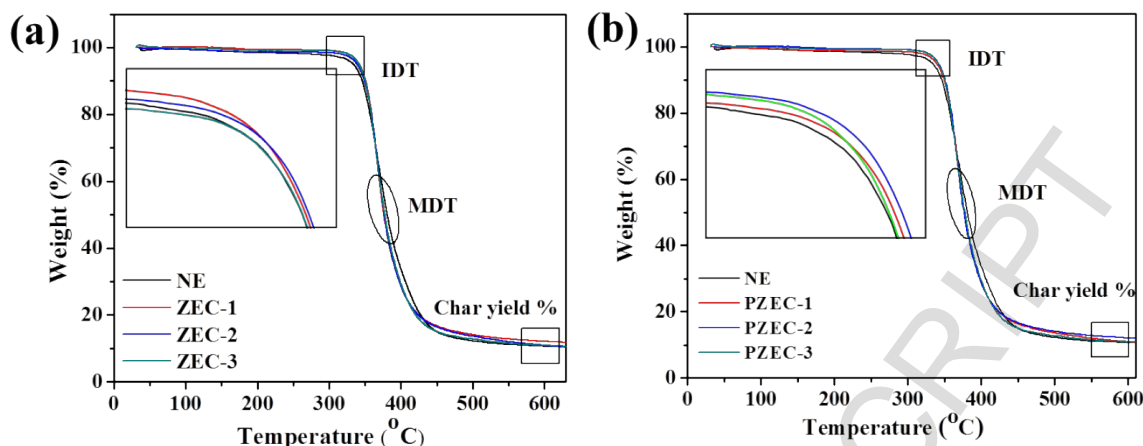


Fig. 8 TGA curve of (a) unmodified ZnO epoxy nanocomposites (b) PVA modified ZnO epoxy nanocomposites.

Fig. 8 shows the addition of both unmodified and modified nanoparticles in epoxy network results in improved thermal resistance of the composite. It was observed that the decomposition temperature of PZEC-2 slightly improves the T_{IDT} (326.3 °C) by 16 °C with respect to NE (310.5 °C). Similar trend was also observed in case of ZEC-3, but comparatively less enhancement in T_{IDT} by 10 °C with respect to NE. Moreover, the increase in T_{IDT} is an indicative of the homogeneous dispersion of ZnO in epoxy matrix which acts as obstacles to heat flow through the matrix [11, 47-49]. The char yield percentage (CYP) of unmodified and modified nanocomposites increases with the increase in particle content. It is obvious that the thermal degradation of inorganic nanoparticles is higher as compared to the polymer matrix. Therefore, due to addition of nanoparticles in the matrix, the overall thermal stability of nanocomposites increases. Moreover, these modified and unmodified ZnO nanoparticles forms residue in nitrogen atmosphere. Maximum CYP of PZEC-2 compared to other composites demonstrates its potential to hinder the decomposition at high temperature. Thus, it infers that the PVA modified ZnO nanoparticles have the potential to restrict the thermal decomposition of the epoxy network, which may be due to their nanorod like morphology with high aspect ratio compared to the unmodified nanoparticles.

3.6. Mechanical properties

3.6.1. Tensile and compressive properties

Fig. 9 shows the tensile and compressive properties of unmodified and modified ZnO nanoparticles reinforced epoxy composites. The tensile strength, tensile modulus and compressive strength was found increased with the increase in filler content up to 2 wt.% but decreased with the further increase in ZnO content. The measured values

of tensile strength, tensile modulus and compressive strength for NE are 55 MPa, 2.85 GPa and 120 MPa, respectively. The maximum enhancements in tensile strength, modulus and compressive strength of PZEC-2 compared to the neat epoxy system are ~24%, ~47% and ~48%, respectively. Furthermore, the decrease in tensile strength, tensile modulus and compressive strength at higher nanoparticle content may be attributed to the agglomeration of nanoparticles in matrix due to high surface energy and large surface area of the ZnO nanoparticles.

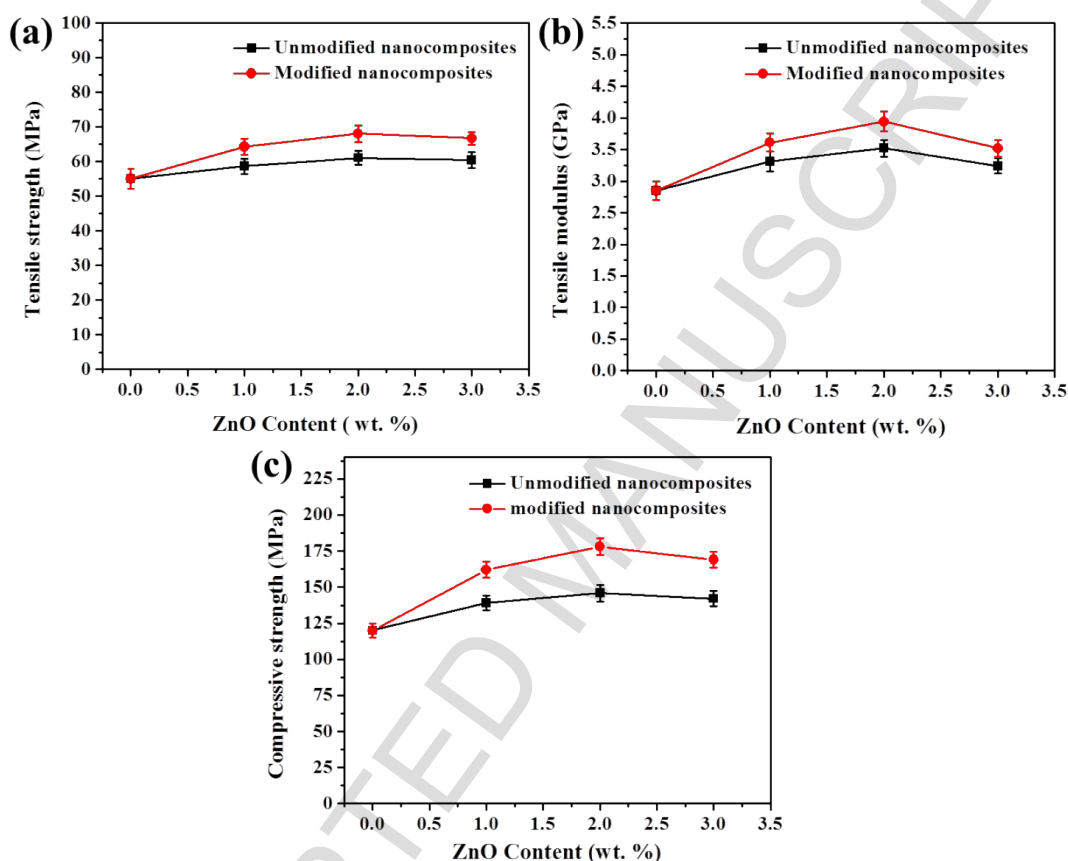


Fig. 9 Tensile and compressive properties of epoxy nanocomposites: variation in (a) tensile strength (b) tensile modulus and (c) compressive strength; with respect to the concentration of ZN and PZN.

The above hypothesis is supported by the FESEM images of the nanocomposites at higher filler contents (3 wt.%). A microscopic image of epoxy nanocomposites reinforced with modified ZnO at particles content 1, 2 and 3 wt.% is shown in Fig. 10 (a-c). From Fig. 10 (b), it is clear that maximum distribution of particles into the matrix is achieved with 2 wt.% loading while at 3 wt.% ZnO nanoparticles tend to aggregate as shown in Fig. 10 (c).

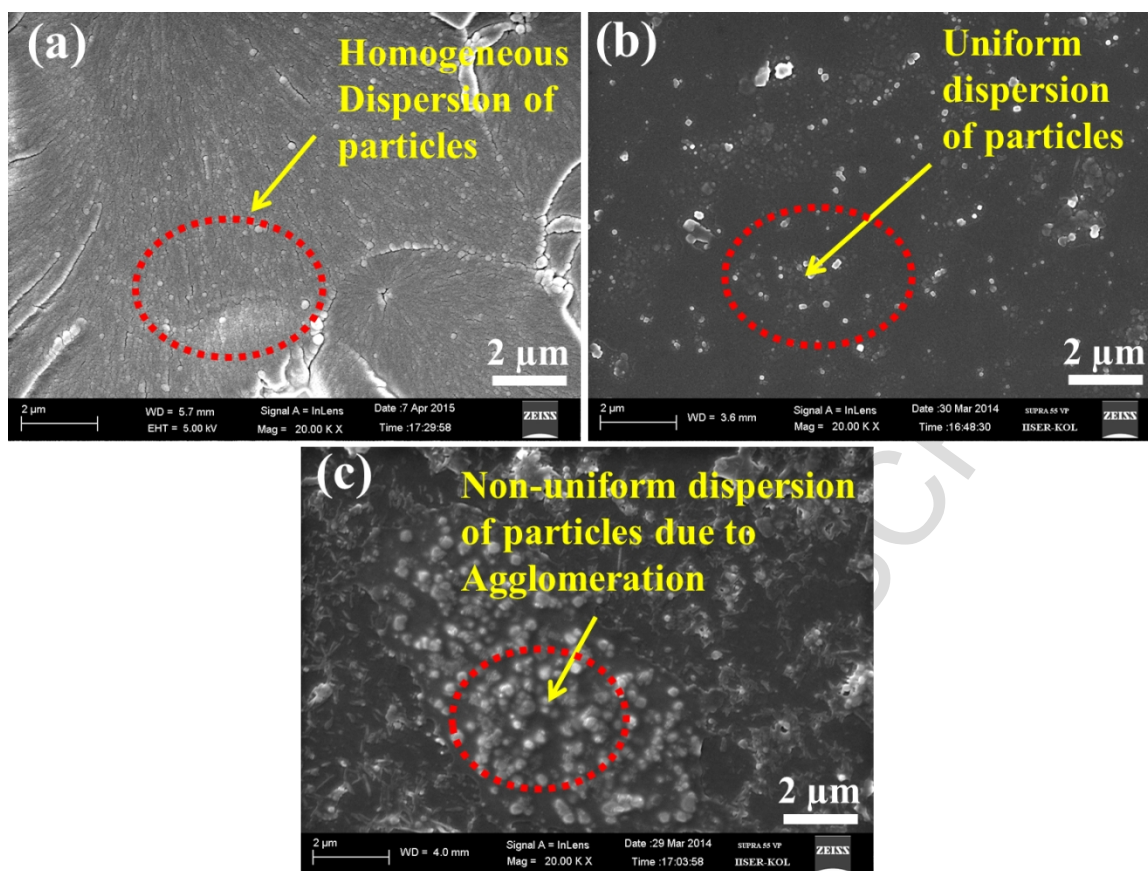


Fig. 10 FESEM images of epoxy nanocomposites containing (a) 1 wt. % PZN, (b) 2 wt. % PZN, (c) 3 wt. % PZN.

It is fact that the reinforcing capability of PZN in epoxy resin system was better than that of ZN. The mechanical properties of the epoxy nanocomposites strongly depend on the proper dispersion of fillers in the polymer matrix along with a good interaction between them. After grafting of ZnO with PVA, the surface energy gets reduced resulting in effective dispersion and strong adhesion at the interface and thereby improving tensile properties of nanocomposites. So, in comparison to unmodified ZnO nanoparticles, the modified ZnO nanoparticles show superior mechanical properties may be due to strong chemical bond formation between modified nanoparticles and the epoxy matrix. It is suggested that the use of PVA modified ZnO nanoparticles in composites can significantly improve effective dispersion and compatibility with polymer matrix.

3.6.2. Flexural property

Fig. 11 shows the flexural properties of unmodified and modified nanoparticle reinforced epoxy composites. The flexural strength and modulus of unmodified and modified ZnO/epoxy nanocomposites increase with the increase in particle content up to 2 wt.% and then decline with the further increase in particle content from 2 to 3 wt.%.

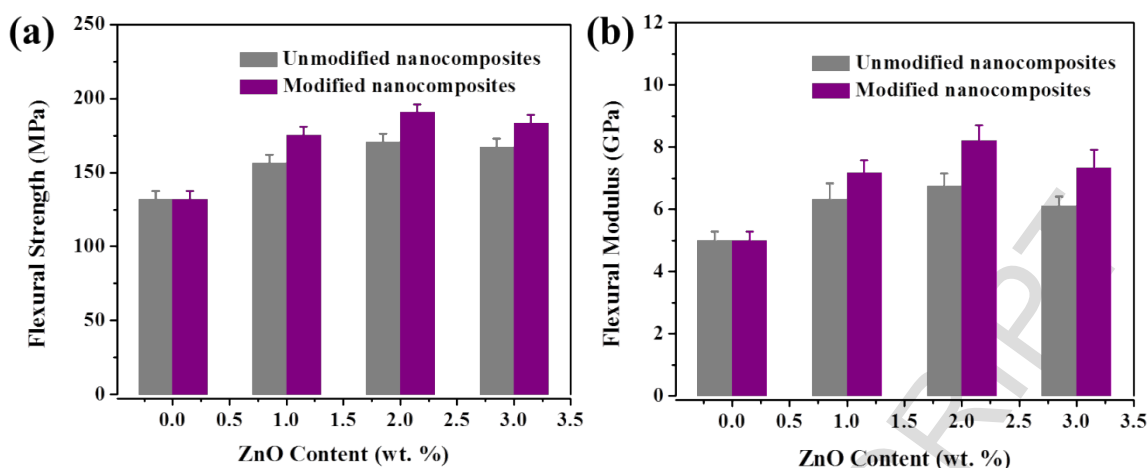


Fig. 11 Flexural properties of epoxy nanocomposites: Variation in (a) flexural strength and (b) flexural modulus; with respect to the concentration of ZN and PZN.

The maximum enhancement in flexural strength and modulus of PZEC-2 compared to the neat epoxy system are ~44% and ~64% respectively. In comparison to unmodified ZnO nanoparticles, the modified ZnO nanoparticles show better flexural properties that may be due to the formation of strong chemical bond at interface between the modified ZnO nanoparticles and epoxy network. Similar results were also reported in the case of other nanofillers based nanocomposites [50]. Thus, it can be inferred that the surface modification of ZnO nanoparticles leads to the significant improvement in mechanical properties of the epoxy nanocomposites.

3.6.3. Fracture toughness and fracture energy

The effect of unmodified and modified ZnO nanoparticles content on fracture toughness (K_{Ic}) and fracture energy (G_{Ic}) of nanocomposites is shown in Fig. 12, where error bars indicate the maximum and minimum values. The K_{Ic} and G_{Ic} values of NE are ~0.46 MPa.m^{1/2} and 37.42 J/m² respectively. The addition of 2 wt.% of modified ZnO particles into the epoxy matrix results an increase in K_{Ic} value of ~0.79 MPa.m^{1/2}, which corresponds to ~49% increase in the fracture toughness. For higher nanoparticle content (3 wt.%), the K_{Ic} reduces to ~0.69 MPa.m^{1/2}, which is found still higher than that of the neat epoxy matrix. Similar trend of K_{Ic} value is observed in case of ZEC and PZEC nanocomposites at 2 wt.% filler concentration. To achieve same enhancement (~49%) in K_{Ic} , the required amount of (wt.%) titania, alumina and silica nanoparticles in epoxy matrix is quite high [51-53]. This reflects the potential of low content of modified ZnO nanoparticles to reinforce the epoxy matrix.

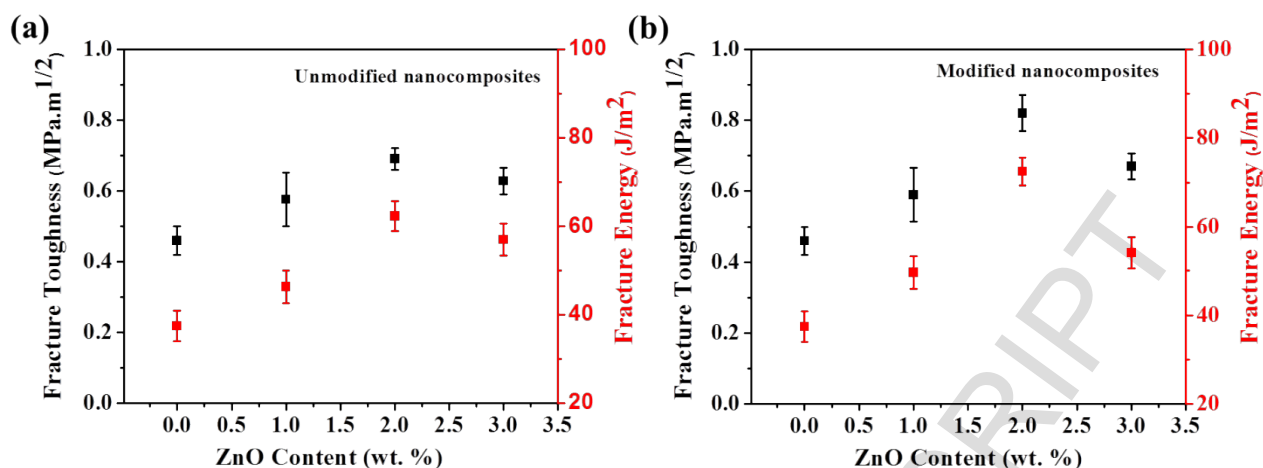


Fig. 12 Fracture toughness of epoxy nanocomposites: Variation of fracture toughness (K_{Ic}) and fracture energy (G_{Ic}) as a function of (a) unmodified and (b) modified ZnO nanoparticle content in epoxy matrix.

The fracture energy of the epoxy nanocomposites have been calculated using the relation: $G_{Ic} = K_{Ic}^2 [(1-\mu^2)/E]$, where μ is Poisson's ratio [54]. The fracture energy (G_{Ic}) enumerates the energy required for crack propagation in a material. The calculated G_{Ic} of NE is 280 J/m². At 2 wt.% loading of PZEC (PZEC-2), the composite shows ~95% enhancement in G_{Ic} (545 J/m²) with respect to that of NE. However, further increase in nanoparticle content up to 3 wt.% (i.e., PZEC-3) leads to a reduction in the fracture energy of the composite. This noticeable enhancement in energy release rate of the nanocomposite is comparable to tough polymers reinforced with other organic or inorganic nanofillers [55]. These results exhibit the potential of low content of modified ZnO nanofiller in toughening the epoxy polymer. All mechanical properties such as tensile strength, compressive strength, elastic modulus, flexural strength, flexural modulus, fracture toughness and fracture energy show a consistent improvement with increase in modified ZnO nanofillers content up to 2 wt.% and on further addition of filler, efficiency gets reduced. This is happening due to clustering of ZnO nanoparticles which aids to diminish the interfacial contact area at ZnO/epoxy matrix interface and hence, it weakens the ability of the ZnO to reinforce the composite. The clustering leads to poor interaction between the nanofiller and epoxy matrix because; the clusters act as defect centers in the epoxy matrix causing a detrimental effect on the mechanical properties.

3.7 Microscopic investigation of fracture surfaces

FESEM images of tensile fracture surface of NE, epoxy nanocomposites containing 2 and 3 wt.% of unmodified and modified ZnO particle content are shown in Fig. 13 (a-e). The fracture surface of NE (Fig. 13a) shows relatively

smooth and glassy texture, which suggests poor absorption of energy during crack propagation **resulting** in brittle fracture. However, the fracture surface of epoxy nanocomposites containing unmodified ZnO nanoparticles such as ZEC-2 and ZEC-3 (Fig. 13(b, d)) shows less rough surface compared to that of epoxy nanocomposites (PZEC-2 and PZEC-3) containing modified ZnO nanoparticles (Fig. 13(c, e)).

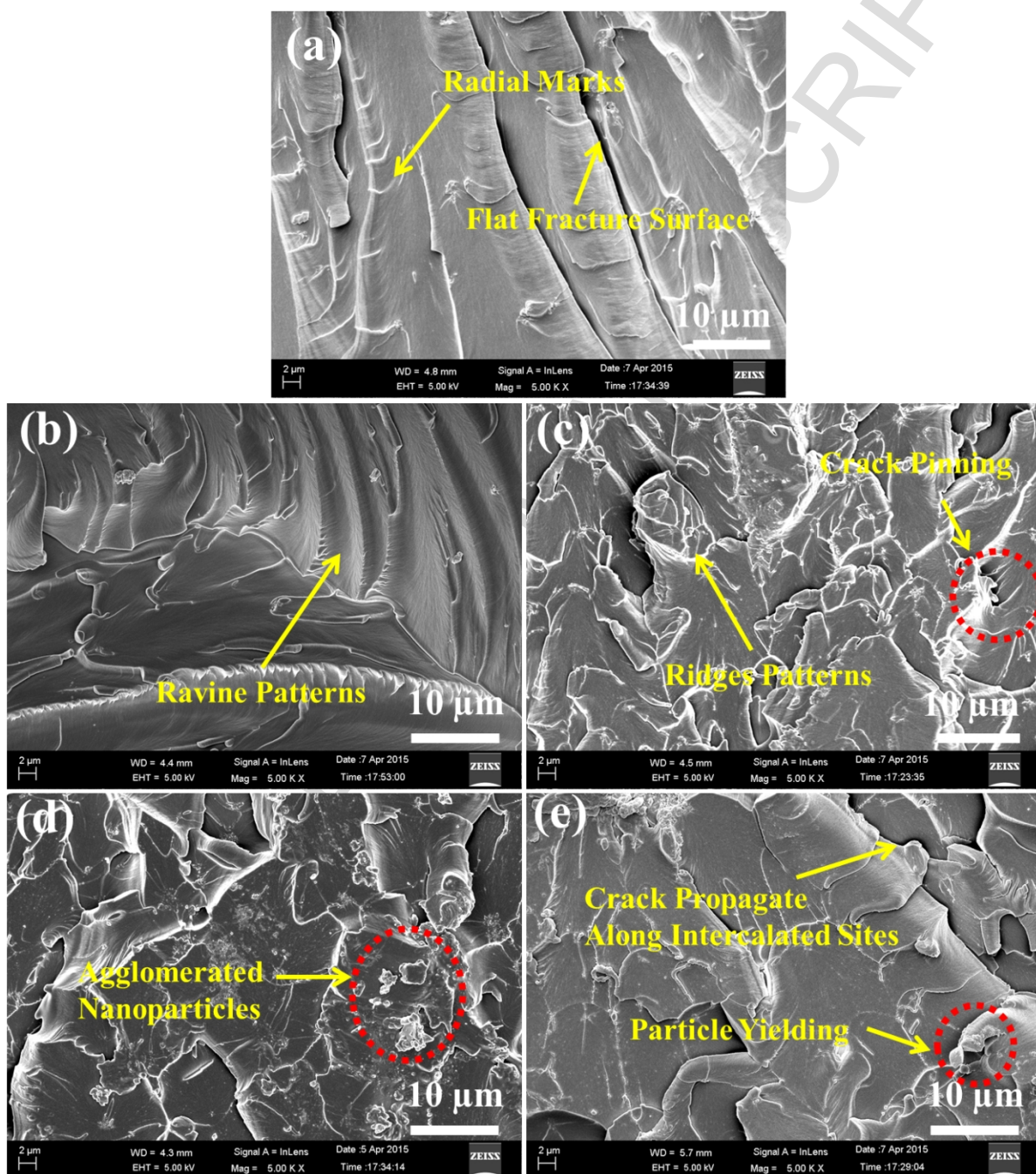


Fig. 13 Tensile fracture surfaces of (a) NE (b) ZEC-2 (c) PZEC-2 (d) ZEC-3 and (e) PZEC-3.

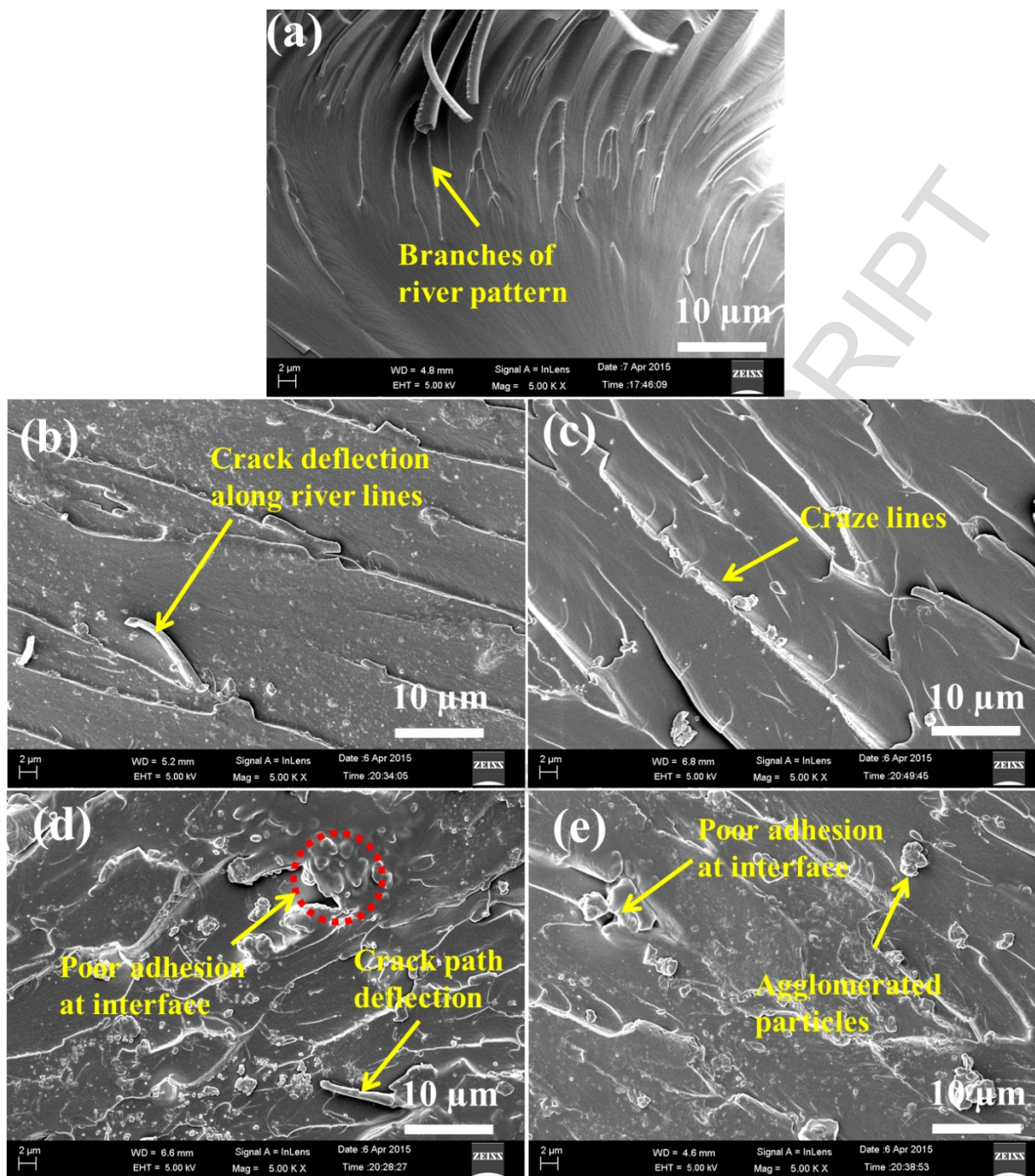


Fig. 14 FESEM micrographs showing the failure mechanisms of the fracture surface obtained from the 3-point bending specimens of (a) NE (b) ZEC-2 (c) PZEC-2 (d) ZEC-3 and (e) PZEC-3.

The fracture surface of the epoxy nanocomposite containing unmodified ZnO nanoparticles shows ravine patterns like texture and the surfaces between the ravine patterns is relatively smooth, which are the characteristic features of brittle fracture behaviour. However, the fracture surfaces of epoxy nanocomposites containing modified ZnO

nanoparticles shows substantial increase in the surface roughness due to the presence of large number of dimples and ridges, which is a measure of enhanced restriction to crack propagation and thus enhanced the toughness of material. The modified ZnO nanoparticles reinforced in highly cross linked polymer network diverts the crack path in different direction thereby delaying the crack propagation and absorbing more energy prior to failure. As a result of crack divergence in random directions, crack pinning and shear yielding zones were observed on the fracture surface which may be attributed to increase in overall tensile strength and stiffness of the nanocomposites [56]. FESEM images of SENB fracture surface of NE, epoxy nanocomposites containing 2 and 3 wt.% of unmodified and modified ZnO particle content are shown in Fig. 14 (a-e). The fracture surface of NE (Fig. 14a) shows the smooth surface with branches of river pattern depicting brittle failure. However, the introduction of ZnO nanoparticles into the epoxy network shows localized shear yielding zones like shear bands, river lines and crazing lines (Fig. 14 (b-e)). These zones may restrict the crack propagation resulting in absorption of more energy prior to failure of the material. But, the craze line features observed on the fracture surfaces of epoxy nanocomposites containing modified ZnO nanoparticles may absorb more energy and result in slow crack propagation and develop more number of shiny river lines and shear yielding zone. Moreover, the appearance of increased number of river lines increases the surface roughness justifying the enhanced fracture toughness of PZEC (Fig. 14 (c, e)) as compared to the ZEC (Fig. 14 (b, d))

Conclusion

Poly vinyl alcohol (PVA)-assisted ZnO nanoparticles with nanorod like morphology was successfully synthesized using sol-gel technique. Better dispersion of 2 wt.% modified ZnO nanoparticles in epoxy polymer was achieved compared to the same concentration of unmodified particles. A lower concentration (~2 wt.%) of unmodified and modified ZnO nanoparticles for obtaining improved thermal and mechanical properties of the epoxy composites was identified. A significant improvement in T_g and minor improvement in thermal stability were achieved for synthesized epoxy nanocomposites. The improvement in modified ZnO nanoparticles reinforced composite was superior to unmodified particle reinforced epoxy composite at same nanofiller content (2 wt.%). The mechanical properties including the tensile strength, tensile modulus, compressive strength, flexural strength and flexural modulus were significantly increased in case of modified ZnO nanoparticles reinforced composite at 2 wt.% nanoparticle concentration. The reduced thermal and mechanical properties of epoxy nanocomposites for higher

concentration of nanoparticles might be due to non-homogeneous particle dispersion in the **highly viscous epoxy resin system**. The tensile and SENB test fracture surfaces of unmodified and modified ZnO/epoxy nanocomposites showed the various toughening mechanisms like branching of river lines, regular river lines and craze lines. The improved roughness of fracture surface of ZnO nanoparticles reinforced composites significantly improved the fracture toughness. These results show the potential of surface modified ZnO nanoparticles to enhance the thermal and mechanical properties of epoxy nanocomposites.

Acknowledgement

This work was supported by start-up research grant from National Institute of Technology Silchar with Grant No. Dean (RC)/457/122. Authors are also thankful to Department of Science and Technology, India to provide support under DST-FIST programme 2014 with Grant No. SR/FST/ ETI-373/2014.

References

- [1] C. Velasco-Santos, A.L. Martínez-Hernández, F.T. Fisher, R.S. Ruoff, V.M. Castaño Improvement of Thermal and Mechanical Properties of Carbon Nanotube Composites through Chemical Functionalization, *Chem. Mater.* 15 (2003) 4470–4475.
- [2] S. Bal, Experimental study of mechanical and electrical properties of cnf/epoxy composites, *Mater. Des.* 31 (2010) 2406–2413.
- [3] S. Halder, P.K. Ghosh, A. Pathak, M.S. Goyat, Influence of nanoparticle weight fraction on morphology and thermal properties of epoxy/TiO₂ nanocomposite, *J. Reinf. Plast. Compos.* 31 (2012) 1180–1188.
- [4] S. Halder, M.S. Goyat, P.K. Ghosh, Influence of ultrasonic dual mode mixing on morphology and mechanical properties of ZrO₂-epoxy nanocomposite, *High Perform. Polym.* 24 (2012) 331–341.
- [5] A. Jumahat, C. Soutis, S.A. Abdullah, S. Kasolang, Tensile properties of nanosilica/epoxy nanocomposites, *Procedia Eng.* 41 (2012) 1634–1640.
- [6] J. Cho, M.S. Joshi, C.T. Sun, Effect of inclusion size on mechanical properties of polymeric composites with micro and nano particles, *Compos. Sci. Technol.* 66 (2006) 1941–1952.
- [7] C. Xu, P. Shin, L. Cao, D. Gao, Preferential Growth of Long ZnO Nanowire Array and Its Application in Dye-Sensitized solar cells, *J. Phys. Chem. C.* 114 (2010) 125–129.
- [8] A. Z. Sadek, S. Choopun, W. Wlodarski, S.J. Ippolito, K.K. Zadeh, Characterization of ZnO Nanobelt-Based Gas Sensor for H₂, NO₂, and Hydrocarbon Sensing, *IEEE Sens. J.* 7 (2007) 919–924.

- [9] T. Goto, S. Yin, T. Sato, T. Tanaka, Morphological Control of Zinc Oxide and Application to Cosmetics, *Int. J. Nanotechnol.* 10 (2013) 1–4.
- [10] A.M. Díez-Pascual, A.L. Díez-Vicente, High-performance aminated poly(phenylene sulfide)/ZnO nanocomposites for medical applications, *ACS Appl. Mater. Interfaces.* 6 (2014) 10132–10145.
- [11] T. Prasad, S. Halder, M.S. Goyat, S.S. dhar, Morphological Dissimilarities of ZnO Nanoparticles and its effect on Thermo-Physical Behavior of Epoxy composites, *Polym. Compos.* (2016) 1–11.
- [12] T. Prasad, S. Halder, Optimization of Parameters and its Effect on Size of ZnO Nanoparticles Synthesized by Sol-gel Method, *Comput. Probl. Solving, Adv. Intelligent Syst. Comput.* 336 (2015) 403–411.
- [13] G.J. Ehlert, H.A. Sodano, Zinc Oxide Nanowire Interphase for enhanced strength in lightweight polymer fiber composites, *Appl. Mater. INTERFACES.* 1 (2009) 1827–1833.
- [14] S. Halder, S. Ahemad, S. Das, J. Wang, Epoxy/Glass Fiber Laminated Composites Integrated with Amino Functionalized ZrO₂ for Advanced Structural Applications, *ACS Appl. Mater. Interfaces.* 8 (2016) 1695–1706.
- [15] M. Murariu, A. Doumbia, L. Bonnaud, A.L. Dechief, Y. Paint, M. Ferreira, C. Campagne, E. Devaux, P. Dubois, High-performance polylactide/ZnO nanocomposites designed for films and fibers with special end-use properties, *Biomacromolecules.* 12 (2011) 1762–1771.
- [16] M. Z. Rong, M.Q. Zhang, W.H. Ruan, Surface modification of nanoscale fillers for improving properties of polymer nanocomposites: a review, *Mater. Sci. Technol.* 22 (2006) 787–796.
- [17] H. baier Zhi-Qiang Yu, Shu-Li You, Effect of Organosilane Coupling Agents on Microstructure and Properties of Nanosilica/epoxy Composites, *Polym. Compos.* (2012) 1516–1524.
- [18] J. Kathi, K.Y Rhee, J.H. Lee, Effect of chemical functionalization of multi-walled carbon nanotubes with 3-APTES on mechanical and morphological properties of epoxy nanocomposites, *Compos. Part A.* 40 (2009) 800–809.
- [19] S. Mallakpour, M. Dinari, E. Azadi, Poly (vinyl alcohol) Chains Grafted onto the Surface of Copper Oxide Nanoparticles : Application in Synthesis and Characterization of Novel Optically Active and Thermally Stable Nanocomposites Based on Poly (amide-imide) Containing N - trimellitylimid, *Int. J. Polym. Anal. Charact.* 20 (2015) 1–16.
- [20] S. Kayal, R. V. Ramanujan, Doxorubicin loaded PVA coated iron oxide nanoparticles for targeted drug delivery, *Mater. Sci. Eng. C.* 30 (2010) 484–490.
- [21] M. Z. Rong, M. Q. Zhang, W. H. Ruan, Surface modification of nanoscale fillers for improving properties of polymer nanocomposites: a review. *Materials Science and Technology* 22 (2006) 787-796.
- [22] K.H. Ding, G.L. Wang, M. Zhang, Characterization of mechanical properties of epoxy resin

- reinforced with submicron-sized ZnO prepared via in situ synthesis method. *Materials and Design* 32 (2011) 3986–3991.
- [23] R. D. Rashmi, T.K. Maji, Effect of Nano-ZnO on Thermal, Mechanical, UV Stability, and Other Physical Properties of Wood Polymer Composites. *Ind. Eng. Chem. Res.* 51 (2012) 3870–3880.
- [24] S. Mallakpour, M. Javadpour, Effective Methodology for the production of Novel Nanocomposites Films based on Poly(vinyl chloride) and Zinc Oxide Nanoparticles Modified With Green PVA, *Polym. Compos.* (2015) 1–10.
- [25] X.L Xie, R.K.Y. Lic, Q.X Liua, Y.W. Mai, Structure-property relationships of in-situ PMMA modified nano-sized antimony tri-oxide filled PVC nanocomposites, *Polymer*. 45 (2004) 2793–2802.
- [26] A. Chatterjee, S. Mishra, Rheological, Thermal and Mechanical Properties of Nano-Calcium Carbonate (CaCO₃)/PMMA Core-Shell Nanoparticles Reinforced Polypropylene Composites, *Polym. Soc. Korea.* (2012) 1–10.
- [27] W. Yang, R. Yi, Xu Yang, M. Xu, S. Hui, X. Cao, Effect of Particle Size and Dispersion on Dielectric Properties in ZnO/Epoxy Resin Composites. *Trans. Electr. Electron. Mater.* 13(3) 116 (2012).
- [28] B. Ramezanzadeh, M.M. Attar. M. Farzam, Effect of ZnO nanoparticles on the thermal and mechanical properties of epoxy-based nanocomposite. *J Therm Anal Calorim* 103 (2011) 731–739.
- [29] D.Q. H. Zhang, D. Yang, D. Li, X. Ma, Controllable Growth of ZnO Microcrystals by a capping-molecules-Assisted Hydrothermal Process, *Cryst. Growth Des.* 5 (2005) 547–550.
- [30] K. Biswas, B. Das, C.N.R. Rao, Growth Kinetics of ZnO Nanorods- Capping-Dependent Mechanism and Other Interesting, *J. Phys. Chem. C.* 112 (2008) 2404–2411.
- [31] K. Yong, Youngio Tak, Controlled Growth of Well-Aligned ZnO Nanorod Array Using a Novel Solution Method, *J. Phys. Chem. C.* 109 (2005) 19263–19269.
- [32] N.R. Yogamalar, R. Srinivasan, A.C. Bose, Multi-capping agents in size confinement of ZnO nanostructured particles, *Opt. Mater. (Amst).* 31 (2009) 1570–1574.
- [33] S. Simon P. Garcia, Controlling the Morphology of Zinc Oxide Nanorods Crystallized from the Aqueous Solutions: The Effect of Crystal Growth Modifiers on Aspect Ratio, *Chem. Mater.* 19 (2007) 4016–4022.
- [34] R. Viswanatha, D.D. Sharma, Study of the Growth of Capped ZnO Nanocrystals: A Route to Rational Synthesis, *Chem Eur. J.* 12 (2006) 180–186.
- [35] C.C. Vidyasagar, Y. Arthoba Naik, Surfactant (PEG 400) effects on crystallinity of ZnO nanoparticles, *Arab. J. Chem.* 9 (2016) 507–510.
- [36] L. Kumari, W.Z. Li, S. Kulkarni, K.H. Wu, W. Chen, C. Wang, C.H. Vannoy, R.M. Leblanc, Effect

- of Surfactants on the Structure and Morphology of Magnesium Borate Hydroxide Nanowhiskers Synthesized by Hydrothermal Route, *Nanoscale Res. Lett.* 5 (2010) 149–157.
- [37] C. Dinh, T. Nguyen, F. Kleitz, T. Do, Shape-Controlled Synthesis of Highly Crystalline Titania Nanoparticles, *ACS Nano*. 3 (2009) 3737–3743.
- [38] K. Akhil, J. Jayakumar, G. Gayathri, S.S. Khan, Effect of various capping agents on photocatalytic, antibacterial and antibiofilm activity of ZnO nanoparticles, *J. Photochem. Photobiol. B: Biology*. 160 (2016) 32–42.
- [39] A.K. Srivastava, M. Deepa, N. Bahadur, M.S. Goyat, Influence of Fe doping on nanostructures and photoluminescence of sol-gel derived ZnO, *Mater. Chem. Phys.* 114 (2009) 194–198.
- [40] A.N. Mallika, A.R. Reddy, K.V. Reddy, Annealing effects on structural and optical properties of ZnO nanoparticles with PVA and CA as chelating agent, *J. Adv. Ceram.* 4 (2015) 123–129.
- [41] R.K. S. Kango, S. kalia, A. celli, J. Njuuna, Y. Habbibi, Surface modification of inorganic nanoparticles for development of organic-inorganic nanocomposites-A Review, *Prog. Polym. Sci.* 38 (2013) 1232–1261.
- [42] S. Mallakpour, M. Dinari, M. Hatami, Dispersion of surface-modified nano-Fe₃O₄ with poly(vinyl alcohol) in chiral poly(amide-imide) bearing pyromellitoyl-bis-l-phenylalanine segments, *J. Mater. Sci.* 50 (2015) 2759–2767.
- [43] P. Li, X. Chen, J.-B. Zeng, L. Gan, M. Wang, Enhancement of the interfacial interaction between poly(vinyl chloride) and zinc oxide modified reduced graphene oxide, *RSC Adv.* 6 (2016) 5784–5791.
- [44] B. Ramezanzadeh, M.M. Attar, M. Farzam, Effect of ZnO nanoparticles on the thermal and mechanical properties of epoxy-based nanocomposite, *J. Therm. Anal. Calorim.* 103 (2011) 731–739.
- [45] K.W. Putz, M.J. Palmeri, R.B. Cohn, R. Andrews, L.C. Brinson, Effect of Cross-Link Density on Interphase Creation in Polymer Nanocomposites, *Macromolecules*. 41 (2008) 6752–6756.
- [46] V.I. Roldughin, O.A. Serenko, E. V Getmanova, N.A. Novozhilova, G.G. Nikifirova, M.I. Buzin, S.N. Chvalun, A.N. Ozerin, A.M. Muzafarov, Effect of Hybrid Nanoparticles on Glass Transition Temperature of Polymer Nanocomposites, *Polym. Compos.* (2016).
- [47] B. Natarajan, Y. Li, H. Deng, L.C. Brinson, Linda S. Schadler, Effect of Interfacial Energetics on Dispersion and Glass Transition Temperature in Polymer Nanocomposites. *Macromolecules* 46 (2013) 2833–2841.
- [48] S. Chandran, J.K. Basu, Effect of nanoparticle dispersion on glass transition in thin films of polymer nanocomposites. *Eur. Phys. J. E* 34 (2011) 1-5.
- [49] Y. Sun, Z. Zhang, K.S Moon, C. P. Wong, Glass Transition And Relaxation Behavior Of Epoxy

- Nanocomposites. *Journal Of Polymer Science: Part B: Polymer Physics*, 42 (2004) 3849–3858.
- [50] Y. Zhou, L. Li, Y. Chen, H. Zou, M. Liang, Enhanced mechanical properties of epoxy nanocomposites based on graphite oxide with amine-rich surface, *RSC Adv.* 5 (2015) 98472–98481.
- [51] H. Zhang, Z. Zhang, K. Friedrich, C. Eger, Property improvements of in situ epoxy nanocomposites with reduced interparticle distance at high nanosilica content, *Acta Mater.* 54 (2006) 1833–1842.
- [52] H. baier Zhi-Qiang Yu, Shu-Li You, Effect of Surface Functional Modification of Nano Alumina Particles on Thermal and Mechanical Properties of Epoxy Nanocomposites, *Adv. Compos. Mater.* 20 (2011) 487–502.
- [53] K. Kumar, P.K. Ghosh, A. Kumar, Improving mechanical and thermal properties of TiO₂-epoxy nanocomposite, *Compos. Part B Eng.* 97 (2016) 353–360.
- [54] K. Wang, J. Wu, L. Chen, C. He, M. Toh, Epoxy Nanocomposites With Highly Exfoliated Pristine Clay: Mechanical Properties and Fracture mechanisms, *Macromolecules.* 38 (2004) 1820–1824.
- [55] N. Domun, H. Hadavinia, T. Zhang, T. Sainsbury, G.H. Liaghat, S. Vahid, Improving the fracture toughness and the strength of epoxy using nanomaterials – a review of the current status, *Nanoscale.* 7 (2015) 10294–10329.
- [56] M.S. Goyat, A.S. Sumit, B. Sumit, S. Halder, P. Ghosh, Thermomechanical response and Toughening mechanisms of a carbon nano bead reinforced epoxy composite, *Mater. Chem. Phys.* 166 (2015) 144–152.

Captions:

Scheme. 1. Schematic representation of reaction mechanism between PVA monomer and hydroxyl groups of ZnO nanoparticles.

Fig. 1 Morphology of (a) unmodified ZnO nanoparticle (ZN) and (b) PVA modified ZnO nanoparticles (PZN).

Fig. 2 (a) Morphology of modified ZnO nanoparticle at high magnification (b) Histogram of particle size distribution of Unmodified ZnO nanoparticle (ZN) and PVA Modified ZnO nanoparticle (PZN).

Fig. 3 (a) FTIR spectra of Unmodified ZnO nanoparticle (ZN) and PVA modified ZnO nanoparticle (PZN).

Fig. 4 TGA thermogram of unmodified ZnO nanoparticles (ZN) and PVA modified ZnO nanoparticles (PZN).

Fig. 5 (a) XRD pattern (b) EDX spectra of Unmodified ZnO nanoparticles (ZN) and PVA modified ZnO nanoparticles (PZN).

Fig. 6 FESEM micrographs of epoxy nanocomposites containing (a) unmodified ZnO nanoparticles (1 wt. %), (b) PVA modified ZnO nanoparticles (1 wt. %).

Scheme. 2. Schematic representation of curing reaction of epoxy network in presence of PZN.

Fig. 7 Variation in glass transition temperature (T_g) with respect to concentration of ZN and PZN in epoxy matrix.

Fig. 8 TGA curve of (a) unmodified ZnO epoxy nanocomposites (b) PVA modified ZnO epoxy nanocomposites.

Fig. 9 Tensile and compressive properties of epoxy nanocomposites: Variation in (a) tensile strength (b) tensile modulus and (c) compressive strength; with respect to the concentration of ZN and PZN.

Fig. 10 FESEM images of epoxy nanocomposites containing (a) 1 wt. % PZN, (b) 2 wt. % PZN, (c) 3 wt. % PZN.

Fig. 11 Flexural properties of epoxy nanocomposites: Variation in (a) flexural strength and (b) flexural modulus; with respect to the concentration of ZN and PZN.

Fig. 12 Fracture toughness of epoxy nanocomposites: Variation of fracture toughness (K_{Ic}) and fracture energy (G_{Ic}) as a function of (a) unmodified and (b) modified ZnO nanoparticle content in epoxy matrix.

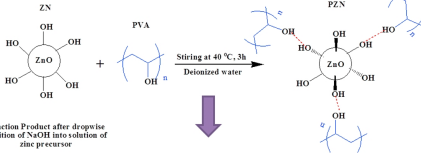
Fig. 13 Tensile fracture surfaces of (a) NE (b) ZEC-2 (c) PZEC-2 (d) ZEC-3 and (e) PZEC-3.

Fig. 14 FESEM micrographs showing the failure mechanisms of the fracture surface obtained from the 3-point bending specimens of (a) NE (b) ZEC-2 (c) PZEC-2 (d) ZEC-3 and (e) PZEC-3.

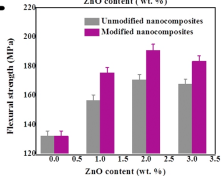
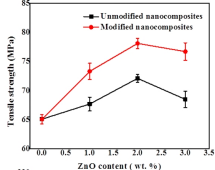
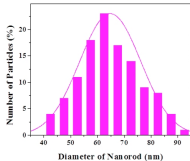
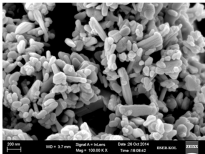
Tables:

Table. 1. Elemental analysis of unmodified and PVA modified ZnO nanoparticles from EDX spectra.

Table. 2. Thermal properties of NE, PVA modified and unmodified ZnO/epoxy nanocomposites.



Reaction Product after dropwise addition of NaOH into solution of zinc precursor



System Reliability Optimization Using Gray Wolf Optimizer Algorithm

Anuj Kumar,^a Sangeeta Pant^a and Mangey Ram^{b,*†}

For the past two decades, nature-inspired optimization algorithms have gained enormous popularity among the researchers. On the other hand, complex system reliability optimization problems, which are nonlinear programming problems in nature, are proved to be non-deterministic polynomial-time hard (NP-hard) from a computational point of view. In this work, few complex reliability optimization problems are solved by using a very recent nature-inspired metaheuristic called gray wolf optimizer (GWO) algorithm. GWO mimics the chasing, hunting, and the hierarchal behavior of gray wolves. The results obtained by GWO are compared with those of some recent and popular metaheuristic such as the cuckoo search algorithm, particle swarm optimization, ant colony optimization, and simulated annealing. This comparative study shows that the results obtained by GWO are either superior or competitive to the results that have been obtained by these well-known metaheuristic mentioned earlier. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: reliability optimization; gray wolf optimizer; metaheuristic approaches; optimization

1. Introduction

Reliability optimization is everywhere, in almost all applications of engineering and industry. Nowadays, computer simulation plays an active role in building the modern engineering design, but it adds some difficulties in the optimization of reliability. Because of the increasing requirement of accuracy and growing complexity in a system's structure, simulation process consumes very much time. It takes several days or even weeks for the evaluation of a single design in many engineering fields. So, a technique is needed to reduce the simulation time and to improve the results so that time and money both can be saved. The tools and techniques that help to achieve the optimized solution are known as search algorithms. In general, optimization problems in reliability theory are non-deterministic polynomial-time (NP)-hard problems. Further, it is more tedious and time-consuming to find the optimal solutions of these problems through exact/heuristic algorithms because these optimization problems generate a very large search space. So the metaheuristic algorithms, which are often nature-inspired, are the most efficient technique for the solution of problems of reliability optimization as they are simple to use and work well even for highly nonlinear, challenging glitches. These algorithms have become so popular because of its simplicity, flexibility, and derivation-free mechanism. Because of the stochastic nature of metaheuristic, they have superior ability to avoid local extrema when compared with its counterpart conventional optimization techniques.

It is worthy to mention No Free Lunch theorem¹ here. According to this theorem, the suitability of a metaheuristics varies from problem to problem. In other words, a metaheuristic that is giving promising results on a particular set of problem may show poor performance on different problems.

The classification of metaheuristics can be carried out on the basis of initial number of candidate solutions. If there is only single candidate solution associated with any metaheuristic, then it can be termed as the single solution-based metaheuristic, for example, simulated annealing (SA). In this class, search process starts with one candidate solution that improves over the course of the iteration. The other class is population-based metaheuristic, in it the search process starts with an initial set of solutions called initial population, which improves over the course of the iteration. Swarm intelligence (SI) is an important branch of metaheuristics, which is based on population. It has been emerging out of the study of ant colonies and swarms of organisms and their social behaviors. These algorithms mostly mimic the social behavior of herds, swarms, or schools of creatures in nature. SI techniques focus on how individuals in groups of organisms behave based on their personal characteristics and social influences. Gray wolf optimizer (GWO),² cuckoo search algorithm (CSA),^{3,4} particle swarm optimization (PSO),⁵ and ant colony optimization (ACO)⁶ are some of the most popular examples of SI techniques.

^aDepartment of Mathematics, University of Petroleum & Energy Studies, Dehradun, India

^bDepartment of Mathematics, Graphic Era University, Dehradun, India

*Correspondence to: Mangey Ram, Department of Mathematics, Graphic Era University, Dehradun, India.

†E-mail: drmrswami@yahoo.com

Exploration and exploitation are two key components for any metaheuristic. Exploration phase refers to exploring the search space more thoroughly and helping in generating diverse solutions for a problem. In order to support exploration phase, an algorithm should have stochastic operators to randomly and globally search the search space. Exploitation uses the local information in the search process for the generation of better solutions. In an algorithm, there should be a fine balance between exploration and exploitation process because the probability of the true optimal outcome will be increased because of much more exploration, and process will slow down. On the other hand, the optimization process converges rapidly by leading the premature convergence because of much more exploration. The rest of the paper is structured as follows. Section 2 presents a literature review of reliability optimization. Section 3 gives the brief overview of GWO algorithm. Problem description and the results and discussion of reliability allocation problems are presented in Sections 4 and 5, respectively. At the end, results conclude in Section 6.

2. Literature review

Generally, the reliability definition is termed as 'the probability that the system will not fail during delivery of service'.⁷ During the designing of a system, a system designer must satisfy different kinds of criteria. These are especially to minimize the cost and to maximize the reliability of the system along with enhancing the comfort and increasing the functional safety. The process of maximizing or minimizing of criteria is termed as optimization process. Reliability optimization problems can be categorized into reliability allocation problems, redundancy allocation problems, and reliability–redundancy allocation problems. This categorization has been performed on the basis of the types of their decision variables.⁸ Analytic integrated factors method based on multicriteria approach has been employed by Di Bona *et al.*⁹ for reliability allocation. Yadav and Zhuang¹⁰ proposed a modified critical measure for allocating system level reliability improvement goal to subsystems. Liaw *et al.*¹¹ used the maximal entropy ordered weighted averaging-based decision-making trial and evaluation laboratory method for reliability allocation, while Silvestri *et al.*¹² proposed a critical flow method for the same. From the viewpoint of mathematical programming, reliability allocation is a continuous nonlinear programming problem (NLP).

Many algorithms, which are categorized into three types, namely, approximate, exact, or heuristic/metaheuristic, are used to find the optimal solutions. In the past two decades, the reliability optimization problems are solved by metaheuristics (Table I), which have been widely treated by many authors.^{13–23} At the very first, to solve the reliability optimization problems, Coit and Smith²⁴ employed a genetic algorithm. Later, Ravi *et al.*²⁰ developed an improved version of non-equilibrium SA called INESA and applied it to solve a variety of reliability optimization problems. Further, Ravi *et al.*²⁵ first formulated various complex system reliability optimization problems with single and multi-objectives as fuzzy global optimization problems. They also developed and applied the non-combinatorial version of another metaheuristic, viz. threshold accepting to solve these problems. Shelokar *et al.*²¹ applied the ACO algorithm to these problems and obtained comparable results to previously reported results. Vinod *et al.*²⁶ applied genetic algorithms (GAs) to risk-informed in-service inspection, which aims at prioritizing the components for inspection within the permissible risk level thereby avoiding unnecessary inspections. A new fuzzy multi-objective optimization problem (MOOP) method is introduced, and it is used for the optimization, decision-making of the series, and complex system reliability with two objectives is presented by Mahapatra and Roy.¹² Mahapatra²⁷ considered a series-parallel system to find out optimum system reliability with an additional entropy objective function. Salazar *et al.*^{28,29} solved the system reliability optimization problem by using several evolutionary algorithms (EAs) and multi-objective evolutionary algorithms (MOEAs). Ravi³⁰ developed an extended version of the great deluge

Table I. Different optimization techniques used in reliability optimization

Model type	Solution techniques	Algorithm description	Source
Redundancy allocation	Heuristic–metaheuristic	SA	Atiqullah and Rao ³⁵
		GA	Coit and Smith ²⁴
		TS	Kulturel-Konak <i>et al.</i> ³⁶
		Variable neighborhood search algorithm	Liang and Chen ³⁷
Reliability allocation	Exact Heuristic–metaheuristic	SA	Wattanapongsakorn and Levitan ³⁸
		Cutting plane algorithm	Majety <i>et al.</i> ³⁹
		Random search algorithm	Mohan and Shanker ³³
		PSO	Pant <i>et al.</i> ⁸
		CSA	Kumar <i>et al.</i> ⁴⁰
Redundancy–reliability allocation	Exact	Surrogate dual problem under DP algorithm	Hikita <i>et al.</i> ⁴¹
		Surrogate constraint algorithm	Hikita <i>et al.</i> ⁴²
		DP	Yalaoui <i>et al.</i> ⁴³

SA = simulated annealing; GA = genetic algorithm; TS = tabu search; PSO = particle swarm optimization; CSA = cuckoo search algorithm; DP = dynamic programming.

algorithm and demonstrated its effectiveness in solving the reliability optimization problems. Deep and Deepti¹³ applied self-organizing migrating genetic algorithm to optimize such type of problems. Recently, Mutingi and Kommula³¹ applied the fuzzy multicriteria genetic algorithm for the reliability optimization of the complex bridge system. Furthermore, different reliability optimization and allocation techniques have been reviewed by Kuo and Prasad¹⁵ and Kuo and Wan.³²

3. Gray wolf optimizer algorithm

Mirjalili *et al.*² introduced the GWO algorithm in 2014. It follows the leadership hierarchy and the chasing policy of gray wolves (*Canis lupus*). The gray wolf is a social animal, and they used to live in a pack of 5–12 members. The whole group can be categorized into four types of gray wolf, which are alpha (α), beta (β), delta (δ), and omega (ω). Alpha holds the highest level in the social hierarchy of the pack, which decreases subsequently up to the omega. A pack of gray wolves can have a male or a female leader, called alpha. The alpha wolf dictates his or her decisions related to hunting, sleeping place, and so on to the pack, and the pack is bound to follow the alpha's decision. The betas, which can be males or females, are subordinate wolves to the alpha, and he or she is probably the best candidate, which replaces an alpha in case of death and aging of an alpha wolf. Betas are responsible for maintaining the discipline in the prey and reinforcing the commands of the alpha. At the same time, they give feedback to the alpha. Omega has the lowest rank in the gray wolf ranking. They are the last wolves that are allowed to eat. If a wolf is not an alpha, beta, or omega, he or she is called delta (or a subordinate wolf). Delta wolves dominate the omega, but they are lower to alphas and betas.

Apart from the social hierarchy, another interesting social behavior of gray wolves is group hunting. They have a well-defined mechanism for hunting. At the very first, they approach the prey by means of tracking and chasing. After that, the pack of gray wolves encircles the prey and harasses it till the movement stop. After that, they attack towards the prey.

The encircling behavior of the pack of gray wolves can be defined with the help of the following mathematical equations²:

$$\vec{D} = \left| \vec{C} \cdot \vec{X}_p(t) - \vec{X}(t) \right| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}_p(t) - \vec{A} \cdot \vec{D} \quad (2)$$

Where \vec{X}_p is the position vector of the prey, while \vec{X} is the position vector of a gray wolf. 't' indicates the current iteration, and \vec{A} and \vec{C} are the coefficient vectors, which can be calculated as follows:

$$\vec{A} = 2\vec{a} \cdot \vec{r}_1 - \vec{a} \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r}_2 \quad (4)$$

Where r_1 and r_2 are random vectors in [0,1], and components of \vec{a} are linearly decreasing from 2 to 0 over the course of iterations.

In the mathematical modeling of the hunting behavior of gray wolves, it is supposed that the alpha (best candidate solution), beta, and delta have better knowledge about the location of the prey. That is why we have saved the first three best solutions obtained so far and oblige the other search agent, including omegas, to update their positions according to the position of the best search agents as per the following formulas²:

$$\vec{D}_\alpha = \left| \vec{C}_1 \cdot \vec{X}_\alpha - \vec{X} \right|, \vec{D}_\beta = \left| \vec{C}_2 \cdot \vec{X}_\beta - \vec{X} \right|, \vec{D}_\delta = \left| \vec{C}_3 \cdot \vec{X}_\delta - \vec{X} \right| \quad (5)$$

$$\vec{X}_1 = \vec{X}_\alpha - \vec{A}_1 \cdot (\vec{D}_\alpha), \vec{X}_2 = \vec{X}_\beta - \vec{A}_2 \cdot (\vec{D}_\beta), \vec{X}_3 = \vec{X}_\delta - \vec{A}_3 \cdot (\vec{D}_\delta) \quad (6)$$



Figure 1. Hunting by pack of gray wolves. This figure is available in colour online at wileyonlinelibrary.com/journal/qre

$$\vec{X}(t+1) = \frac{\vec{X}_1 + \vec{X}_2 + \vec{X}_3}{3} \quad (7)$$

In order to propagate exploration and exploitation, the parameter 'a' is decrease from 2 to 0. Mathematically, $|A| > 1$ force the gray wolves to diverge from the prey in search of a fitter prey (exploration), while $|A| < 1$ force the wolves to attack towards the prey (exploitation) (Figures 1 and 2).

```

Initialize the grey wolf population  $X_i (i=1,2,\dots,n)$ 
Initialize a, A and C
Calculate the fitness of each search agent
 $\vec{X}_\alpha$  = the best search agent
 $\vec{X}_\beta$  the second best search agent
 $\vec{X}_\delta$  = the third best search agent

while (t < Max number of iterations)
  for each search agent
    Update the position of current search agent by equation (7)
  end for
  Update a, A, and C
  Calculate the fitness of all search agents
  Update  $\vec{X}_\alpha$   $\vec{X}_\beta$   $\vec{X}_\delta$ 
  t = t+1
end while
return  $\vec{X}_\alpha$ 

```

Figure 2. Pseudo code of the gray wolf optimization algorithm²⁰

4. Problem description

Reliability engineers generally deal with complex systems that have mixed configuration, that is, connected neither in series nor in parallel configuration. In order to examine the performance of the GWO algorithm on reliability optimization, two reliability allocation problems are considered: they are the complex bridge and life-support system (LSS) in a space capsule.

4.1. Complex bridge system problem

A system whose redundant unit is not in a purely series configuration is considerably more difficult to solve. Figure 3 represents the block diagram of the complex bridge system problem.³³ This system has five components, each having component reliability r_i ($i = 1, 2, 3, 4, 5$). The system in the form of complex bridge has two subsystems. First subsystem comprises components 1 and 4, which are connected in series, and second subsystem has components 2 and 5 in the series. The two subsystems have connected in parallel configuration, and component 3 is inserted.

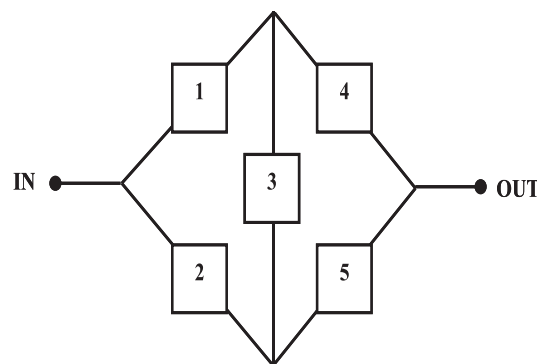


Figure 3. Complex bridge system

The system reliability (R_s) that is the probability of success of the system and system cost (C_s) is given by Equations (8) and (9), respectively. Thus, in general, R_s is the nonlinear function of the variable r_i ($i = 1, 2, 3, 4, 5$). The objective of this problem is to minimize the system cost.

$$R_s = r_1 r_4 + r_2 r_5 + r_2 r_3 r_4 + r_1 r_3 r_5 + 2r_1 r_2 r_3 r_4 r_5 - r_1 r_2 r_4 r_5 - r_1 r_2 r_3 r_4 - r_2 r_3 r_4 r_5 - r_1 r_2 r_3 r_5 - r_1 r_3 r_4 r_5 \quad (8)$$

$$C_s = \sum_{i=1}^5 a_i \exp \left[\frac{b_i}{(1 - r_i)} \right] \quad (9)$$

Here, the authors have considered the optimization of this system with nonlinear constraint as per the mathematical expressions given as follows:

Minimize C_s

subjected to

$$0 \leq r_i \leq 1, \quad i = 1, 2, 3, 4, 5$$

$$0.99 \leq R_s \leq 1,$$

$a_i = 1$ and $b_i = 0.0003$, for $i = 1, 2, 3, 4, 5$.

4.2. Life-support system in space capsule problem

It is imperative to design, create, and analyze a physical habitat for the person who goes on space exploration to live in and be shielded from the rigors of space. At the same time, the LSS for the space capsule must be regenerative, so that the elements that we humans need in order to survive must be supplied. Figure 4 shows the block diagram of LSS in space capsule.³⁴ Here, the system has four components, each having reliability r_i ($i = 1, 2, 3, 4$). The system, which requires a single path for its success, has two redundant subsystems, each comprising components 1 and 4. Each of the redundant subsystems is in series with component 2, and the resultant pair of series-parallel arrangement forms two equal paths. Component 3 is inserted as a third path and backup for the pair. Here, component 1 is backed up by a parallel component 4. There are two equal paths, each of which has component 2 in series with the stage formed by components 1 and 4. These two equal paths operate in parallel so that if one of them is good, the output is assured.

Let U_s denote the probability of system failure, R_i is the probability that the component i is good, and U_i is the probability that the component i is bad. Then, the system unreliability is given by

$$U_s = U_s(\text{given } i \text{ is good})R_i + U_s(\text{given } i \text{ is bad})U_i$$

The corresponding system reliability R_s is $R_s = 1 - U_s$.

For obtaining the reliability of LSS in space capsule presented in Figure 4, component 3 has been taken as the key component i . Thus, the unreliability of the LSS in space capsule is given by

$$U_s = U_s(\text{given } 3 \text{ is good})R_3 + U_s(\text{given } 3 \text{ is bad})U_3$$

If component 3 is good, the system can fail if the stage formed by units 1 and 4 fails. Thus,

$$U_s(\text{given } 3 \text{ is good}) = [(1 - r_1)(1 - r_4)]^2$$

If unit 3 is bad, then

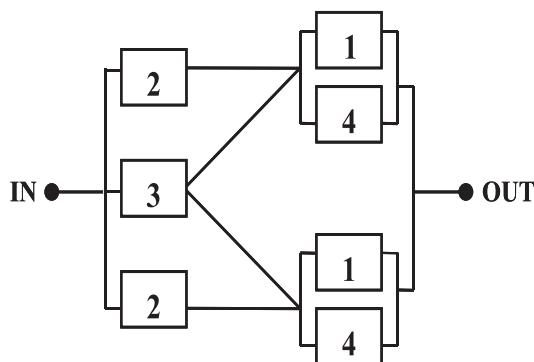


Figure 4. Life-support system in a space capsule

$$U_s(\text{given 3 is bad}) = [1 - r_2\{1 - (1 - r_1)(1 - r_4)\}]^2$$

So, the unreliability of the system is

$$U_s = r_3[(1 - r_1)(1 - r_4)]^2 + (1 - r_3)[1 - r_2\{1 - (1 - r_1)(1 - r_4)\}]^2$$

Thus, the system reliability R_s is given by

$$R_s = 1 - r_3[(1 - r_1)(1 - r_4)]^2 - (1 - r_3)[1 - r_2\{1 - (1 - r_1)(1 - r_4)\}]^2 \quad (10)$$

While, the system cost C_s is

$$C_s = 2K_1r_1^{\alpha_1} + 2K_2r_2^{\alpha_2} + K_3r_3^{\alpha_3} + 2K_4r_4^{\alpha_4} \quad (11)$$

where $K_1 = 100$, $K_2 = 100$, $K_3 = 200$, $K_4 = 150$, and $\alpha_i = 0.6$, $i = 1, 2, 3, 4$.

This problem is continuous nonlinear in nature, where the objective is to find the minimum cost of this complex system subject to the constraints on reliability. The problem formulation is as follows:

Minimize C_s

subjected to

$$0.5 \leq r_i \leq 1 \quad i = 1, 2, 3, 4$$

$$0.9 \leq R_s \leq 1$$

where R_i is i^{th} component's reliability.

5. Results and discussion

In GWO, there is no direct relation between the search agent and fitness function. Various search agents of GWO update their position subjected to the location of the alpha, beta, and delta. Due to this nature of GWO, any kind of constraint handling can be employed without modifying its mechanism. For the aforementioned reliability allocation problems, the simplest constraints handling method,

Table II. Comparison results for complex bridge system problem

Algorithm	Optimum variables					Optimum system cost (C_s)	System reliability (R_s)	FE
	r_1	r_2	r_3	r_4	r_5			
GWO	0.934100	0.936350	0.791370	0.933880	0.935650	5.019900	0.990028	9000
CSA	0.934900	0.935546	0.788534	0.941231	0.927708	5.019980	0.990000	60,000
PSO	0.934821	0.935028	0.791948	0.935005	0.934735	5.019918	0.990000	120,000
ACO	0.933869	0.935073	0.798365	0.935804	0.934223	5.019923	0.990001	80,160
SA	0.935660	0.936740	0.792990	0.938730	0.928160	5.019970	0.990010	

GWO = gray wolf optimization; CSA = cuckoo search algorithm; PSO = particle swarm optimization; ACO = ant colony optimization; SA = simulated annealing.

Table III. Comparison results for life-support system in space capsule problem

Algorithm	Optimum variables				Optimum system cost (C_s)	System reliability (R_s)	FE
	r_1	r_2	r_3	r_4			
GWO	0.500000	0.838920	0.500000	0.500000	641.823600	0.900000	50,000
CSA	0.500000	0.838920	0.500000	0.500000	641.823563	0.900000	15,000
PSO	0.500000	0.838920	0.500000	0.500000	641.823562	0.900000	2040
ACO	0.500000	0.838920	0.500000	0.500000	641.823562	0.900000	20,100
SA	0.500950	0.837750	0.500250	0.500150	641.903000	0.900010	

GWO = gray wolf optimization; CSA = cuckoo search algorithm; PSO = particle swarm optimization; ACO = ant colony optimization; SA = simulated annealing; FE = function evaluated.

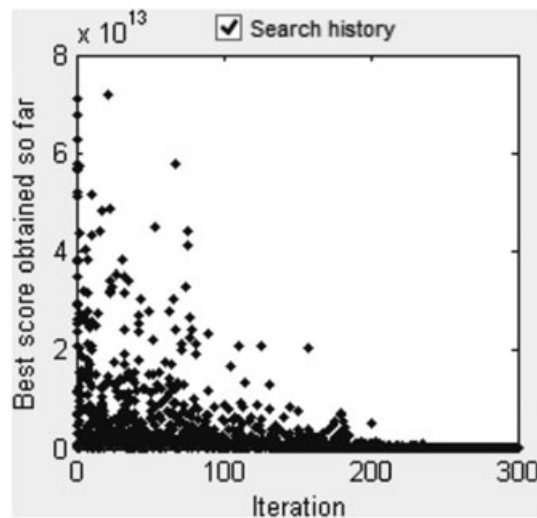
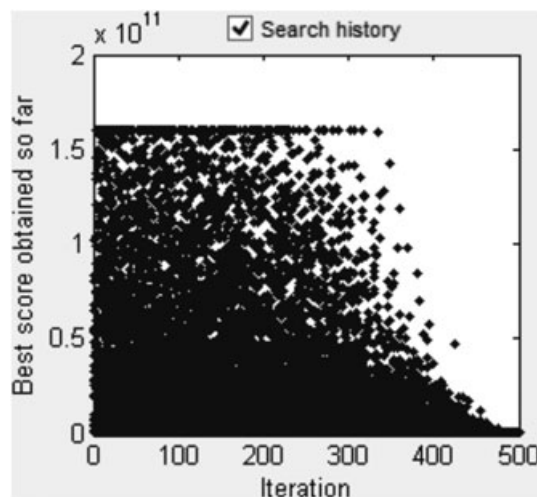
Table IV. Convergence results of C_5 by gray wolf optimization algorithm

	Minimum (best)	Maximum (worst)	Mean	Standard deviation
Complex bridge system	5.019900	5.108700	5.034300	0.028380
Life-support system in space capsule	641.823600	641.842600	641.826000	0.003542

penalty functions have been employed, and the GWO algorithm, which has been implemented in MATLAB, was run and the best results obtained are reported in Tables II and III. The results of the average and standard deviation are revealed in Table IV.

For complex bridge system problem, the number of gray wolves has been fixed at 30 and GWO algorithm executed for 300 iterations. Figure 5 depicts the search history of complex bridge system, and Table II presents the results obtained by GWO algorithm. Apart from that, few previously reported results have also been included in Table II. It shows that GWO provides the most relevant enhancement for the previous solutions obtained by PSO,⁸ SA,²⁰ ACO,²¹ and CSA.⁴⁰ GWO obtained minimum system cost 5.019900 at only 9000 function evaluations, which is the lowest among all, and it kept system reliability at 0.990028, which is higher when compared with system reliability obtained by other methods.

For the LSS in a space capsule, the number of gray wolves has been fixed at 100 and GWO algorithm executed for 500 iterations. Figure 6 depicts the search history for LSS in a space capsule. As mentioned in Table III, a total of 50,000 evaluations of cost function are obtained, and GWO gave 641.823600 as the optimal cost and 0.900000 optimized reliability of the system. Results achieved from

**Figure 5.** Search the history of the complex bridge system**Figure 6.** Search history for life-support system in a space capsule

the present study have been listed in Table III along with the results obtained for the same problem using different metaheuristics. The simulation result indicates that GWO algorithm provides very competitive results.

6. Conclusion and further research

In this work, a new nature-inspired algorithm is implemented for reliability optimization problems. Generally, reliability optimization problems are very complex in nature and NP-hard from the computational point of view. Hence, they are very much difficult to solve as compared with general nonlinear optimization problem. The results obtained by GWO for the complex bridge and LSS in the space capsule show that GWO has high performance on complex reliability optimization problems. Furthermore, the comparative study with the literature having the same problems shows that GWO algorithm has better efficiency as it can provide a solution that is either superior or comparable to the best available results.

Therefore, GWO is proved to be a promising and viable tool to solve the reliability allocation optimization problems. Currently, the authors are investigating a number of improvements and extensions related to the benchmark problems in reliability optimization field.

References

1. Wolpert DH, Macready WG. No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation* 1997; **1**:67–82.
2. Mirjalili S, Mirjalili SM, Lewis A. Grey wolf optimizer. *Advances in Engineering Software* 2014; **69**:46–61.
3. Yang, X. S., Deb, S. Cuckoo search via Levy flights. In *the Proceeding of World Congress on Nature & Biologically Inspired Computing (NaBiC)*, India, 2009, 210–214.
4. Yang XS, Deb S. Engineering optimization by cuckoo search. *International Journal of Mathematical Modelling and Numerical Optimization* 2010; **1**(4):330–343.
5. Kennedy, J., Eberhart, R. Particle swarm optimization, in neural networks, In: *Proceedings, IEEE international conference on*; 1995. pp. 1942–1948.
6. Dorigo M, Birattari M, Stutzle T. Ant colony optimization. *IEEE Computational Intelligence Magazine* 2006; **1**:28–39.
7. Kumar A, Singh SB. Reliability analysis of an n-unit parallel standby system under imperfect switching using copula. *Computer Modeling and New Technologies* 2008; **12**(1):47–55.
8. Pant S, Anand D, Kishor A, Singh SB. A particle swarm algorithm for optimization of complex system reliability. *International Journal of Performability Engineering* 2015; **11**(1):33–42.
9. Di Bona G, Forcina A, Petrillo A, De Felice F, Silvestri A. A-IFM reliability allocation model based on multicriteria approach. *International Journal of Quality & Reliability Management* 2016; **33**(5):676–698.
10. Yadav OP, Zhuang XA. Practical reliability allocation method considering modified critically factors. *Reliability Engineering and System Safety* 2014; **129**:57–65.
11. Mahapatra GS, Roy TK. Fuzzy multi-objective mathematical programming on reliability optimization model. *Applied Mathematics and Computation* 2006; **174**:643–659.
12. Mahapatra GS, Roy TK. Fuzzy multi-objective mathematical programming on reliability optimization model. *Applied Mathematics and Computation* 2006; **174**:643–659.
13. Deep K, Deepthi. Reliability optimization of complex systems through C-SOMGA. *Journal of Information and Computing Science* 2009; **4**:163–172.
14. Kishore A, Yadav SP, Kumar S. A multi objective genetic algorithm for reliability optimization problem. *International Journal of Performability Engineering* 2009; **5**(3):227–234.
15. Kuo W, Prasad VR. An annotated overview of system-reliability optimization. *IEEE Transactions on Reliability* 2000; **49**:176–187.
16. Pant, S., and S. B. Singh. Particle swarm optimization to reliability optimization in complex system. *IEEE Int. Conf. on Quality and Reliability*, Bangkok, Thailand, Sept 14–17, 2011; pp. 211–215.
17. Pant, S., Kumar, A., Kishor, A., Anand, D., & Singh, S.B. Application of a multi-objective particle swarm optimization technique to solve reliability optimization problem. In *The Proceeding of IEEE Int. Conf. on Next Generation Computing Technologies*, September 4–5, 2015; pp. 1004–1007.
18. Rani M, Sharma SP, Garg H. A novel approach for analyzing the behaviour of industrial systems under uncertainty. *International Journal of Performability Engineering* 2013; **9**(2):201–210.
19. Ravi V. Optimization of complex system reliability by a modified great deluge algorithm. *Asia-Pacific Journal of Operational Research* 2004; **21**(4):487–497.
20. Ravi V, Murty BSN, Reddy J. Nonequilibrium simulated-annealing algorithm applied to reliability optimization of complex systems. *IEEE Transactions on Reliability* 1997; **46**(2):233–239.
21. Shelokar PS, Jayaraman VK, Kulkarni BD. Ant algorithm for single and multiobjective reliability optimization problems. *Quality and Reliability Engineering International* 2002; **18**(6):497–514.
22. Yeh WC. A two-stage discrete particle swarm optimization for the problem of multiple multi-level redundancy allocation in series systems. *Expert Systems with Applications* 2009; **36**(5):9192–9200.
23. Zio E, Di Maio F, Martorelli S. Fusion of artificial neural networks and genetic algorithms for multi-objective system reliability design optimization. *Journal of Risk and Reliability* 2008; **222**(2):115–126.
24. Coit DW, Smith AE. Reliability optimization of series-parallel systems using a genetic algorithm. *IEEE Transactions on Reliability* 1996a; **45**:254–260.
25. Ravi V, Reddy PJ, Zimmermann HJ. Fuzzy global optimization of complex system reliability. *IEEE Transactions on Fuzzy Systems* 2000; **8**:241–248.
26. Vinod G, Kushwaha HS, Verma AK, Srividya A. Optimisation of ISI interval using genetic algorithms for risk informed in-service inspection. *Reliability Engineering & System Safety* 2004; **86**:307–316.
27. Mahapatra GS. Reliability optimization of entropy based series-parallel system using global criterion method. *Intelligent Information Management* 2009; **1**:145–149.
28. Salazar DE, Rocco S, Claudio M. Solving advanced multi-objective robust designs by means of multiple objective evolutionary algorithms (MOEA): a reliability application. *Reliability Engineering & System Safety* 2007; **92**:697–706.
29. Salazar D, Rocco CM, Galván BJ. Optimization of constrained multiple-objective reliability problems using evolutionary algorithms. *Reliability Engineering & System Safety* 2006; **91**:1057–1070.

30. Ravi V. Modified great deluge algorithm versus other metaheuristics in reliability optimization. *Computational Intelligence in Reliability Engineering* 2007; **40**:21–36.
31. Mutingi, M., Kommula, V. P., Reliability optimization for the complex bridge system: fuzzy multi-criteria genetic algorithm In *The proceedings of Fifth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing* 437, 2016, pp. 651–663. DOI 10.1007/978-981-10-0451-3_58.
32. Mutingi, M., Kommula, V. P., Reliability optimization for the complex bridge system: fuzzy multi-criteria genetic algorithm In *The Proceedings of Fifth International Conference on Soft Computing for Problem Solving, Advances in Intelligent Systems and Computing* 437, 2016, pp. 651–663. DOI 10.1007/978-981-10-0451-3_58.
33. Mohan C, Shanker K. Reliability optimization of complex systems using random search technique. *Microelectronics Reliability* 1988; **28**(4):513–518.
34. Tillman FA, Hwang CL, Kuo W. Optimization of Systems Reliability. Marcel Dekker Inc.: New York, 1980.
35. Atiqullah MM, Rao SS. Reliability optimization of communication networks using simulated annealing. *Microelectronics Reliability* 1993; **33**:1303–1319.
36. Kulturel-Konak S, Smith AE, Coit DW. Efficiently solving the redundancy allocation problem using tabu search. *IIE Transactions* 2003; **35**:515–526.
37. Liang YC, Chen YC. Redundancy allocation of series-parallel systems using a variable neighborhood search algorithm. *Reliability Engineering & System Safety* 2007; **92**:323–331.
38. Wattanapongsakorn N, Levitan SP. Reliability optimization models for embedded systems with multiple applications. *IEEE Transactions on Reliability* 2004; **53**:406–416.
39. Majety SRV, Dawande M, Rajgopal J. Optimal reliability allocation with discrete cost-reliability data for components. *Operations Research* 1999; **47**:899–906.
40. Kumar A, Pant S, Singh SB. Reliability optimization of complex system by using cuckoos search algorithm. In *Mathematical Concepts and Applications in Mechanical Engineering and Mechatronics*. IGI Global Publisher: Hershey, Pennsylvania, USA, 2016; 95–112.
41. Hikita M, Nakagawa Y, Nakashima K, Yamato K. Application of the surrogate constraints algorithm to optimal reliability design of systems. *Microelectronics and Reliability* 1986; **26**:35–38.
42. Hikita M, Nakagawa Y, Nakashima K, Narihisa H. Reliability optimization of systems by a surrogate-constraints algorithm. *IEEE Transactions on Reliability* 1992; **41**:473–480.
43. Yalaoui A, Châtelet E, Chu C. A new dynamic programming method for reliability & redundancy allocation in a parallel-series system. *IEEE Transactions on Reliability* 2005; **54**:254–261.

Authors' biographies

Anuj Kumar is an Assistant Professor of Mathematics at University of Petroleum and Energy Studies, Dehradun, India. He has obtained his Master's and doctorate degree in Mathematics from G. B. Pant University of Agriculture and Technology, Pantnagar, India. His area of interest is reliability analysis and optimization.

Sangeeta Pant received her doctorate from G. B. Pant University of Agriculture and Technology, Pantnagar, India. Presently, she is working with the department of Mathematics of the University of Petroleum and Energy Studies, Dehradun, as an Assistant Professor. Her area of interest is numerical optimization, evolutionary algorithms, and nature inspired algorithms.

Mangey Ram received the BSc degree in science from Chaudhary Charan Singh University, Meerut, India, in 2000; the MSc degree in mathematics from Hemwati Nandan Bahuguna Garhwal University, Srinagar (Garhwal), India, in 2004; and the PhD degree major in Mathematics and minor in Computer Science from G. B. Pant University of Agriculture and Technology, Pantnagar, India, in 2008. He has been a Faculty Member for around 8 years and has taught several core courses in pure and applied mathematics at undergraduate, postgraduate, and doctoral levels. He is currently Professor with the Department of Mathematics, Graphic Era University, Dehradun, India. Before joining the Graphic Era University, he was a Deputy Manager (Probationary Officer) with Syndicate Bank for a short period. He is Editor-in-Chief of *International Journal of Mathematical, Engineering and Management Sciences*; Executive Associate Editor of the *Journal of Reliability and Statistical Studies* and the Guest Editor and Member of the editorial board of many journals. He is a regular Reviewer for international journals, including IEEE, Elsevier, Springer, Emerald, John Wiley, Taylor and Francis, Inderscience, and many other publishers. He has published 75 research papers in IEEE, Springer, Emerald, World Scientific, Inderscience and many other national and international journals of repute and also presented his works at national and international conferences. His fields of research are reliability theory and applied mathematics. Dr Ram is a member of the IEEE, Operational Research Society of India, Society for Reliability Engineering, Quality and Operations Management in India, International Association of Engineers in Hong Kong, and Emerald Literati Network in the U.K. He has been a member of the organizing committee of a number of international and national conferences, seminars, and workshops. He has been conferred with "Young Scientist Award" by the Uttarakhand State Council for Science and Technology, Dehradun, in 2009. He has been awarded the "Best Faculty Award" in 2011 and recently "Research Excellence Award" in 2015 for his significant contribution in academics and research at Graphic Era University.

Wavelet Variance, Covariance and Correlation Analysis of BSE and NSE Indexes Financial Time Series

Anuj Kumar^{1*}, Sangeeta Pant¹, Lokesh Kumar Joshi²

¹Department of Mathematics, University of Petroleum & Energy Studies, Dehradun-248007, India

²Department of Applied Mathematics, Faculty of Engineering and Technology
Gurukul Kangri Vishwavidyalaya, Haridwar-249404, India

**Corresponding author: anuj4march@gmail.com*

(Received February 26, 2016; Accepted March 8, 2016)

Abstract

The mostly used measure to analyze the stock market behavior is wavelet correlation analysis. Cross-country correlations have been largely used to obtain a static estimate of the comovements of actual returns across country. In this paper wavelet based variance, covariance and correlation analysis of BSE and NSE indexes financial time series have been done using index data from April 1990 to March 2006.

Keywords-stock markets, MODWT, financial time series, BSE & NSE indexes, wavelet variance and covariance.

1. Introduction

A time series is a sequence of data points, measured typically at successive times, spaced at uniform time intervals. Time series analysis comprises methods that attempt to understand such time series, often either to understand the underlying theory of data points or to make forecasts (predictions). Examples of time series are the gross national product, steel production, income per capita, etc. Wavelets are a relatively new way of analyzing time series. Wavelet analysis is in some cases complementary to existing analysis techniques (e.g. correlation and spectral analysis) and in cases capable of solving problems for which little progress has been made prior to the introduction of wavelets (Percival and Walden, 2000).

Traditional time series analysis techniques can be represented as autoregressive integrated moving average models (Bowerman and Connell, 1987; Box and Jenkins, 1976). The traditional models can provide good results when the dynamic system under investigation is linear or non-linear. However, for cases in which the system dynamics are highly nonlinear, the performance of traditional models might be very poor (Cichocki and Unbehauen, 1993; Weigend and Gershenfeld, 1994). The analysis of time series has often been difficult when data do not conform to well study theoretical concepts. One of the most common statistical properties violated by time series data is stationarity. A time series is considered (weakly or second-order) stationary when it has a mean and auto covariance sequence that do not vary with time. It is not uncommon to encounters departures from stationarity in financial time series. Its effects are not limited to the mean of a financial time series, but may also enter into the variance. Other time series exhibit a persistence of correlation much longer than can be explained by short memory (ARIMA) models; they are known as long memory process. The existence of data, such as these, that defy current statistical methods motivates researchers to develop better theories and better tools with which to analyze them. Another concept which arises in the time series analysis is the notion of ‘multi-scale features’ i. e. , an observed financial time series may contain several phenomena, each occurring in different time scales (these correspond to ranges of frequencies in the Fourier

domain). Wavelet techniques possess a natural ability to decompose financial time series into several sub-series which may be associated with particular time scales. Hence, interpretation of features in complex financial time series may be alleviated by first applying a wavelet transform and subsequently interpreting each individual sub-series (Kumar et al., 2010, 2011).

2. Methodology

Variability and association structure of certain stochastic processes can be represented with the help of wavelet methods on a scale-by-scale basis. For a given stationary process $\{X\}$ with variance σ_X^2 , the wavelet variance $\sigma_{X,j}^2$ at scale j have the relationship (Saiti et al., 2014):

$$\sum_{j=1}^{\infty} \sigma_{X,j}^2 = \sigma_X^2 \quad (1)$$

Thus, as $\sigma_{X,j}^2$ represents the contribution of the changes at scale j to the overall variability of the process.

With the help of the above relationship the variance of a time series can be decomposed into components that are associated to different time scales. Spectral density decomposes the variance of the original series with respect to frequency f in the similar manner the wavelet variance decomposes the variance of a stationary process with respect to the scale at j^{th} level i.e.

$$\sum_{j=1}^{\infty} \sigma_{X,j}^2 = \sigma_X^2 = \int_{-1/2}^{1/2} S_X(f) df \quad (2)$$

where $S(\cdot)$ denotes the spectral density function.

By definition the time independent wavelet variance at scale j , $\sigma_{X,j}^2$ is given by the variance of j -level wavelet coefficients $\sigma_{X,j}^2 = \text{var} \{ \tilde{W}_{j,t} \}$.

A time-independent wavelet variance may be defined not only for stationary processes but also for non-stationary processes with stationary d^{th} order differences with local stationarity (Gallegati, 2008). As the wavelet filter $\{h_l\}$ represents the difference between two generalized averages and is related to a difference operator, wavelet variance is time-independent in case of non-stationary processes with stationary d^{th} order differences, provided that the length L of the wavelet filter is large enough. $L \geq d$ is a sufficient condition to make the wavelet coefficients $\tilde{W}_{j,t}$ of a stochastic process stationary whose d^{th} order backward difference is stationary.

As MODWT employs circular convolution, the coefficients generated by both beginning and ending data could be spurious. Thus, if the length of the filter is L , there are $(2^j - 1)(L - 1)$ coefficients affected for 2^{j-1} -scale wavelet and scaling coefficients (Schleer-van and Gellecom, 2014). If $N - L_j \geq 0$, then an unbiased estimator of the wavelet variance based on the MODWT may be obtained by removing all coefficients affected by the periodic boundary conditions using

$$\tilde{\sigma}_{X,j}^2 = \frac{1}{\tilde{N}_j} \sum_{t=L_j}^N \tilde{W}_{j,t}^2, \text{ where } \tilde{N}_j = N - L_j + 1 \text{ is the number of maximal overlap coefficients at}$$

scale j and $L_j = (2^j - 1)(L - 1) + 1$ is the length of the wavelet filter for level j . Thus, the j^{th} scale wavelet variance is simply the variance of the non-boundary or interior wavelet coefficients at that level. Scaling of BSE index financial time series has been done and shown that it is monofractal and can be represented by a fractional Brownian motion (Razadan, 2002).

The MODWT-based estimator has been shown to be superior to the DWT-based estimator although both can decompose the sample variance of a time series on a scale-by-scale basis via its squared wavelet coefficients.

To determine the magnitude of the association between two financial time series of observations X and Y on a scale-by-scale basis the notion of wavelet covariance is used. The wavelet covariance at wavelet scale j can be defined as the covariance between scale j wavelet coefficients of X and Y , i.e.

$$\tilde{\gamma}_{XY,j} = \text{cov}[\tilde{W}_{j,t}^X, \tilde{W}_{j,t}^Y] \quad (3)$$

An unbiased estimator of the wavelet covariance using MODWT can be obtained by removing all wavelet coefficients affected by boundary conditions and given by

$$\tilde{\gamma}_{XY,j} = \left(\frac{N-1}{\tilde{N}_j} \right) \sum_{t=L_j-1}^{N-1} \tilde{W}_{j,t}^X \tilde{W}_{j,t}^Y \quad (4)$$

The MODWT estimator of the wavelet cross-correlation coefficients for scale j and lag τ may be obtained by making use of the wavelet cross-covariance $\tilde{\gamma}_{\tau,XY,j}$, and the square root of the wavelet variances $\tilde{\sigma}_{X,j}$ and $\tilde{\sigma}_{Y,j}$ by

$$\tilde{\rho}_{\tau,XY,j} = \frac{\tilde{\gamma}_{\tau,XY,j}}{\tilde{\sigma}_{X,j} \tilde{\sigma}_{Y,j}} \quad (5)$$

The wavelet cross-correlation coefficients $\tilde{\rho}_{\tau,XY,j}$, just as the usual unconditional cross-correlation coefficients are between 0 and 1 and provide the lead/lag relationships between the two processes on a scale-by-scale basis.

Starting from spectrum $S\omega_{X,j}$ of scale j wavelet coefficients, it is possible to determine the asymptotic variance V_j of the MODWT-based estimator of the wavelet variance and construct a random interval which forms a $100(1-2p)$ % confidence interval.

3. Results and Conclusion

Here, the authors have presented the variance of a process on a scale basis with the help of wavelet analysis. Plot of $\tilde{\sigma}_{X,j}^2$ against scale j indicates which scales contribute more to the process variance. Fig. 1 shows the MODWT-based variance of the BSE Index and NSE Index plotted on a log-log scale. In this figure the straight line “U” and “L” represent the upper and lower bounds for the 95% approximate confidence interval and the straight line shows the valued wavelet variance. Reflection boundary condition has been applied for the calculation of wavelet

variance. Due to this, we have sufficient number of non-boundary coefficients to approximate wavelet variance up to scale 6.

As wavelet analysis have the ability to decompose a financial time series into its time scale components. It is also advantageous in analyzing conditions in which the degree of association between two financial time series is likely to change with the time-horizon. The lead/lag relationship between two financial time series of BSE and NSE has been analyzed on a scale-by-scale source by using wavelet cross-correlation analysis.

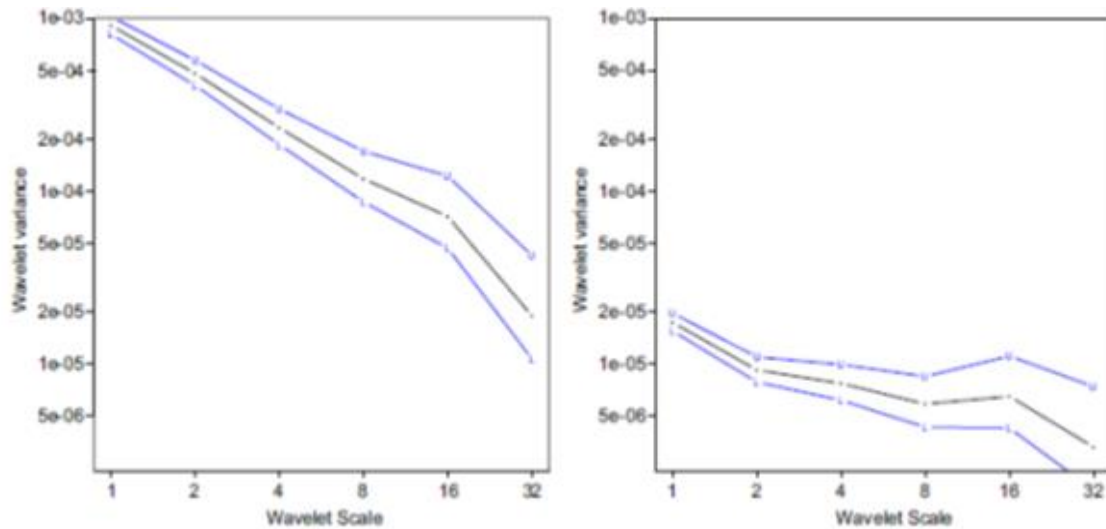


Fig. 1. Wavelet variance for the BSE index and NSE index at log-log scale

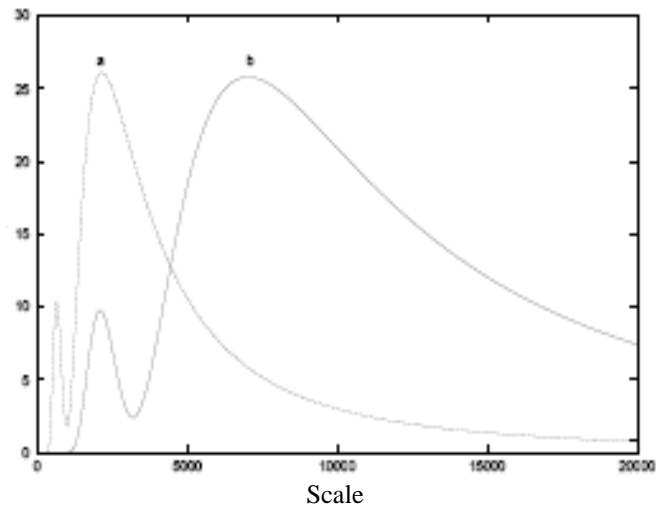


Fig. 2. Wavelet spectrum of NSE (curve a) and BSE (curve b) indices

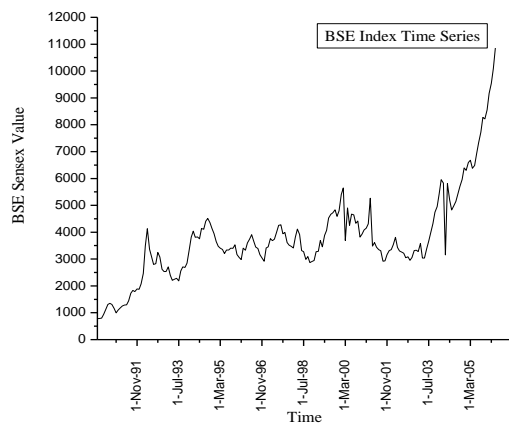


Fig. 3. Financial time series of BSE index

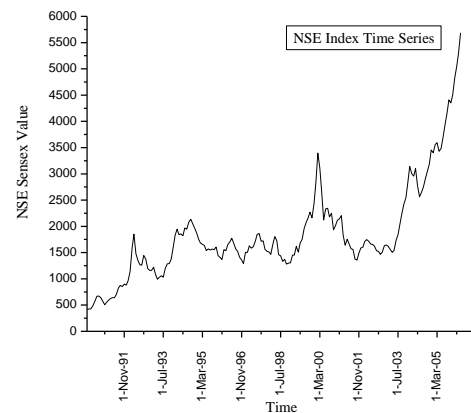


Fig. 4. Financial time series of NSE index

Fig. 2 shows the Wavelet spectrum of NSE and BSE indices and Fig. 3 & Fig. 4 shows the financial time series of BSE and NSE indexes. Fig. 5 exhibits the MODWT-based wavelet correlations and cross-correlation coefficients with the corresponding imprecise confidence intervals. For example, scale 1 is associated to 2–4 month periods, scale 2 to 4–8 month periods, scale 3 to 8–16 month periods, and so on.

The magnitude of the association between the two variables at the shortest scales; i.e. scales 1 to 2; is generally close to zero at all leads and lags, whereas at scales 4 and 5, such connection become stronger. There is a low magnitude of association between BSE and NSE indexes at scales 3 and 4 as the value of wavelet correlation coefficient at lag zero indicates.

On the other hand, the cross-correlation wavelet coefficients 0.5 and 0.7 at scale 4 and 5 reveal high positive foremost relationship between BSE and NSE indexes with the leading period increasing as the time scale increases. It is clear that the largest cross-correlation coefficients going on at leads 6 for wavelet scale 4, that is 16–32 month periods, and 10 for wavelet scale 5, that is 32–64 month periods.

Table 1 and Table 2 represents the monthly average of BSE and NSE Sensex closing index

The averages are based on daily BSE Sensex closing index.

(Base: 1978-79 = 100)

The averages are based on daily closing index.

(Base: 1983-84 = 100)

So, it is clear from the above discussion that correlation between BSE and NSE indices is scale dependent.

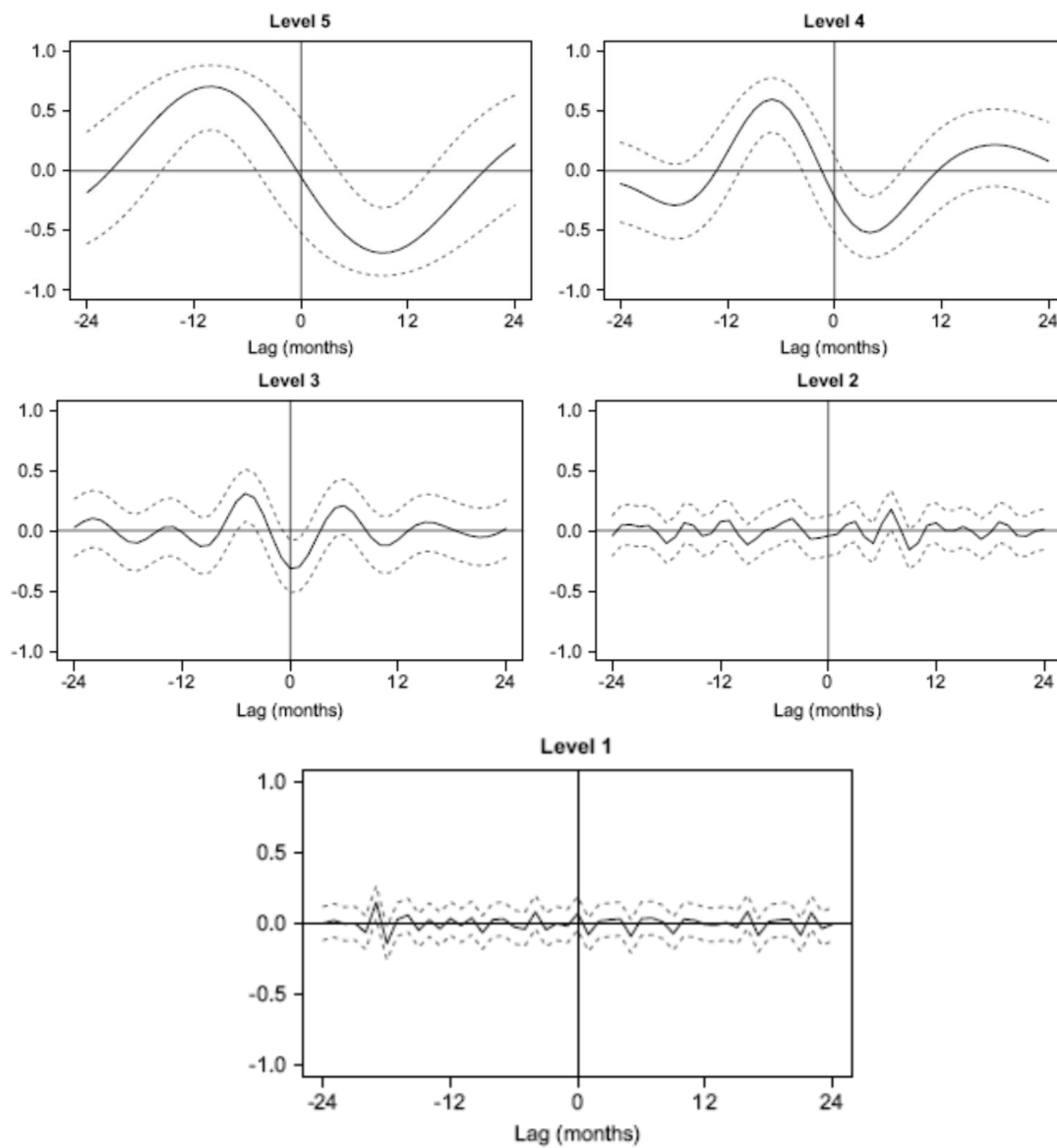


Fig. 5. Wavelet cross-correlations between BSE and NSE indexes

Year/Month	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
1	2	3	4	5	6	7	8	9	10	11	12	13
1990-91	780.18	785.57	802.45	938	1116.19	1307.87	1354.02	1306.09	1161.87	996.45	1100.78	1180.7
1991-92	1255.25	1291.74	1295.15	1440.72	1723.82	1833.34	1789.5	1890.09	1872.31	2073.6	2464.74	3487.19
1992-93	4131.01	3366.55	3088.59	2797.27	2829.96	3243.19	3075.28	2618.2	2535.64	2532.86	2708.72	2398.27
1993-94	2205.37	2248.01	2281.95	2190.34	2556.16	2708.39	2688.51	2850.35	3301.85	3813.74	4039.42	3811.25
1994-95	3824.75	3756.1	4135.67	4106.95	4407.4	4511.34	4351.16	4139.06	3949.78	3651.59	3474.92	3408.29
1995-96	3359.29	3206.86	3336.46	3334.86	3402.81	3396.37	3528.1	3172.02	3060.05	2979.3	3405.56	3327.33
1996-97	3599.66	3732.2	3906.72	3668.21	3449.17	3390.11	3159.79	3044.28	2918.68	3410.3	3453.24	3762.52
1997-98	3681.5	3740.95	4001.47	4256.11	4276.31	3944.79	3991.75	3611.83	3515.54	3472.87	3413.49	3816.87
1998-99	4114.66	3911.95	3317.49	3271.73	2988.4	3089.88	2866.55	2912.39	2945.99	3275.05	3289.24	3689.42
1999-00	3455.05	3880.37	4066.84	4526.25	4662.84	4724.96	4835.47	4588.53	4802.02	5407.14	5650.66	3689.42
2000-01	4905.3	4253.11	4675.4	4647.34	4330.31	4416.61	3819.69	3928.1	4081.42	4152.39	4310.13	5261.77
2001-02	3480.94	3613.84	3439.01	3346.88	3304.99	2918.28	2933.55	3164.25	3314.88	3353.31	3528.58	3807.64
2002-03	3435.13	3302.91	3257.03	3214.87	3053.16	3085.53	2949.76	3058.19	3315.84	3327.66	3278.85	3580.73
2003-04	3036.66	3033.47	3386.89	3665.46	3977.86	4314.74	4742.32	4951.1	5424.67	5954.15	5826.74	3155.7
2004-05	5809.01	5204.65	4823.87	4972.88	5144.17	5423.27	5701.61	5960.75	6393.83	6300.76	6595.05	6679.18
2005-06	6379.29	6482.72	6925.86	7336.7	7726.03	8272.32	8220.45	8552.09	9162.07	9539.67	10090.08	10857.03

Table 1. Monthly average of BSE sensex

Year/Month	Apr.	May	Jun.	Jul.	Aug.	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
1	2	3	4	5	6	7	8	9	10	11	12	13
1990-91	417.99	423.79	428.68	486	573.49	669.06	672.79	644.79	573.77	506.56	553.24	593.69
1991-92	627.72	641.32	639.17	703.02	825.33	872.98	853.77	898.09	880.5	960.14	1138.17	1579.04
1992-93	1850.94	1481.01	1351.06	1267.88	1260.28	1444.95	1376.05	1194.63	1162.92	1160.02	1221.62	1081.17
1993-94	993.63	1029.84	1057.45	1030.1	1199.31	1283.5	1292.72	1368.88	1589.25	1827.17	1945.4	1842.82
1994-95	1855.81	1822.25	1967.76	1947.56	2080.67	2133.49	2054.54	1968.58	1876.13	1755.38	1683.04	1658.97
1995-96	1631.55	1539.44	1570.48	1550.57	1568.33	1555.07	1603.84	1442.44	1406.95	1369.94	1556.09	1539.14
1996-97	1649.6	1701.15	1771.88	1676.6	1575.49	1522.67	1409.83	1356.4	1290.21	1502.66	1504.97	1629.43
1997-98	1586.13	1610.98	1716.56	1844.63	1863.62	1717.52	1722.58	1563.46	1525.78	1512.7	1467.54	1654.92
1998-99	1804.55	1728.93	1459.27	1437.4	1333.8	1371.49	1281.38	1298.19	1307.34	1452.71	1450.6	1620.74
1999-00	1506.84	1682.65	1755.07	1960.83	2075.59	2156.82	2272.13	2161.39	2429.71	2822.05	3394.88	3109.03
2000-01	2663.53	2120.93	2334.27	2344.29	2180.79	2249.43	1931.61	2017.59	2113.84	2140.09	2203.99	1829.32
2001-02	1641.89	1753.46	1661.26	1572.67	1559.95	1373.77	1357.64	1486.33	1587.92	1601.92	1711.43	1746.78
2002-03	1715.11	1661.21	1658.78	1623.07	1536.74	1521.96	1466.79	1510	1632.19	1642.07	1622.58	1559.54
2003-04	1504.62	1538.65	1729.15	1843.86	2055.64	2242.97	2423.87	2543.09	2813.58	3142.23	3003.89	2956.07
2004-05	3101.76	2772.81	2563.78	2653.04	2748.23	2908.81	3049.82	3188.92	3455.28	3403.25	3558.11	3595.2
2005-06	3431.03	3483.19	3697.22	3920.12	4139.35	4407.48	4353.07	4508.71	4825.99	5048.54	5303.59	5686.04

Table 2. Monthly average of NSE index

References

- Bowerman, B. L. & O' Connell, R. T. (1987). *Time series forecasting*. New York PWS.
- Box, G. E. P. & Jenkins, G.M. (1976). *Time series analysis, forecasting, and control*. San Francisco, CA: Holden-Day.
- Cichocki, A. & Unbehauen, R. (1993). *Neural networks for optimization and signal processing*. New York: Wiley.
- Gallegati, M. (2008). Wavelet analysis of stock returns and aggregate economic activity. *Computational Statistics & Data Analysis*, 52(6), 3061-3074.
- Kumar, A., Joshi, L. K., Pal, A. K. & Shukla, A. K. (2010). A new approach for generating parametric orthogonal wavelet. *Journal of Wavelet Theory and Applications*, 4(1), 1-8.
- Kumar, A., Joshi, L. K., Pal, A. K. & Shukla, A. K. (2011). MODWT based time scale decomposition analysis of BSE and NSE indexes financial time series. *International Journal of Mathematical Analysis*, 5(27), 1343-1352.
- Percival, D. B. & Walden, A. T. (2000). *Wavelet methods for time series analysis*. Cambridge University Press.
- Razadan, A. (2002). Scaling in the Bombay stock exchange index. *Pramana*, 58(3), 537-544.

- Saiti, B., Bacha, O., & Masih, A. (2014). *Is the global leadership of the US financial market over other financial markets shaken by 2007-2009 financial crisis? Evidence from Wavelet Analysis*. University Library of Munich, Germany.
- Schleer-van Gellecom, F. (2014). *Advances in non-linear economic modeling- theory and applications*. In *Dynamic Modeling and Econometrics in Economics and Finance*. (Vol.17), Springer.
- Weigend, A. S. & Gershenfeld, N.A. (1994). *Time series prediction: Forecasting the future and understanding the past*. Reading, MA: Addison-Wesley.