We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



2

Applications of Principal Component Analysis (PCA) in Materials Science

Prathamesh M. Shenai[1], Zhiping Xu[2] and Yang Zhao[1]
[1]*Nanyang Technological University*
[2]*Tsinghua University*
[1]*Singapore*
[2]*China*

1. Introduction

Nowadays we are living in the information age with the fast development of computational technologies and modern facilities. Larger data sets are produced by experiments and computer simulations. In contrast to conventional scientific approaches where simple models are built to fit the data, automated procedures are urged to obtain insights into the core messages carried by the large volume of data.

Many problems encountered in materials science involve complicated data models. For example, in biological materials, the collective motion of protein domains usually defines the structural and biological activity of proteins, which should be separated from the irrelevant localized motion of atoms and molecules with high-frequencies. An efficient approach to capture the essential subspace of protein dynamics can remarkably reduce the complexity and directly uncovers the underlying physics (Amadei et al., 1993). On the other hand, nanostructures, which are widely used in nanoscale devices, also have several functional modes that are closely tied to their operation. To visualize them in a thermal and noisy environment requires some insightful treatment (Xu et al., 2008).

Principal component analysis (PCA), as invented by Karl Pearson in 1901, is a procedure to convert a set of correlated variables into uncorrelated ones called principal components (Joliff, 2002). Using mathematical algorithms such as eigenvalue decomposition of the covariance tensor or single value decomposition (SVD), PCA methods find successful applications in many fields as covered in this book. Figure 1 shows the principal modes of ubiquitin in solvent and carbon nanotubes (CNTs) under water flow, as mined from their correlated dynamics in solvents.

In this chapter we will introduce the applications of PCA method in materials science, which not only assist to find useful patterns from the detailed dynamics of atoms and molecules, but also advances the development of PCA technique itself.

2. The mathematics and algorithms of PCA

There are many areas of scientific explorations that lead to enormous quantities of data. Post-processing of such a huge data to extract only the most valuable information is often a

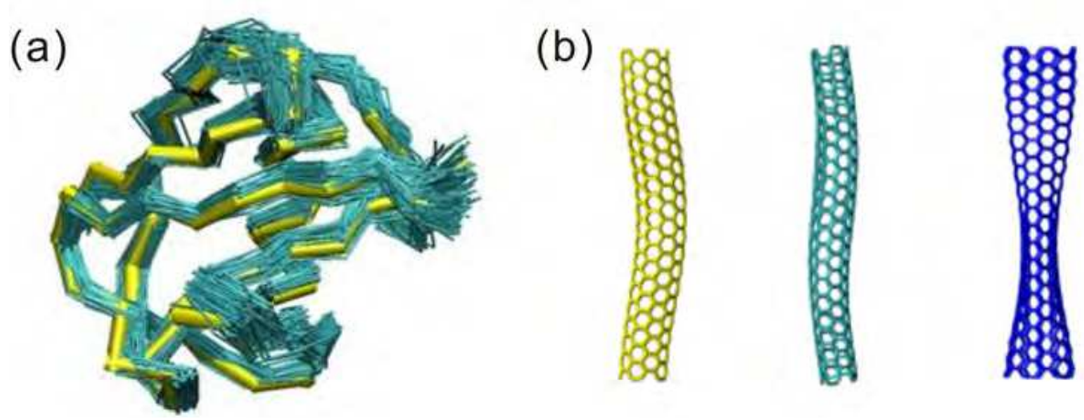


Fig. 1. Applications of principal component analysis (PCA) methods in (a) protein dynamics (Yang et al., 2009) and (b) dynamics of carbon nanotubes under water flow (Chen & Xu, 2011).

tedious task. In a very broad perspective, PCA belongs to a particular set of techniques aimed at reducing a large dataset to a smaller one which can describe the essential characteristics of the underlying system at hand. Molecular dynamics (MD) is a powerful and widely utilized approach in simulating various materials properties and in this chapter, we will focus on the usefulness of PCA in analyzing trajectories generated by MD.

2.1 PCA on MD trajectories

A typical MD trajectory consists of the information of time-evolution of the coordinates of all the constituent atoms forming the system being studied. Commonly used MD timesteps are on the order of 1 fs while the simulation time may range from a few to tens of nanoseconds, in any moderately sized configuration. A single resultant trajectory can thus easily contain a huge amount of data. For an *N*-atom system, the input dataset for PCA can be constructed as a trajectory matrix in which each column contains a cartesian coordinate for a given atom at each output timestep ($x(t)$). Prior to performing PCA, it is ususally necessary to remove any net translational or rotational motion of the system by fitting the coordinate data to a reference structure to obtain the proper trajectory matrix (X). The standardized trajectory data is then utilized to generate a covariance matrix (C), elements of which are defined as

$$C_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{1}$$

where $\langle \ldots \rangle$ denotes an average performed over the all the timesteps of the trajectory. The next step consists of diagonalization of the symmetric $3N$x$3N$ covariance matrix and can be achieved via eigenvector decomposition method as

$$C = T \Lambda T^{\mathrm{T}} \tag{2}$$

where T is a matrix of column eigenvectors and Λ is a diagonal matrix containing the corresponding eigenvalues. This procedure thus transforms the original trajectory matrix in a new orthornormal basis set composed of the eigenvectors. The eigenvalues themselves are indicative of the mean squared displacements of atoms along the corresponding eigenvector. There will be $3N$ resulting eigenvalues if the number of configurations (M) is greater than $3N$. If M<$3N$ there will be the number of eigenvalues will be reduced to M.

The simplest manner of visualizing these results requires sorting the eigenvectors in a descending order in their eigenvalues. The plot of eignevalues against the index of the corresponding eigenvector can then be obtained and is called a 'scree plot'. Characteristically, a scree plot shows that only a few first eigenvectors possess large eignevalues with the higher indexed vectors having eignevalues many orders of magnitude smaller. As a result, most of the variance in the original data is contained and described by only a few first modes. It is then imperative to presume that the motions along these 'essential eigenmodes' dominate the dynamics of the systems and contain the most important global information.

In simple systems, visualization of the components of individual eigenvector can be helpful to gauge the nature of the eigenmode. Followed by identification of a subset of important eigenmodes, further analysis detailing each mode can be undertaken by projecting the original trajectory along a given (or a set of) eigenvector. The corresponding projection matrix (P) can be obtained as

$$P = XT. \tag{3}$$

The time evolution given by the projection matrix yields a manner in which the excitation amplitude of a given eigenvector can be examined. The column vectors in P ($\mathbf{p}$(t)) are called as the 'principal components'.

To analyze the motion along any given eigenvector, the column vector from P multiplied by the corresponding eignevector in T^{T} yields a reduced trajectory containing motion only along the selected mode. Such filtering of modes can be performed for a single or more than one eigenmodes as well and the resulting trajectory provides a visual guidance to the nature of the mode.

A quantitive measure of similarity (S) between different principal modes can be obtained by taking inner product of the corresponding eigenvectors ($\boldsymbol{v}_i$ amd $\boldsymbol{w}_j$) from $\boldsymbol{T}$ as follows:

$$S_{ij} = \boldsymbol{v}_i . \boldsymbol{w}_j \tag{4}$$

The same concept can be further extended to calculate a measure of overlap ($O(\boldsymbol{v},\boldsymbol{w})$) between an essential subspace spanned by eigenvectors $\boldsymbol{w}_j$ (j=1,2,..,n) and another spanned by eigenvectors $\boldsymbol{v}_i$ (i=1,2,..,m) as (Amadei et al. 1999; Hess 2002):

$$O(v,w) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} S_{ij} \tag{5}$$

The overlap will be equal to unity if the subspace spanned by $\boldsymbol{v}_i$ is a subset of $\boldsymbol{w}_j$.

2.2 Computational implementations

Apart from long-time MD simulations to generate sufficient trajectory data, the diagonalization of the 3N X 3N covariance matrix poses the most computationally exhaustive step during PCA. The computational expense as well as memory requirements increase roughly with the square of the number of atoms in the system. As a result, for quite large systems (which can easily be the case when considering large biomolecules), use of efficient algorithms such as QR decomposition is required for matrix diagonalization. Due

to the widespread use of PCA, some existing molecular dynamics programs including open source packages such as GROMACS (Hess et al., 2004) and AMBER (Case et al., 2005) and commercially available Accelrys Materials Studio have incorporated implementations of PCA. Another helpful utility is Interactive Essential Dynamics (IED) which can use the output of PCA performed with GROMACS/AMBER to visualize filtered trajectories via a graphical user interface (Morgan, 2004).

2.3 Demonstrative calculations on a single walled carbon nanotube

Emergence of CNTs and graphene as potential candidates for nanoscale machines has led to their exhaustive probing by using molecular dynamics. It is likely that PCA can prove extremely useful in uncovering many novel dynamical features in such scenarios. In this section, we thus apply PCA to MD simulations of a single walled carbon nanotube (SWNT) with its chirality specified as (5,5). Two different approaches viz. fine-grained and coarse-grained models are studied. The fine-grained approach consists of the regular full atomistic simulations on the SWNT configuration. The other approach adopted from Buehler et al. consists of approximating the structure of the SWNT as finite-sized beads connected with stiff springs (Buehler, 2006).

2.3.1 Fine-grained (fully atomistic) approach

A long (5,5) SWNT configuration with lengths ~ 100 nm (8000 atoms) is considered, a schematic of which is shown in figure 2(a). The intratube C-C interactions are described by Adaptive Intermolecular Reactive Bond Order (AIREBO) potential (Stuart et al., 2000) and MD simulations are performed on the equilibrated structures in a canonical ensemble at 300 K. Temperature control is exercised through the use of Berendsen thermostat (Berendsen et al., 1984).

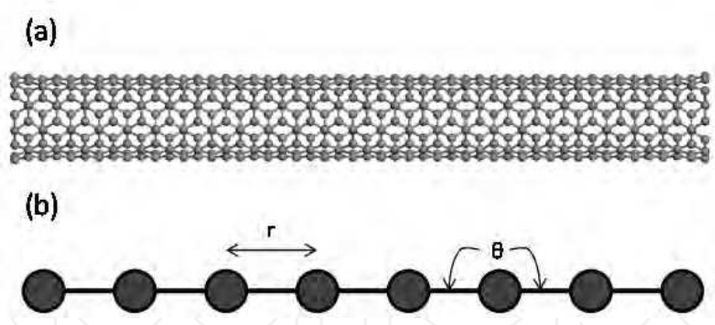


Fig. 2. (a) A schematic of atomistic model of a (5,5) SWNT and (b) a corresponding coarse-grained bead-spring model.

All the simulations are performed using the massively parallelized open source MD software LAMMPS (http://www.cs.sandia.gov/~sjplimp/lammps.html) with a timestep of 1 fs (Plimpton, 1995). At first, the system is thermalized at 300 K for 100 ps. The production run is carried out for 10 ns and the obtained trajectories are subjected to PCA using various tools available in GROMACS. For analyzing the long tube, the production run trajectory is sampled every 50 ps. This sampling rate is chosen to focus on low frequency bending modes and to match the time-scale for a fair comparison with coarse-grained model described in the next subsection.

2.3.2 Coarse-grained approach

Fully atomistic simulations become increasingly computationally prohibitive as the number of atoms in the system grows. As a result, especially when the study of the structural properties at micro-scale is required, the precise atomistic information is rendered redundant. A coarse-graining approach delineated in the work of Buehler et al. can be useful to circumvent the computational expenses and allow for investigation on longer scales. In this approach, a SWNT is essentially modeled as a linear chain of beads connected via springs as depicted in figure 2(b). The properties of individual beads (such as mass) and the springs (such as tensile stretch, angle bend and torsion) can be determined from full molecular dynamics. In this work, we adopt the same approach and coarse-grain a 100 nm long (5,5) CNT as a 100 beads-chain with an equilibrium inter-bead separation distance of 1 nm. All the required parameters can be found in (Buehler 2006) and the dynamics of this system is simulated using LAMMPS. The time-step chosen for the dynamics is 50 fs and the production run is carried out for 10 ns. Using a sampling rate of every 1000 timesteps, PCA is performed on the coordinates' data of all the beads in an analogous way as explained before.

2.3.3 PCA results

Figure 3 shows the scree plot for both the coarse-grained and the atomistic model of the CNT for the first 30 modes. It can be observed in either case that only a few of the first modes occupy high eigenvalue position and thus contain the essential information of the bead dynamics. Modes at higher indices correspond to smaller eigenvalues. Although both the models show a high-eigenvalue first mode, as compared to the coarse-grained model the atomistic model shows a more gradual decrease in the eigenvalues. One of the reasons in the difference is that the correspondance between the similarly indexed modes from the two models is not strictly perfect and as described later, the atomistic system displays a slighly more complicated hierarchy of modes.

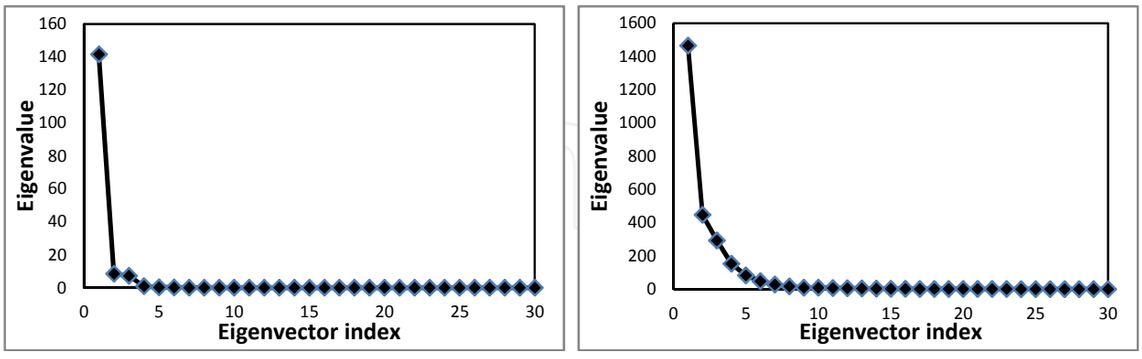


Fig. 3. Eigenvalues against the index of eigenvector obtained from PCA on (a) coarse-grained model of (5,5) SWNT and (b) full atomistic model.

As the coarse-grained system is quite simple, it is intuitive to take a look at the components of individual eigenvectors. For the first five modes, the eigenvector components corresponding to x, y and z coordinates of each bead's center are shown in figure 4. It becomes apparent that the eigenvectors indeed represent the fundamental vibrational modes of the bead-spring system

and its harmonics which resemble to that of the vibrational modes of a strectched spring. Being a slightly more complex system than a vibrating string, the individual principal modes in the bead-spring model can further be seen as the superimposed vibrational modes along X and Y directions. A rough estimation of the mode frequency can also be obtained from their projections on the original trajectory. Note that within a coarse-grained model, any of the modes constituting radial displacements cannot be present and thus, the top principal modes revealed are the bending-like oscillatory modes.

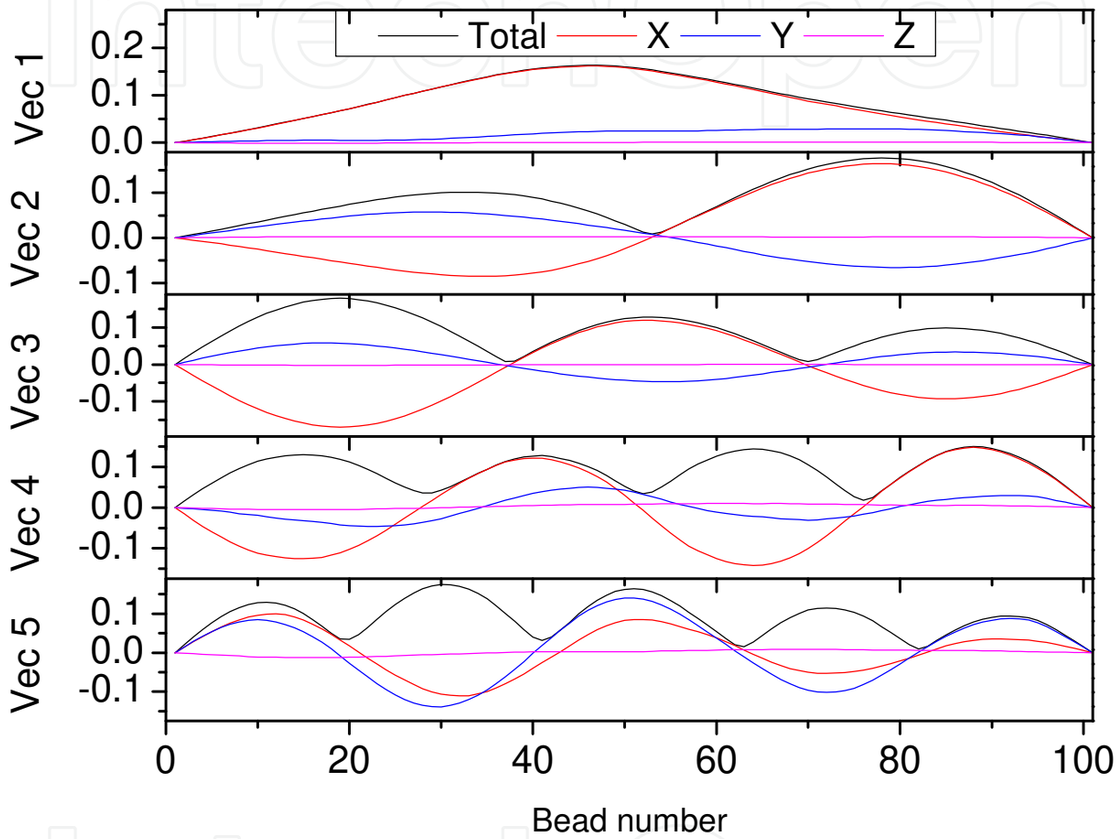


Fig. 4. The components of the most significant five eigenmodes for the (5,5) SWNT. The principal modes resemble fundamental vibrational mode and its harmonics for a stretched string.

The components of the first five eigenvectors of the atomistic model can also be examined in a similar manner and are shown in figure 5 such that the atom numbers are indexed consecutively along the circumference and length. With respect to full atomic contributions (all the X, Y and Z variables) certain qualitative similarities between the mode patterns among the two models can be easily observed. Similar to the bead-spring model, the first eigenmode is the fundamental bending mode of the CNT while the next higher modes represent more or less the sequentially higher harmonics as well. However, a closer look reveals a slightly more complexity, e.g. in the nature of 3rd and 4th principal modes. It can be noted that unlike the bead-spring model, the third mode here appears to be a superimposition of the third harmonic along X-axis and second harmonic along the Y –axis of CNT.

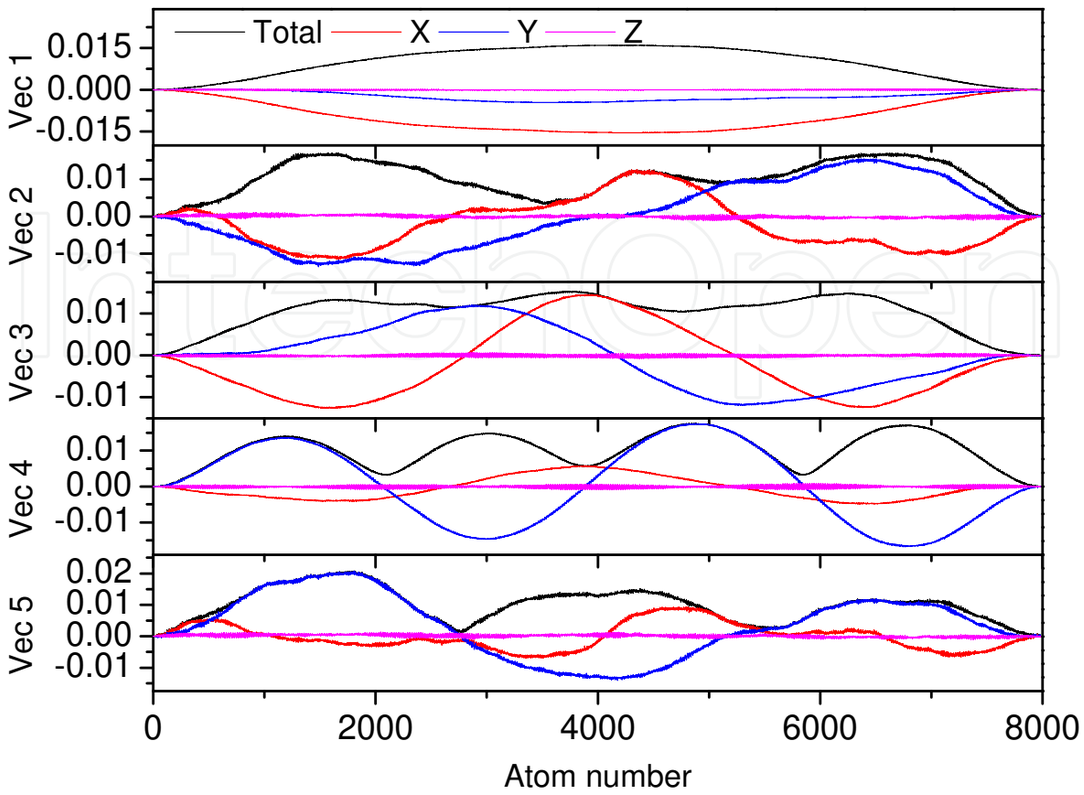


Fig. 5. The components of the first five significant eigemodes for a 100 nm long (5,5) SWNT.

These simple representative calculations thus demonstrate how PCA can help identify various essential modes in a molecular system. In addition to MD simulations, mesoscale simulations of CNTs have started to appear in recent literature. Here, we have purposefully chosen comparisons of atomistic simulations with coarse-grained model of a long carbon nanotube to focus mainly on the out-of plane bending modes. Comparisons between different simulation models for studying material properties at different scales can thus be seen greatly assisted with the use of PCA.

3. Applications in biology, advantages and limitations of PCA

3.1 PCA in biomolecular MD

Biological systems are of immense research interest not only because of the fundamental mysteries of living systems involved, but also because of the possibilities of imitating principles of natural designs in advanced technologies. Proteins, which are the basic building blocks of life, exhibit a striking functional dependence on their conformation. At the cellular level, a variety of biological machinery work precisely amidst extremely noisy environments. Extraction of physical principles that govern such directional dynamics may prove crucial in constructing their artifical counterparts at the molecular level. MD simulations of large biomolecules in fact presented the need of introduction of PCA (Amadei et al., 1993). Amadei et al. proposed that except for the degrees of freedom that belong to the 'essential subspace' of proteins, all the other modes are largely irrelevant

Gaussian fluctuations. The technique quickly became popular in analyzing MD simulations of a large number of biomolecules.

Folding of a protein in a well defined characteristic three-dimensional structure from a random coil structure is one of the most crictical biophysical processes, and PCA has proved vastly useful in its exploration using MD. Ligand binding in proteins such as Myoglobin is strongly influenced by very specific conformational changes near the binding sites. Touriner and Smith investigated MD simulations of hydrated myglobin and found a single principal mode primarily responsible for a dynamical transition appearing at about 180 K (Tournier & Smith, 2003). A class of proteins called membrane proteins such as gramicidin, serves a crucial role of formation of selective ion channels that regulate fluxes across the cell membrane. Recently, Kurylowicz et al. probed the role of anharmonic principal modes such as tilting of peptide planes towards its function as ion channel in gramicidin-A (Kurylowicz et al., 2010). Here we resctrict ourselves to enlist only a few representative examples from the vast literature exists pertaining to PCA in biololecules.

3.2 Inadequacies of PCA

PCA can be performed on either a long single MD trajectory or an ensemble of short trajectories. The latter route is usually advocated since in biomolecular MD simulations, since it is well known that PCA presents difficulties with respect to proper sampling (Balsera et al. 1996; Caves et al. 1998). An excellent analysis about reliability of PCA with respect to sampling issues can be found in the work of Skjaerven et al. (Skjaerven et al., 2011). PCA perfomed on multiple independent runs of the protein systems under the same simulation conditions except the initial atomic velocities, revals noticable differences (de Groot et al., 1998; Skjaerven et al., 2011). While PCA on a single trajectory unambiguosly identifies essential modes during the simulation time, significant differences that can be found among independent runs suggest inadequancy of the sampling of dyamics in the trajectory.

Computational limitations make MD possible on ns timescale while many conformational transitions in proteins, nucleic acids may occur on ms or greater scale. As a result, a single MD trajectory may not entail all possible modes that are essential towards dynamical conformational changes. A direct consequence of such considerations is that even for a single trajectory, the principal modes obtained during one observation window may differ from an another window. While this remains true, a very long (few hundreds of ns) MD simulation may not necessarily yield highly convergent eigenvectors from PCA as compared to simulations with a timespan on the order of tens of ns. Even though efforts have been put in devising methods of enhanced sampling of essential dynamics (Amadei et al., 1999; Hess, 2002), convergence of eigenmodes remains a critical issue.

3.3 Variants of standard PCA

In certain cases, it is also possible and perhaps more useful to disregard less important internal coordinates such as bond lengths and restrict the consideration to dihedral angles. Implementation of PCA based on dihedral angles is commonly referred to as dihedral PCA (dPCA) and was introduced by Mu et al. (Mu et al., 2005). The developement of this approach is mainly aimed at reduction in the dimensionality of the input covariance matrix itself. It has been shown that the dPCA yields results generally equivalent to those obtained

with the conventional cartesian PCA (Altis et al., 2007). Further, instead of MD simulations, it is also possible to use the experimentally generated structural data such as from Nuclear Magnetic Resonance (NMR) or X-ray techniques, for performing PCA. In such a case, an ensemble containing a sufficient number of structural models of the biomolecule needs to be determined from the aforementioned experiments (Howe, 2001; Yang et al., 2009). Although analysis on structural analysis cannot resolve precise atomic motion and is thus of 'coarse' nature as compared to MD simulations, it can still provide a crude approach to compare MD models with experimental data (van Aalten et al., 1998).

3.4 Comparison with Normal Mode Analysis (NMA)

In its standard form, NMA is essentially a harmonic analysis technique which relies on an assertion that the functionally important modes can be extracted as the low frequency normal vibrational modes. The underying assumption is that the conformational energy surface for a given system is approximately parabolic at the global energy minimum. NMA has also been vastly used in structural biology to gain an understanding of the fundamental functional modes in macromolecules such as proteins, lipids and nucleic acids. For a comprehensive account of NMA in reference to biological simulations, reader is referred to an excellent review by Hayward and Go (Hayward & Go 1995).

As NMA demands the structure to be in its lowest energy state, it first needs to be subjected to thorough energy minimization. The next step consists of evaluation of the 'Hessian' (H), which is a matrix of second derivatives of the energy (U) with respect to displacements along cartesian coordinates (x_i), and is calculated as

$$H_{ij} = \frac{\partial U}{\partial x_i \partial x_j} \tag{6}$$

The diagonalization of the mass weighted Hessian ($M^{-1/2}HM^{-1/2}$) where the diagonal matrix M contains the information of atomic masses then yields the eigenvectors and corresponding eigenfrequencies. As opposed to PCA, the normal modes are sorted in ascending manner according to their frequencies.

The fundamental difference between NMA and PCA is in the harmonicity of the resulting modes. Due to the underlying assumption, NMA invariably is restricted to small amplitude harmonic fluctuations around the energy minimum. PCA on the other hand, deals with the positional fluctuations and is thus well suited to study anharmonic vibrations. Furthermore, existing evidence suggests the functional modes in biomolecules to be anharmonic in nature (Amadei et al., 1993; Hayward et al., 1995), which implies that at the physiological temperatures, the underlying assumption of NMA becomes too drastic to be relevant. As a result PCA can be viewed as the more apt technique among the two for exploring dynamical transitions. Yet, standard NMA and its variants such as elastic network NMA have been quite extensively utilized in understanding low frequency functional modes in proteins. Due to the need of long-time MD trajectories, PCA is much more computationaly exhaustive as compared to NMA whereas NMA simply requires a single lowest energy configuration.

4. Applications of PCA in nanomaterials

Compared to biological systems, application of PCA in materials simulations has been sparse. The possible reasons include a lack of material systems for which detailed molecular

motions influence the macroscopic dynamical behavior significantly. However, past couple of decades have witnessed a huge increase in interest in nanoscale transport and mechanical phenomena such as carbon nanotube based hyper-GHz mechanical oscillators, resonators, rotational bearings and actuators (Bourlon et al., 2004; Cumings & Zettl 2000). Double walled CNTs (DWNTs) and multiwalled CNTs are blessed with a rare combination of strong mechanical elements in the form of constituent SWNTs which interact weakly via van der Waal's forces. This sets up an ideal scenario to construct devices in which relative inter-tube rotation or translation can be achieved at the expense of negligible frictional loss. While the required technology at the atomic level is yet to mature, theoretical and molecular dynamics based approaches have opened up proactive paths of investigating characteristics of such nanomachines. This is one of the promising fields of materials research that PCA can fruitfully debut in.

4.1 Analyzing dynamics of CNT based nanomachines

In our previous work, we were able to deduce analogies between dynamics of a rapidly translating SWNT inside a larger SWNT and an aircraft flying near supersonic speed (Xu et al., 2008). It was discovered that for most of the travelling speeds, the core tube can translate without any significant frictional dissipation. However, at certain specific values of axial velocities, abrupt increase in frictional effects can take place. Such kind of energy dissipation points to possibility of resonance effects at particular travelling velocities and is in contrast with the phononic friction commonly observed in nanodevices. We used PCA to gain insights into the nature of modes present in the nanotube-shuttle systems, in one of the first direct applications of PCA in analysing nanodevices.

Using detailed PCA, the underlying principal modes constituting the total motion in the MD trajectories were identified as shown in figure 6. The striking feature in the scree plots corresponding to those initial velocities (1000 m/s and 1900 m/s) at which frictional enhancements appear can be observed in the excitation of high indexed vibrational modes. It was found that at the detrimental critical speed range, a resonance occurs between the 'washboard frequency' and the radial breathing mode (RBM) frequency of the constituent DWNT. The coupling of RBMs with other non-rigid body modes such as bending modes

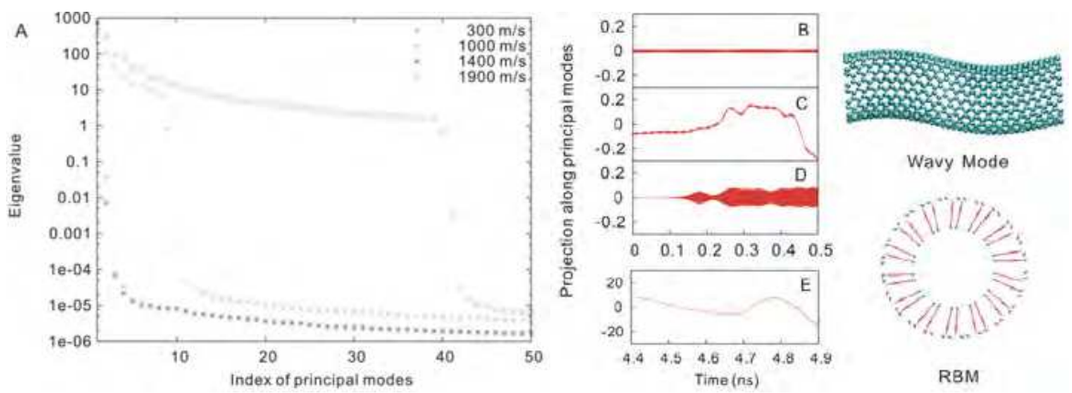


Fig. 6. (A) The scree plot for (7,7)/(12,12) DWNT configuration with different initial travelling speeds of inner nanotube, of which 1000 m/s and 1900 m/s lead to resonant frictional effects. (B)-(D) The projections of wavy, RBM and bending modes for 1000 m/s. (E) Projection of RBM-like mode towards the end of simulation (Xu et al., 2008).

further ensures a nonreversible energy dissipation. Resonant excitation of RBM is evident in figure 6(C) and figure 6(E) that promotes the excitation of various non-rigid body modes such as wavy or bending modes (see figure 6B and 6D). As a result, uncovering new resonant frictional regimes in nanoscale devices by using PCA was demonstrated successfully.

In the case of rotational nanobearings based on DWNTs, Shenai et al. found operational behavior for short sleeved configuration reminiscent of the trans-phonon effect in the translational counterparts (Shenai et al., 2010). It was found that the rotational bearing exhibits a step-like dissipative operation in which at certain angular speeds, the bearing appears to rotate in nearly frictionless manner. The stable rotation, can get hampered however, in such a manner that the angular velocity dissipates more or less abruptly until the bearing stabilizes in a next lower favorable angular speed range. In this case as well, by application of PCA to the bearing trajectory during stable operation and during dissipative operation, it was detected that excitation of dissipative wavy modes takes place during the decay period shown as the 4th eigenmode in figure 7(d).

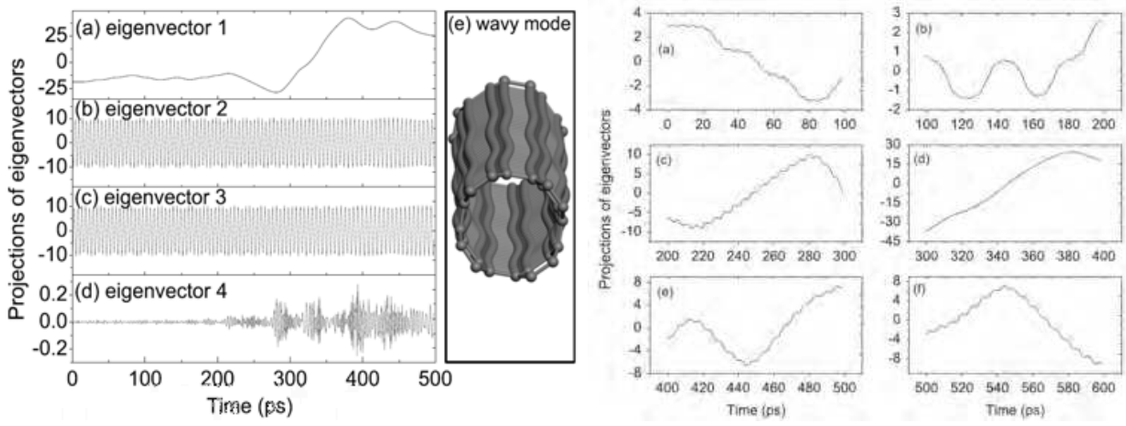


Fig. 7. (a)-(d) Projections of first four eigenvectors on the trajectory of a DWNT based rotational nanobearing. While axial translation reveals itself as the first eigenmode, the rotation is represented by 2nd and 3rd modes together. (e) Depiction of the dissipative wavy mode as the 4th eigenmode. Right panel shows similar analysis for the first eigenvector in different time periods from the initial 600 ps (Shenai et al., 2010).

As a rotational nanobearing was studied in this work, the three lowest eigenmodes are the rigid body modes such as translation (1st eigenvector) and rotation of sleeve (2nd and 3rd eigenvector in combination). More interestingly, it was found that the leakage of the rotational kinetic energy of the sleeve to dissipative wavy modes occurs via another channeling mode – the axial translation of the sleeve. Due to the atomic arrangement of a DWNT, the interaction energy surface between the two tubes exhibits periodic corrugation with respect to relaive axial displacement. Due to typically small corrugation against axial sliding and the small mass of the sleeve, excitation of such translational mode can take place through extraction of a small part of the rotational kinetic energy. When the energy occupation of the axial sliding mode is low, the motion occurs in step-like manner between adjacent energy wells. However when it acquires highly enough translational energies, its enhanced coupling with the higher indexed wavy modes leads to chanelling of the excess

energy as shown in the right panel of figure 7. As soon as the axial oscillation dies, the undesirable channel to the wavy modes gets closed as well, thereby suppressing further decay in rotational kinetic energy. The intricacies of the axial sliding motion and its role in the corresponding excitation of wavy modes was thus successfully resolved.

Negi et al. performed a rigorous study of normal modes via singular value decomposition (SVD) to analyze MD trajectories of single walled carbon nanotubes (Negi & Chaturvedi, 2010) under NVE and NPT conditions. Their approach essentially produces results similar to those with the standard PCA. The full spectrum of principal modes including RBMs and other non-rigid body modes was successfully extracted, see for example, figure 8. In the detailed analysis, they categorized the principal modes according to the uniformities in the displacement characteristics of tube atoms along radial, axial and angular directions. In another subsequent study involving rotational nanomotors driven by external electric field, similar SVD analysis was put to use in understanding the operational regimes and characteristics (Negi et al., 2010).

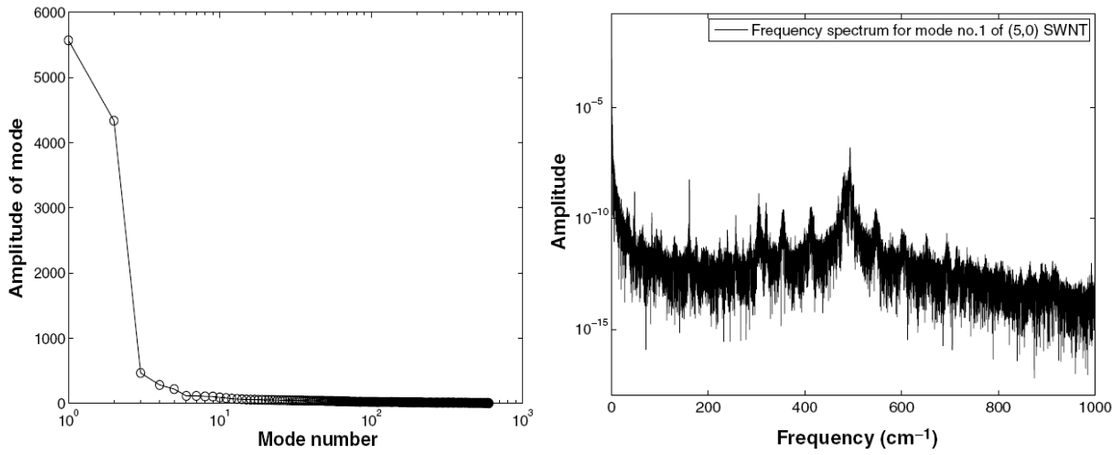


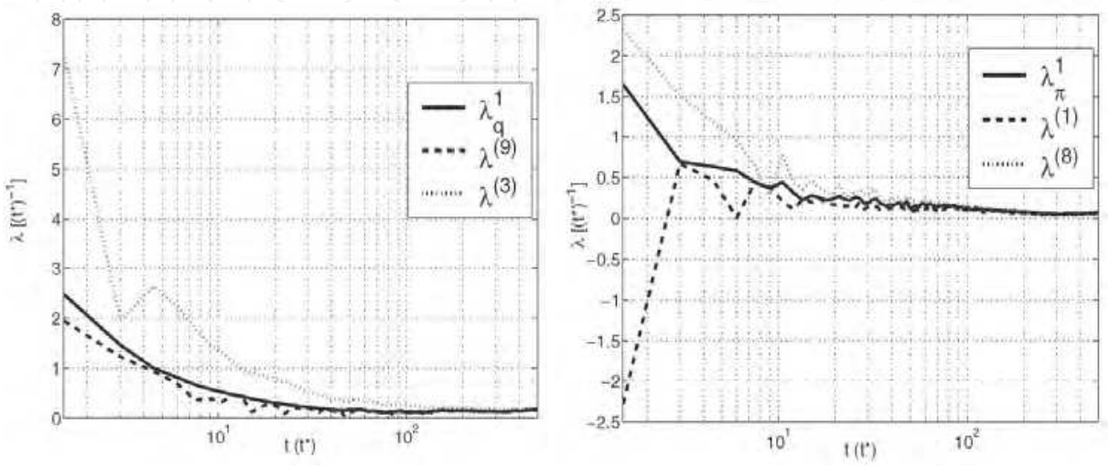
Fig. 8. (a) A typical scree plot obtained from PCA on a (5,0) SWNT. (b) Corresponding power spectrum showing a peak corresponding to frquency of RBM, which is first principal mode (Negi & Chaturvedi 2010).

The aforementioned studies involving the most basic types of CNT based nanomachines demonstrate the usefulness of PCA in analyzing their dynamical features. In future studies probing frictional dissipation or energy channeling between different modes in various nanomechanical devices, PCA can be expected to prove significantly helpful in providing valuable insights.

4.2 Applications in non-linear dynamics of other materials

In another interesting approach by Battisti et al., PCA was innovatively used to study coherent and chaotic dynamics of a small molecule, butane (Battisti et al., 2009). Characterization of chaoticity in the butane molecule was based on evaluation of Lyapunov exponent (LE), which essentially determines the exponential rate of divergence between two trajectories in the phase space separated by a very small distance in their initial conditions.

In this study, a conjecture that in chaotic systems at low energies, different degrees of freedom may entail different degrees of chaoticity was examined. The 'essential' degrees of freedom were obtained as the eigenmodes obtained from PCA on the MD trajectories of butane. Using the individual trajectories reconstructed by projecting different eigenvectors on the original trajectory, it was possible to calculate LE for individual degree of freedom (in terms of principal modes). It was revealed that depending upon the system temperature, there exists a hierarchy of degrees of freedom with respect to 'coherence time' – a measure of degree of order. Certain degrees of freedom exhibit more chaoticity than the system as whole, may exhibit lower chaoticity as shown in figure 9.



(a) (b)

Fig. 9. Differential evolution of Lyapunov exponent during short time for the trajectories filtered along different principal modes for a butane molecule investigated with molecular dynamics in (a) coordinate space at 180 K, and (b) velocity space at 147 K (Battisti et al., 2009).

In general, at relatively high temperatures, the first two eigenmodes turn out to be the most coherent degrees of freedom. Such hierarchy was further shown to vary significantly at varying temperatures as well as under the particular subspaces (coordinate or velocity) at which calculations are performed.

5. Conclusions and perspectives

In this chapter, we have presented a comprehensive account of PCA and its applications in different fields of materials science, in particular. An overview of the underlying theory is presented followed by demonstration of its applications in the study of SWNTs, based on two different approaches – coarse-grained simulations and fully atomistic fine-grained simulations. The results emphasizing the importance of 'essential subspace' and identification of lowest principal modes are presented with respect to the two models along with comparisons between them.

While it has been extensively applied in the studies of biomolecules over the past two decades, possibilities of its usage in the study of materials, have started to emerge only

recently. The vast applications of PCA in structural biology have led to developements of its variants such as coarse-grained PCA or dihedral-PCA. Despite being highly successful, concerns regarding accuracy and robustness of PCA such as sampling issues must be addressed very carefully. Such limitations have not been thoroughly investigated when PCA is employed in the study of materials. Strictly speaking, direct application of PCA in core materials science is sill quite limited. Yet, the emerging field of nanomachines based on carbon nanotubes or graphene focussed on MD, undoubtedly stands out as the most promising area where PCA appears to be quite useful in understanding dynamical characteristics.

6. References

Altis, A.; Nguyen, P. H.; Hegger, R. & Stock, G. (2007). Dihedral angle principal component analysis of molecular dynamics simulations. *Journal of chemical physics,* Vol. 126, pp. 244111

Amadei, A.; Linssen, A. B. M. & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics,* Vol.17, No.4., pp. 412-425

Amadei, A.; Ceruso, M. A. & Nola, A. D (1999). On the convergence of conformational coordinates basis set obtained by the essential dynamics analysis of proteins' molecular dynamics simulations. *Proteins*, Vol.36, pp. 419-24

Balsera, M. A.; Wriggers, W.; Oono Y. & Schulten, K. (1996). Principal Components Analysis and Long Time Protein Dynamics. *Journal of Physical Chemistry.* Vol.100, pp. 2567-72

Battisti, A.; Lalopa, R. G.; Tenenbaum A. & and D'Alessandro M. (2009). Ordered and chaotic dynamics of collective variables in a butane molecule. *Phys. Rev.* E, Vol.79, pp.46206

Berendsen H. J. C.; Postma J. P. M.; van Gunsteren W. F.; DiNola A. & Haak J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* Vol.81, pp. 3684-90

Bourlon B.; Glattli, D. C.; Miko, C.; Forro, L. & Bachtold A. (2004). Carbon Nanotube Based Bearing for Rotational Motions. *Nano Letters*, Vol.4, No.4, pp. 709-712

Buehler, M. J. (2006). Mesoscale modeling of mechanics of carbon nanotubes: self-assembly, self-folding, and fracture. *J. Mater. Res.,* Vol.21, pp. 2855-69

Case, D. A.; Cheatham III, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz Jr., K.M. ; Onufriev, A.; Simmerling, C.; Wang, B. & Woods R. (2005). The Amber biomolecular simulation programs. *J. Computat. Chem.,* Vol.26, pp. 1668-88.

Caves, L. S. D.; Evanseck, J. D. & Carplus, M. (1998). Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin. *Protein Science,* Vol.7, No.3, pp. 649-66

Chen, C.; Ma, M.; Jin, K.; Liu, J. Z.; Shen, L. M.; Zheng, Q. S. & Xu, Z. (2011) Nanoscale fluid-structure interaction: Flow resistance and energy transfer between water and carbon nanotubes. *Physical Review E*, Vol. 84, pp. 046314

Cumings J. & A. Zettl (2000). Low-friction nanoscale linear bearing realized from multiwall carbon nanotubes. *Science,* Vol.289, No.5479, pp. 602-604

de Groot, B.; Hayward, S.; van Aalten, D.; Amadei, A. & Berendsen, H. (1998). Domain motions in bacteriophage T4 lysozyme: A comparison between molecular dynamics and crystallographic data. *Proteins,* Vol.31, pp. 116-27

Hayward S. & Go, N. (1995). Collective variable description of native protein dynamics. *Annu. Rev. Phys. Chem.,* Vol.46, pp.223-50

Hayward S.; Kitao, A.; & Go, N. (1995). Harmonicity and anharmonicity in protein dynamics: a normal modes and principal components analysis. *Proteins: Structure, Function, and Bioinformatics,* Vol.23, No.2, pp. 177-86

Hess B. (2002). Convergence of sampling in protein simulations. *Phys. Rev. E,* Vol.65, pp. 031910

Hess, B., Kutzner, C.; van der Spoel, D. & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, Vol.4, pp. 435-47

Howe ,P. W. (2001). Principal components analysis of protein structure ensembles calculated using NMR data. *J. Biomol NMR* Vol.20, No.1, pp. 60-70

Jolliffe, I. T. (2002). *Principal Component Analysis*, Springer

Kurylowicz, M.; Yu, C. H. & Pomes, R. (2010). Systematic Study of Anharmonic Features in a Principal Component Analysis of Gramicidin A. *Biophys. J.* Vol.98, No.3, pp. 386-95

Morgan, J (2004). Ineractive essential dynamics. *J. Computer-Aided Molecular Design,* Vol.18, pp. 433-36

Mu, Y.; Nguyen, P. H. & Stock, G. (2005). Energy Landscape of a Small Peptide Revealed by Dihedral Angle Principal Component Analysis. *Proteins,* Vol.58, pp.45

Negi, S. & Chaturvedi, S. (2010). Normal mode analysis of a single-walled carbon nanotube based on molecular dynamic: A singlular value decompopsition study. *Int. J. Nanosci.,* Vol.9, No.5, pp. 471-86

Negi, S.; Warrier, M. & Chaturvedi, S. (2010). Determination of useful parameter space for a double-walled carbon nanotube based motor subjected to a sinusoidally varying electric field. *Comp. Mater. Sci.,* Vol.50, pp. 761-70.

Plimpton, S. (1995). Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.,* Vol.117, pp. 1-19.

Skjaerven, L.; Martinez, A. & Reuter, N. (2011). Principal component and normal mode analysis of proteins; a quantitative comparison using the GroEL subunit. *Proteins: Structure, Function, and Bioinformatics,* Vol.79, No.1, pp. 232-243

Shenai P. M.; Ye, J. & Zhao, Y. (2010). Sustained smooth dynamics in short sleeved double-walled carbon nanotubes. *Nanotechnology,* Vol.21, No.49, pp. 495303

Stuart S. J.; Tutein A. B. & Harrison J. A. (2000). A reactive potential for hydrocarbons with intermolecular interactions . *J. Chem. Phys.,* Vol.112, pp. 6472-86

Tournier, A. L. and Smith, J. C. (2003). Principal components of the protein dynamical transition. *Phys. Rev. Lett.,* Vol. 91, No.20, pp. 208106

van Aalten, D. M. F.; Grotewold, E. & Joshua-Tor, L. (1998). Essential dynamics from NMR clusters: Dynamic properties of the Myb DNA-binding domain and a hinge-bending enhancing variant. *Methods-A Companion to Methods in Enzymology,* Vol.14, No.3, pp 318-28

Xu, Z.; Zheng, Q.; Jiang, Q.; Ma, C.-C.; Zhao, Y.; Chen, G.; Gao, H. & Ren, G. (2008). Trans-phonon effects in ultrafast nano-devices. *Nanotechnology,* Vol.19, No.25, pp. 255705-255705

Yang, L.-W.; Eyal, E.; Bahar, I. & Kitao, A. (2009). Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. *Bioinformatics,* Vol.25, No.5, pp. 606-614

IntechOpen

IntechOpen



Principal Component Analysis - Engineering Applications

Edited by Dr. Parinya Sanguansat

ISBN 978-953-51-0182-6
Hard cover, 230 pages
Publisher InTech
Published online 07, March, 2012
Published in print edition March, 2012

This book is aimed at raising awareness of researchers, scientists and engineers on the benefits of Principal Component Analysis (PCA) in data analysis. In this book, the reader will find the applications of PCA in fields such as energy, multi-sensor data fusion, materials science, gas chromatographic analysis, ecology, video and image processing, agriculture, color coating, climate and automatic target recognition.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Prathamesh M. Shenai, Zhiping Xu and Yang Zhao (2012). Applications of Principal Component Analysis (PCA) in Materials Science, Principal Component Analysis - Engineering Applications, Dr. Parinya Sanguansat (Ed.), ISBN: 978-953-51-0182-6, InTech, Available from: http://www.intechopen.com/books/principal-component-analysis-engineering-applications/applications-of-principal-component-analysis-pca-in-materials-science

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

© 2012 The Author(s). Licensee IntechOpen. This is an open access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IntechOpen

IntechOpen